UCL Working Papers in Linguistics

19

2007

Department of Phonetics and Linguistics
University College London
Gower Street
London WC1E 6BT
http://www.phon.ucl.ac.uk

Contents

Semantics and Pragmatics

Appendices

Editorial note

UCL Working Papers in Linguistics features research reports by staff and graduate students of the Department of Phonetics and Linguistics, University College London. Reports on phonetic research in the Department appear in our sister publication, *Speech, Hearing and Language*.

This is the first volume of *UCL Working Papers in Linguistics* that is published on-line only. The Internet edition of *UCLWPL* includes an archive of viewable abstracts and downloadable copies of all the papers in this volume.

Richard Breheny and Nikolaos Velegrakis
London, November 2007

http://www.phon.ucl.ac.uk/home/PUB/WPL/uclwpl.html

Linguistic Theory

*Twangling Instruments: Is parametric variation definitional of human language?**

NEIL SMITH & ANN LAW

Be not afeard: the isle is full of noises,
Sounds and sweet airs that give delight and hurt not.
Sometimes a thousand twangling instruments
Will hum about mine ears; and sometimes voices …
[Caliban in *The Tempest* Act III Scene 2]

Abstract

We investigate the hypothesis that what is unique to human language is not recursion but the existence of parametric variation (PV), where this is a way of formulating and uniting two problems: Plato's problem and the limits of typological variation. We suggest identity criteria for parametric variation in language, and see if any of the properties generalise first to other human cognitive domains - particularly music and morality, and second to animal cognition, especially birdsong. We consider several possible outcomes:

a. PV is unique to human language
b. PV is unique to humans but not just to language
c. PV is common to human language and birdsong, but not the rest of cognition
d. PV is common to everything – language, cognition, birdsong …
e. There is no coherent (or uniform) notion of PV.

Our tentative conclusion is (a) above.

1 Introduction

What makes human language unique? For Hockett (1958) and Hjelmslev (1961) it is 'double articulation'; for Jackendoff et al. (2006) it is a rich vocabulary; but the most influential recent suggestion is recursion: either recursion *tout court* or 'phase recursion' (Postma & Rooryck 2007). Recursion has featured prominently since Hauser et al (2002) proposed as an empirical hypothesis that what is unique to

* This paper is a revised version of one presented in Utrecht in April 2007 at the workshop on "Birdsong, Speech and Language: Converging Mechanisms?" For comments, conversation and correspondence we are grateful to Misi Brody, Noam Chomsky, Annabel Cormack, Tecumseh Fitch, Paul Harris, Gary Marcus, Rita Manzini, Amahl Smith and Deirdre Wilson. None of these should be taken as agreeing with anything we say. As we are still in the process of making further revisions any reaction to this instar would be welcome.

language and unique to humans is '(probably) recursion'.[1] This answer is (probably) wrong: recursion is not unique to language, but is characteristic of the Language of Thought (in Fodor's 1975 sense; cf. also Fitch et al, 2005, Smith, 2004) and it may not be unique to humans given the hierarchical structure of canary song (Gardner et al, 2005), the improvisation found in whales (Payne, 2000:135), and perhaps the properties of animal navigation. Another potential answer that we investigate is parametric variation (PV) where this is a way of formulating and uniting two linguistic problems: the puzzle of first language acquisition (Plato's problem, Chomsky, 1986) and the limits of typological variation. Despite Hauser's (2006) spirited promotion of the claim that PV is characteristic of moral judgement, and Smith's (2007) parallel suggestion for music, we think it is plausible to suggest that PV is unique to human language and that the variation found in other cognitive domains and in animal vocalisations is not 'parametric'. We do not wish to exclude other possibilities: the putative uniqueness of PV to human language may be derivative from other characteristics, such as an immensely rich lexicon, and the uniqueness of human language itself may well reside in the constellation of a number of different properties.

The structure of the rest of the paper is as follows: in section 2 we outline the properties of PV in language, concentrating on the difference between parametric and non-parametric variation; in section 3 we see if these properties generalise to other domains of human cognition, in particular morality and music; and in section 4 we then see if they generalise to domains of animal cognition, in particular birdsong. Finally, in section 5 we entertain a number of alternative conclusions as in (1):

(1) a. PV is unique to human language
b. PV is unique to humans but not just to language
c. PV is common to human language and birdsong, but not the rest of cognition
d. PV is common to everything – language, cognition, birdsong …
e. There is no coherent (or uniform) notion of PV.

For historical reasons we exclude logical possibilities such as (f):

f. PV characterises e.g. birdsong but **not** human language

We tentatively endorse (1a) and suspect that, if PV is definitional of human language, it is because PV is a solution to Plato's problem and only human language confronts the learner with this problem in its full complexity. Irrespective

[1] It is not easy to interpret such claims precisely: Tomalin, 2007, points out that 'recursion' is used in at least five different ways so that evaluating Hauser et al's (2002) claim is problematic.

of the validity of restricting PV to human language it is clear that each of the domains discussed is regulated by its own universal principles.

2 Parametric variation in the language domain

2.1 'Principles and Parameters' theory

PV is part of 'Principles and Parameters' theory (Chomsky, 1981a; for overviews and history see: Williams, 1987; Roberts, 1997; Baker, 2003). The human language faculty is standardly described in terms of a contrast between the Faculty of Language in the 'broad' sense (FLB) and a proper subpart of that faculty referred to as the Faculty of Language in the 'narrow' sense (FLN) (Hauser et al, 2002). The former includes a variety of performance mechanisms for parsing and producing utterances as well as our strictly grammatical ability. Many parts of FLB are shared with other organisms from bumblebees to bonobos, but FLN – which may be empty - is by hypothesis unique to humans and unique to language. FLN is characterised in terms of Universal Grammar (UG), the innate endowment which allows children to learn their first language and which defines the basic format of human language. It specifies that human languages consist of a Lexicon and a 'Computational system' (referred to as C_{HL} – the computation for human language). The lexicon consists of a set of lexical entries, each of which is a triple of phonological, morpho-syntactic and semantic features, and with a link to associated encyclopaedic information. Thus a word like *bumblebee* is specified as being stressed on the first syllable (phonological information), consisting of, two elements *bumble* and *bee* (morphological information), being a (count) noun (syntactic information), referring to an insect of the genus *Bombus* (semantic information). The link to encyclopaedic memory will relate this entry to such idiosyncratic information as that bumblebees buzz, are large and hairy, etc. Every natural language has a lexicon containing tens of thousands of such entries whose essential function is to link representations of sound to representations of meaning.

UG also provides a set of exceptionless principles such as structure dependence (Chomsky, 1971), (strict) cyclicity (Freidin, 1999; Chomsky, 2002), the Extended Projection Principle (Chomsky, 1995), etc. which constrain the operation of the computations. Structure dependence is a principle which states that all grammatical operations – phonological, morphological and syntactic – have to be defined in all languages over structures rather than simple linear sequences of element. That is, the possibility of counting the number of words or constituents is excluded *a priori*. For instance, the formation of a question from a congeneric statement as in (2) can refer to syntactic categories, such as 'auxiliary verb', and their movement to a particular position (the Infl node), but not to the 'third word'. In (2) the effect might appear to be the same, but the more complex (3), with the two possible questions in (4), shows that only one process is licit.

(2) The man is in the room – Is the man in the room?

(3) The man who is tall is in the room

(4) a. Is the man who is tall in the room?
b. *Is the man who tall is in the room?

It is significant that this principle acts as a constraint on language acquisition: children learning their first language have their 'hypothesis space' constrained with the result that they never make mistakes like that in (4b).

The extended projection principle (EPP) stipulates that all clauses must have a subject, so (5a) is acceptable but (5b) is impossible:

(5) a. John came home early
b. *Came home early

The status of this principle is somewhat different from that of structure dependence in that some 'pro-drop' languages seem to be systematic exceptions. As we shall see, this exception is only apparent.

In addition to a set of universal principles, UG also provides a set of parameters which jointly define the limits of variation. This is typically conceptualised as the setting of a number of 'switches' – on or off for particular linguistic properties. Typical examples of such parameters in syntax are the head-direction parameter (whether heads, such as Verb, Noun and Preposition, precede or follow their complement), the null-subject (or 'pro-drop') parameter (whether finite clauses can have empty pronominal subjects), and the null-determiner parameter (whether noun phrases can have empty determiners). English and Hindi have opposite values for each of these parametric choices, as illustrated in (6) to (8):

(6) Head-first – "on the table" – Head-last "mez par"
table on

(7) Non-pro-drop – *"Is working" – Pro-drop – "Ø kaam kartaa hai"
work doing is

(8) Non-null D – *"boy has come" – Null D – "Ø laṛkaa aaya hai"
boy come is

Typical examples in phonology are provided by the stress differences characteristic of English and French, and the possibility of complex consonant clusters found in English but not in Japanese. English stress is 'Quantity-sensitive' whereas French

stress is 'Quantity-insensitive', with the result that words with the same number of syllables may have different stress in English but uniform stress in French, as shown in (9):

(9) *América / Manitóba // endurcissement / sentimental*

In English words may begin with clusters of consonants in a way which is impossible in Japanese, with the result that English loans into Japanese appear with the clusters separated by epenthetic vowels, as shown in (10):

(10) *screwdriver // sukuryūdoraibā*

The theory unifies two different domains: typology and acquisition. Variation among the world's languages (more accurately the set of internalised I-languages, Chomsky, 1986) is defined in terms of parametric differences such as whether verbs precede their objects as in English, or vice versa as in Hindi. In first language acquisition the child's task is reduced to setting the values of such parameters on the basis of the stimuli it is exposed to – utterances in the ambient language. Given the strikingly uniform success of first language acquisition, "the set of possibilities [must] be narrow in range and easily attained by the first language learner" (Smith, 2004:83). By hypothesis, the principles do not vary from child to child or from language to language: as Chomsky (2006:183) puts it: "acquisition is a matter of parameter setting, and is therefore divorced entirely from … the principles of UG".

The theory is at once 'internalist' (i.e. it is a theory of states of the mind/brain) pertaining to knowledge which is largely unconscious, and universalist. An immediate implication of this position is that the range of parametric choices is known in advance and, as a corollary, it claims that acquisition is largely a process of 'selection' rather than instruction and that such acquisition is likely to take place in a critical period or periods.

The claim that PV is unique to human language could be trivially true – PV presupposes UG and – by hypothesis – neither animals nor other human domains have UG. However, Hauser (2006) and Mikhail (2007) have both proposed a 'Universal Moral Grammar' akin to UG[2], and Hauser is explicit in his defence of PV in the moral domain. Similarly, Trehub & Hannon's (2006:82) observation that infants react to and learn music spontaneously and rapidly is compatible with a parameter-setting account. In any case, we attempt to make the claim non-trivial by extracting from the wealth of views in the literature some putative identity criteria for PV and seeing if they generalise. That such an extension is plausible is

[2] For instance Hauser (2006:419-420) writes: "Underlying the extensive cross-cultural variation we observe ... is a universal moral grammar that enables each child to grow a narrow range of possible moral systems".

suggested by a number of intuitive commonalities between language, music and morality on the one hand, and between language and birdsong on the other. In all areas there are putative universals, suggesting an innate basis; there are simultaneously clear effects of the environment, suggesting an interplay between genes and learning; and there are parallels in the various stages that organisms pass through in mastering the complexity of the system they are acquiring. We shall claim nonetheless that PV is indeed unique to human language. We look first at PV in syntax and then cast a brief look at PV in phonology.

2.2 PV in Syntax

The main discussion of PV has been in the domain of syntax. Originally, parameters were associated with the principles of UG but following Borer's (1984) work were later located in the lexicon (see Smith, 2004, for discussion). As Kayne puts it: "syntactic parameters are … necessarily features, or properties, of elements of the lexicon" (Kayne, 2005:4). Moreover, for Chomsky and many others (e.g. Chomsky, 1995), the relevant part of the lexicon is restricted to the functional lexicon, where this refers to that subset of the whole which deals with functional categories such as Tense, Complementisers and Determiners, in contradistinction to the conceptual lexicon which deals with substantive categories such as Noun, Verb, Adjective and (perhaps) Prepositions. Thus *bumblebee,* illustrated above, belongs to the conceptual lexicon, whereas an item such as *the* belongs to the functional lexicon. Two major differences are associated with this distinction: members of the functional lexicon characteristically have no encyclopaedic link (except, for the literate, their spelling) and, more importantly, may also lack any phonological content. As illustrated above, some languages (like Latin or Russian) allow empty determiners, others (like Spanish or Greek) allow empty (null) subjects, whereas others (like English and German) allow neither of these possibilities. Such differences are a matter of PV. In contrast, a cross-linguistic difference in the subcategorisation properties of a verb such as *convince*, where *John convinced Mary to go* is grammatical (for some) or ungrammatical (for us) is a matter of (non-parametric) variation in the conceptual lexicon.

In the 'old' system the pro-drop parameter was a possible setting of the EPP; in the 'new' system the pro-drop parameter is instantiated as the condition that (the functional category) AGR = PRO (for pro-drop languages) *versus* AGR ≠ PRO (for non-pro-drop languages). Recently, with the demise of 'AGR' in Minimalism the form of the parameter might be: "[Spec, IP] is obligatory". In either case, pro-drop is a property of particular functional categories.

There is a general consensus on the need for such a (traditional) distinction between functional and substantive categories, but little agreement on how to draw the boundary lines between the two (for discussion, see Muysken, in press). The account of functional categories which is closest in spirit to ours is that provided by

Roberts & Roussou (R&R) (2003:28) who define them on the basis of their behaviour at each of the two levels of representation common to all theories: LF and PF. LF represents the interface with the conceptual-intensional system – i.e. where the language faculty connects with the representation of thought, PF represents the interface with the sensori-motor system where the language faculty connects with systems of audition and articulation. For a linguistic representation to be well-formed all its constituent elements must be interpretable at one or both of these interfaces. Lexical items are designated +p and +l to indicate that they must have an interpretation at PF or LF respectively. R&R define functional categories as "that class of syntactic categories which is not obligatorily +p", whereas lexical categories such as Noun and Verb are always +p, +l. In other words lexical categories must be interpretable at both interfaces, but functional categories such as C(omplementiser), T(ense), D(eterminer) may be overt or covert. If they are 'covert' or empty, they have no interpretation at PF, they are syntactically and semantically present but are not pronounced.[3] The set of interpretable features is provided as a set of substantive universals by UG. A functional head, such as D, is then marked for some feature and will be further specified in a language-particular fashion as having this feature manifest overtly or covertly. The relevant functional category will then be realised in the syntax by either Merge or Move (R&R, 2003:30). To make the discussion concrete they illustrate with a simple example of differences among yes/no questions, marked syntactically by the presence of an abstract morpheme Q. In colloquial French Q is unmarked and questions are indicated just by intonation (*Jean a vu Marie?*). In Welsh, Q is marked and also overt so a question particle *a* is merged in initial position (*A welodd John Mary?*). In English Q is marked and covert so the auxiliary moves to the front (*Did John see Mary?*). Many theories restrict PV to a choice between the values [+/- strong] of functional categories. For such theories parameters are necessarily binary, but such a restriction is independent of the general conceptual claims of the theory.

A further difference between competing theories of PV revolves around the deductive consequences of the choice of a particular value for some parameter. This difference is frequently labelled as a distinction between 'macro-parametric' and 'micro-parametric' variation. 'Macro'-PV is typically exemplified by the head-direction (head first/head last) parameter (Chomsky, 1981a) or Baker's (1996) polysynthesis parameter which determines the overall morphological structure of the language. Each of these parameters has a wide variety of effects, whereas 'micro-PV' – of the sort exemplified by the choice of auxiliary to accompany unaccusative verbs (Perlmutter, 1978; Burzio, 1986) or case realignment in Albanian causatives (Manzini & Savoia, 2007) is characteristically more restricted and has correspondingly fewer repercussions. Roberts (1997:273) lists a number of

[3] Though they may have phonological effects. See Smith, 2005, ch. 15, for discussion.

such parameters and their interrelations, though without adverting to the macro-/micro- distinction.

For our purposes, the distinction is not crucial, though it *is* important for us to determine what constitutes PV and what is non-parametric, a distinction that is again a matter of controversy in the literature. Many, perhaps most researchers deny a principled distinction between macro-, micro- and non-parametric variation. For instance, Kayne is sceptical of validating a distinction between micro- and macro-parametric variation. Indeed, he suggests that "[e]very parameter is a microparameter" (Kayne, 2005:10), and probably "[e]very functional element made available by UG is associated with some syntactic parameter" (p.11) – perhaps never more than one (pp.14-15). He suggests about 100 functional elements. Further, if a language "visibly has some functional element, then all languages must have it" (p.16). The problem is to decide what the features involved which must be "simple and limited in type" (ibid) can be. Kayne uses the term to "characterize all cross-linguistic syntactic differences" (ibid p.6) and exemplifies the concept with the contrast in word order between the English and French pairs: *too rich/rich enough* versus *trop riche/assez riche*. Similarly, Manzini & Savoia (2007, Rita Manzini – pc) deny any distinction both between macro- and micro-parametric variation and also between parametric and non-parametric variation. Their respective positions differ in that while Kayne arrogates all change to the functional lexicon by postulating large numbers of abstract (silent) categories (cf. Roberts & Roussou, 2003:24), Manzini & Savoia do it by treating all variation on a par, as simply 'lexical', with no special status for any functional domain, no silent categories and, crucially, no non-parametric variation.

We have two reactions to this position: first, we believe that it is too broad, as not all syntactic differences seem to us to merit the description 'parametric'; second, we are not convinced that the contrast between macro and micro should be given up so easily. Baker argues that there is no principled difference among types of PV – micro-PV is that which is "localized to a particular configuration" whereas macro-PV shapes "every phrase of every clause", but we share Chomsky's intuition that "there seems to be a major distinction between macroparameters like the head parameter … and what some call 'microparameters'", and his "doubt that there will ever be much of a theory about these micro-variations" as they are "basically accidents" – "what a person cannot know in advance" (pc December 06). We are happy to remain non-committal on the 'macro-/micro-' dimension, but it is important that we differentiate the parametric and the non-parametric and we will attempt to give identity criteria for PV.

2.3 PV in Phonology

The domain of PV in syntax is, crucially, the set of functional categories, but there is no comparable constraint in phonology. Rather, as indicated in the examples of

stress and possible onset clusters above, parameters are associated with words, syllables, vowels, etc. (cf. Dresher & Kaye, 1990, Dresher, 1999). It follows that PV is not restricted to the functional lexicon and could be operative in areas without such a category. Accordingly we need to identify more abstract properties of the concept. A crucial property of PV is that it gives rise to a situation in which language acquisition is cued or triggered rather than 'learned'. Learning in the traditional psychological sense (i.e. a process involving hypothesis formation and testing, association, generalisation, and so on) plays little role in first language acquisition. As Chomsky (1980:134) put it: "in certain fundamental respects we do not really learn language; rather, grammar grows in the mind." Language acquisition, however, is clearly dependent on external input, and this input is said to trigger or cue linguistic development. We think that the notion 'cue' is likely to be more fruitful than 'trigger', but the basic distinction is that between either of these and traditional learning.

2.4 Typology and acquisition

We need to recapitulate a little. PV is variation within a narrow range defined by universal principles – languages are not like Joos (1957:96) suggested: they cannot "differ from each other without limit and in unpredictable ways". It distinguishes among possible mentally represented states of the language faculty: i.e. it defines the range of possible I-languages (Chomsky, 1986) and makes available a typology of the world's E-languages. PV unites – as an empirical claim - two domains: typology and first language acquisition (cf. e.g. Fanselow, 1993:xvi), but it is important to note that acquisition has conceptual priority, with the typological exploitation of the claims being derivative. If it is correct to claim that first language acquisition consists in setting parametric choices these must have priority over any taxonomy based upon them. The typological claims being made are nonetheless not trivial, accounting as they do for the interdependence of particular properties across languages: e.g. the fact that SOV, PostP, Suffixation, etc. cluster together. If acquisition is a matter of parameter setting and if there is no negative evidence available to the child then all the possible alternatives are antecedently known or innate[4] and the child's task in learning its first language is a matter of selecting a grammar on the basis of the particular properties of the input, rather than needing instruction (cf. Piattelli-Palmarini, 1989).

This claim of 'antecedent knowledge' or 'knowledge without experience' has a number of implications. The first of these is that parametric choices may give rise

[4] The term 'innate' is problematic because of different usages (see Marler, 1999, for discussion). An obvious contrast between (human) language and e.g. birdsong is that there are no known linguistically relevant genetic differences between humans, whereas there are known and experimentally manipulable differences among different populations of canaries (Mundinger, 1999).

to cascade effects: coming to know one fact (e.g. that Verbs precede their objects) licenses knowledge by the learner of other facts (e.g. that Prepositions precede their objects) without further exposure. 'Cascades' have become unfashionable because of the dissociation of the properties associated with (notably) the pro-drop parameter (for discussion, see Ackema et al, 2006). This reaction may have been hasty: cascades could be operative in the domain of acquisition even if there is such dissociation. That is, the child leaps to the 'cascade conclusion' (i.e. selects one parameter on the basis of the setting of a distinct but related parameter) unless there is evidence to the contrary, thereby solving Plato's problem (in part) (for discussion, see Smith & Cormack, 2002).[5] A second implication of 'selection' is that parametric choices allow no analogical formations ("analogy works only for non-parametric choices", Smith, 2004: 121), and license only a subset of the logically possible mistakes a learner could make. Further, as pointed out originally by Hyams (1987:18), this conception also makes available a wider selection of triggering data, and gives rise to the kind of 'network' described by Roberts (1997:275) where parameters are not independent of each other. The details are complex, contentious and of only limited relevance to our current concerns, so we turn next to a more detailed discussion of the distinction between parametric and non-parametric variation.

2.5 Identity criteria for parametric variation

The theory of PV stipulates that the range of choices is 'antecedently known', and this basic property, our first criterion, correlates with a number of others which distinguish PV from non-parametric variation, and allow us to provide identity criteria for PV. Our second criterion is that variants licensed by parametric choice must be cognitively represented. Consider acclimatisation, specifically sweating. We have a critical period for setting our sweating switch: experiencing hot and humid weather in the first three years of life leads to a different setting from exposure to different conditions. These settings cannot be significantly altered after the critical period. Despite a certain superficial similarity, this is not PV because the different states are not (mentally) represented and have no cognitive effects. Comparable remarks obtain with regard to Hauser's (1997) discussion of the notion 'deception'. It is perhaps mildly surprising to learn that rhesus monkeys practise systematic deception but the evidence is cogent. It is less plausible to think that stomatopods (shrimps) practise it simply because we do not attribute to them the 'cognitive processes' necessary for deceit, even if their behaviour could be interpreted as deceptive if they were intellectually comparable to us in other respects. For the same reason, Hauser's 'sceptical honeybees' are probably best

[5] A clearer example of a cascade may be provided by the contrast 'Signed/Spoken' where choice of the value 'Signed' entails a range of other choices such as +Pro-Drop. In fact, the variation in the syntax of signed languages seems to be much narrower than in spoken languages.

treated behaviourally rather than cognitively. Differences in neurobiology and the absence of relevant experimental results make it difficult to determine where songbirds come in this hierarchy.

Our third criterion is systematicity. A simple example is provided by irregular morphology of the type exemplified by the impossibility of **amn't* in (most varieties of) English, or the kind of defective paradigm seen in Latin *vis-vim-vi.* We do not consider this to be PV because it is by definition not systematic and hence we could not plausibly acquire knowledge of it by any process of triggering. Our fourth criterion is dependence on the input. An example is provided by the individual variation in e.g. consonant harmony in phonological development (cf. Smith, 1973:163), or the variation in the choice of initial or final negation in syntactic development (cf. Smith, 2005:29). For instance, two children in essentially the same environment may produce the adult *duck* as [gʌk] and [dʌt] respectively. These are both manifestations of consonant harmony, but they do not count as PV because the particular variants chosen appear to be independent of the input. There is, of course, some dependence on the input: the form of the vowel, the absence of a labial articulation in either initial or final position, but the choice of [gʌk] or [dʌt] is idiosyncratic to a particular child. It instantiates a general strategy (consonant harmony) which consistently produces variants none of which occurs in the adult language. 'Dependence on the input' thus stipulates that parametric choices must reflect possibilities in the target (ambient) language. To give a syntactic rather than a phonological example, the development of negation typically goes through a stage in which the negator is peripheral, either initial or final. Individual children then differ such that one child learning English may say '*no like cabbage*' and another '*like cabbage no*'. Again we take this variation to be non-parametric, even though it may be hard to differentiate it from UG-licensed errors of the sort described by Crain and his colleagues (cf. Crain & Pietroski, 2002), where a child produces a form which never occurs in the input (e.g. '*What do you think what pigs eat?*') because the structure is licensed by UG and so occurs in other languages.[6] To our knowledge, no language allows peripheral negation with the negator either initial or final. Despite this potential difficulty, we think that the case of consonant harmony in phonology and negation in syntax makes the conceptual contrast between parametric and non-parametric variation clear.

Our fifth and final criterion is that PV must be deterministic: that is, the input to the child must be rich enough and explicit enough to guarantee that a parameter can be set. If the input does not meet this requirement we are dealing with non-parametric, random, variation. This criterion arises from a reinterpretation of Smith & Cormack's (2002) discussion of 'parametric poverty'. They suggested that there are 'random settings': given the same input, different children might assign the

[6] We take it that such over-generalisation is a sign that the child has, temporarily, mis-set the relevant parameter.

same parameter different values. They illustrate this with the possibilities for 'sequence of tense' phenomena. For many speakers *Did you know that Emily is ill?* is simply ungrammatical, and only *Did you know that Emily was ill*" is possible; for us both are fine, though with interpretational differences. Smith & Cormack suggest that the parameter has been fixed at random; we prefer to think that such 'randomness' reveals that the variation is non-parametric. Comparable examples can be found in phonology: Moira Yip (2003:804) gives an example where some people treat a post-consonantal glide as a secondary articulation of the consonant, others as a segment in its own right: "the rightful home of /y/ [is] underdetermined by the usual data, leaving room for variation". Her conclusion is that "speakers opt for different structures in the absence of conclusive evidence for either". Again that indicates for us that the variation is non-parametric.

In Table 1 we summarise and exemplify these criteria, all of which are common to syntax and phonology.

1. The range of choices must be antecedently known; hence acquisition is a matter of selection rather than instruction	
Parametric Movement; ellipsis	Non-parametric Irregular morphology; lexical exceptions (*likely/probable)*
2. Parametric choices must be mentally represented	
Parametric Stress; word-order.	Non-parametric Sweating; consonant harmony (e.g. 'duck' ⇨ [gʌk] or [dʌt])
3. Choices must be systematic - variations are not accidents	
Parametric Wh-movement (covert or overt)	Non-parametric Defective paradigms (**amn't*; Latin *vis-vim-vi*)
4. Choices must be dependent on the input and hence correspond to a possible state of the adult language	
Parametric Quantity-sensitivity Word order – head direction	Non-parametric Consonant harmony Early negation (*no computer on/ computer on no*)
5. Choices must be deterministic	
Parametric Pro-drop Complex onsets in phonology	Non-parametric Sequence of tense Post-consonantal glides

Table 1: Identity criteria for Parametric Variation

The restriction to the functional lexicon of syntactic variation[7] is an epiphenomenon of the fact that PV operates over elements of the appropriate domain: the units manipulated in the syntax are words (i.e. elements of the lexicon which appear in the numeration); the units manipulated in the phonology are syllables and feet (and sometimes phonological words). The criteria are intended to be jointly sufficient and individually necessary to identify PV. We should add that some phenomena (e.g. cascade effects) may be sufficient to license the conclusion that there is PV even though it is not possible to make this a necessary condition. Hence we need to distinguish 'cascade effects' - knowledge without *further* input - from the fixing of a form as in consonant harmony[8] or peripheral negation where the input does not provide any basis for the choice. Assuming the continuity hypothesis (Pinker, 1984) and the idealisation to "instantaneity" (Chomsky, 1981b:224) this fixing reduces to the claim that the grammar the child has attained must generate forms of the ambient language. As a final point it should be noted that occurring in a critical period is compatible with PV, but is not evidence for it.

Assuming that, despite the different interpretations available in the literature, the nature of PV is clear, we now proceed to see if and how it generalises. We shall look in each case at the range of variation involved to see whether it is comparable to parametric variation in language, and whether the acquisition of the relevant ability is reminiscent of language acquisition.

3 Generalisation of Parametric Variation in the domain of human cognition

It might appear most plausible to generalise PV to areas which, like language, rely on auditory input: music and the analysis of environmental sounds (noise). However, the parallels may not be quite as obvious as this suggests: music and noise are necessarily tied to the auditory channel whereas signed languages show that the language faculty is not restricted in this way. We look briefly at the analysis of noise, before concentrating on moral judgement and music.

The case of 'noise' is interestingly parallel to language. Smith (1999) observes that dissociation patterns in the various agnosias (cortical deafness, auditory agnosia, pure-word deafness and phonagnosia) reveal remarkable richness of structure, and speculates that growing up in different auditory environments (a jungle, by the sea, in the desert) might result in the setting of different parametric values for auditory discrimination. Work by Saygin et al (2005) shows an appreciable amount of parallelism between the two domains but also some notable differences, (e.g. in the brain areas involved) and the authors are sceptical that PV

[7] It is possible that some variation – e.g. that for which the subset principle and markedness are irrelevant – is still non-lexical. (cf. Safir, 1987:87).

[8] If Smith, 2005 ('gucks'), is correct, consonant harmony would not be mentally represented.

is at work (Fred Dick, p.c.). In the absence of solid ontogenetic evidence we leave the subject for future work.

3.1 Moral judgement

Moral judgement is not an axiomatic system with the deductive structure of a theory of language and is correspondingly resistant to formalisation. However, there are striking parallels, including such properties as being 'internalist', exploiting a competence/ performance distinction, consisting in knowledge which is largely unconscious, being acquired under poverty of the stimulus[9] - triggered rather than learned; universal – hence uniform and innate; and dependent in part on the development of a Theory of Mind (ToM) module (cf. Papafragou, 1998). Building on Chomsky's insights about language and his suggestions about morality (cf. Chomsky, 1975), e.g. that "Hume ... pointed out that the foundation of morals must be what we nowadays call generative grammar" (Chomsky, 2003:40), Hauser (2006, cf. Hauser et al, 2007) and Mikhail (2007) have elaborated an analysis of moral judgement which has many of the hallmarks of our knowledge of language. After postulating a universal principle of justice as 'fairness' Hauser states that "[a]dopting the analogy to language, one would expect a universally held principle of fairness that varies cross-culturally as a function of parametric variation; experience with the native environment triggers the culture's specific signature of fairness and fair exchange" (2006:72). More worryingly, he writes that "[i]n the same way that our universal grammar provides a toolkit for building a specific grammar, in which certain principles and parameters hold and others do not, our universal moral grammar provides a different toolkit, enabling us to implement particular principles and parameters but not others" (2006:74). This formulation is unfortunate in that it suggests that the *principles* can vary from culture to culture rather than just the parametric choices.

Hauser's statement is inconsistent in claiming that Universal Moral Grammar works "in the same way" as UG, but allowing that certain principles can vary. Interestingly, he also raises the question whether parametric moral cascades occur, asking "Do moral systems work like language in the sense that choosing to set certain parameters influences subsequent settings?" (2006:300). He is also explicit that his position is one where the moral system children construct "depends upon their local culture and how its sets the parameters that are part of the moral faculty" (2006:303).

Apart from Hauser's heterodox interpretation of principles and parameters, there are also systematic asymmetries between language and morality which lead to doubts about the validity of the parallel between the two domains. First, the building blocks of moral judgement are actions, causes, consequences and their

[9] This may be contentious, but the stimuli which give rise to our largely unconscious moral judgments may be concealed in the linguistic input; see Smith, 2007, for discussion.

arrangement over a period of time. Second, the proprietary lexicon of the moral system is not comparable to that of a natural language. That is, one might expect that any cognitive module should have its own lexicon disjoint from that of any other. However, once one goes beyond universal principles such as 'justice as fairness' (see below) the units of moral judgement are a subset of the items in the conceptual lexicon of the culture the system is embedded in. Such systems dispose of a huge set of concepts from the language of thought and these are subject to variation, but the variation lies more in the detail of encyclopaedic knowledge and is idiosyncratic rather than parametric. For these reasons we are sceptical of the force of Hauser's suggestions: we think the typological claims may be plausible but that there is no evidence for PV in the domain of acquisition. Although there is some literature on children's development of moral judgement (e.g. Harris, 2000) there is almost no systematic cross-cultural work on anything that might be considered parameter-setting. Accordingly, we look first at some putative universal principles, as in (11), and specific cultural parameters, as in (12), and evaluate how close they are to what is found in language. In a nutshell, we are happy with Hauser's claims about the existence of universal principles (though these should not be subject to cross-cultural variation) and their parallels with what is found in language; we are not convinced that he has made a case for PV in anything like the current linguistic sense.

(11) Universal moral principles:

a. Justice as 'fairness' in the distribution of advantages. Hauser (2006:83) states that "fairness is a universal principle with the potential for parametric variation and constraints". Following Lakoff (1996) Hauser (2006:75) suggests that such variation might be exemplified by the difference between equal or scalar distribution of responsibility ('effort is equitably shared' *versus* 'those who can do more have greater responsibilities') or a needs-based *versus* scalar distribution of rewards ('those who need more, get more' *versus* 'those who work harder, get more').
b. Obedience to authority
c. Incest taboo; nature of 'shame'.

(12) Putative cultural parameters of moral judgement:

a. There are clearly differences in the underlying philosophy motivating the fairness of distribution: for instance, a deontological versus utilitarian basis. This will be manifest in the existence of a social versus an individual basis of the kind (supposedly) characteristic of communism and capitalism. There will also be differences in the power

relations which are effective in any such distribution, as seen most clearly in the limits of the franchise. Everyone may have the vote, only men, only property-owners, only property-owning men over 21; only whites, only members of a particular religion, and so on.

b. Differences in the nature of authority: e.g. Theocratic/Secular, leading to further differences such as 'sin' versus 'crime'.

b´. Differences in the degree of obedience to authority, ranging from the anarchist to the totalitarian.

c. Definition of kin for incest; context in which 'shame' is felt.

The crucial issue for us is whether moral judgement satisfies the identity criteria for PV. Table 2 gives a summary overview.

1. Antecedently known	Dubious. The principles, such as 'Fairness' seem to be known in advance, but not the cultural variants.
2. Mentally represented	Yes
3. Systematic	In part: there is apparently systematic and non-systematic variation.
4. Dependent on the input	Yes but with differences between adult and child
5. Deterministic	Not consistently

Table 2 - Parametric Variation in moral judgement?

That the elements of the moral system are mentally represented seems straightforward. Whether this knowledge is antecedently known seems plausible for universal principles and the existence of categories such as 'good' and 'bad', but that parametric choices are available beforehand seems highly unlikely. Moral choices are usually systematic but as in language there is apparently arbitrary variation as well, as in the degree of 'moral disgust' felt by different individuals confronted with the same phenomena such as incest (see Jones, 2007). Here and more generally dependence on the cultural input is crucial. One's moral imperatives and one's judgements more generally are clearly likely to differ depending on whether one was schooled in a madrasa, a convent or a school in Beijing. The caveat expressed under 'dependent on the input' is prompted by the ontogenetic development of moral concepts, notably by the child's developing notion of lying. Children go through a stage in which they consider all untruths to be lies; only when they have developed a functioning Theory of Mind do they master the complexity of the adult concept, where some intent to deceive is necessarily included. That is, the child's use of the term 'lie' is unlike the adult's,

though it may correspond to something different in the adult language such as the phrase 'say something untrue'. This is unlike the kind of UG licensed error mentioned with regard to the Crain example above in the following respect. Children's moral misconception with regard to lying results from their representing only a subset of the information present in the adult form, whereas their syntactic error results from hypothesising a grammar partially disjoint from that of the adult language.

Two final points should be made: first, the unification of typology and acquisition seems not to be salient or even possible as yet. Second, the validity of the parallels Hauser draws in his (2006) relies on Chomsky's original notion of PV in which parameters were associated with principles of UG (*mutatis mutandis* with principles of UMG (Universal Moral Grammar), rather than the current one where they are associated with (features of) lexical items. In neither case would it be possible for the *principles* to differ from case to case in the way Hauser suggests. This makes comparison difficult, but on either interpretation PV seems to be different here than in language.

3.2 Music

We turn next to the scope for parametric analysis in the musical domain, looking at a number of properties characteristic of knowledge of music and its acquisition. First, we consider some obvious background differences between music and language. In music, the building blocks are (sequences of) notes arranged (as in language) into metrically articulated "hierarchical recursive structure" (Jackendoff & Lerdahl, 2006:38), there are fixed and discrete pitches for each mode, there is typically isochronicity, and there is no semantics. It is also striking that there is no musical equivalent of a natural language lexicon. It is true that Peretz & Coltheart (2003:690) refer to a 'musical lexicon', but this is characterised as "a representational system that contains all the representations of the specific musical phrases to which one has been exposed during one's lifetime" and a processing device designed to match incoming stimuli with stored representations. This makes it look like a parser and radically unlike the language lexicon. Given the point of departure of this paper, it is also worth putting on record that we do not find Jackendoff & Lerdahl's claim of recursion for music persuasive. What they describe is iteration rather than recursion.

Our focus of interest is the 'musical idiom' (Jackendoff & Lerdahl, 2006) and how hearers become 'familiar' with it. Specifically, "How does a listener acquire the musical grammar of MI [musical idiom – NS/AL]" and "What pre-existing resources in the human mind/brain make it possible for the acquisition of musical grammar to take place?" (ibid p.34). They also raise the issue of the cognitive specificity of the musical capacity which is explicitly parallel to the contrast between FLB and FLN, though their paper has surprisingly little discussion of any

of these issues, simply raising – and dropping – the observation that "each idiom will have its own characteristic structures, created out of the interaction of idiom-specific tonal and metrical principles with universal principles of tension and relaxation." (ibid p.58).

In looking at parallels and differences between language and music, the first observation to make is that the perceptual apparatus underpinning music is the same as that for language, and is moreover shared by some animals, and "the most parsimonious interpretation of the available evidence is that infant skills are a product of general perceptual mechanisms that are neither music- nor species-specific" (Trehub & Hannon, 2006:91). Only some skills, however: they point out that perception of relative duration is necessary to understanding rhythmic structure in both language and speech, even though these differ (ibid p.85), and musical meter is species-specific (ibid p.87). Moreover, as they emphasise (Trehub & Hannon 2006:82) infants react to and learn music spontaneously and rapidly in a way that animals do not.

This largely common perceptual foundation gives rise to a variety of properties shared by the two systems. Even though their status may be controversial (cf. Nettl, 2000, passim), there are clear musical universals – the use of octaves, scales of less than or equal to 7 pitches, the use of tonal scale systems, rhythm based on 2 or 3 (Brown et al, 2000:13-14; see also Peretz 2006). Moreover, as Nettl (2005:56-57) puts it: "People tend to absorb the fundamental grammars of their own language and their own music very early, to know very quickly whether a word belongs to their own language, and whether a particular interval or chord is proper in their own music". Similarly, "the compatibility of Western and traditional African music" (Nettl, 2005:58) suggests the operation of a typological parameter which is shared by these two areas. Nettl goes on (2005:67) to locate this compatibility in terms of harmony and draws a contrast in this respect between African and Middle Eastern musics.

In at least some cases (e.g. the octave) there may be a simple physical explanation for the universality, but there are interesting data on infants' sensitivity to different aspects of music (see e.g. Trehub, 2000) which are suggestive of parametric variation. Synchronised movement to music is universal (Trehub & Hannon, 2006:86); there are aspects of musical ability such as the possession of absolute pitch which seem to require exposure during a critical period (Brown et al, 2000:13); grouping, rhythm and meter are all underpinned by innate abilities (Trehub & Hannon, 2006:82), and musical grouping eventuates in a "hierarchical recursive structure" (Jackendoff & Lerdahl, 2006:38; but cf. the reservation above): that is, there is a putative 'cognitive homology' of metrical grids in language and music (Jackendoff & Lerdahl, 2006:42f.). Trehub & Hannon (2006:81) report that infants outperform adults on detecting contextually appropriate changes (a change in the correct key as opposed to an incorrect key – adults only detect the latter, infants detect both). This ability is strikingly parallel to the abilities of infants to

make categorial discriminations such as the l/r distinction which they lose unless the ability is reinforced by ambient input (cf. Jusczyk, 1997).

To make the discussion explicit, a specific example of what might constitute a parametric choice with obvious triggers for acquisition is provided by the variety of scales found around the world. There is a clear contrast among the Diatonic (Western heptatonic) scale, the Slendro (Indonesian pentatonic) scale used for the gamelan, and characteristic also of Chinese music, and the Śruti (Indian) scale with (usually) 22 śruti per octave.

The commonalities can be generalised. 'Musical idiom' is parallel to language not only in exploiting the same perceptual apparatus but in being 'internalist', largely unconscious, in becoming established in a critical period for (e.g.) absolute pitch, in showing the possession and loss of categorical discrimination (Trehub & Hannon, 2006; cf. Jusczyk, 1997), etc. But these similarities may be only superficial. Metrical structure is peculiar to music and language, whereas 'grouping' structure is also characteristic of other aspects of cognition.

Moreover, all this is neutral with regard to the question of PV in the acquisition of these abilities. However, Trehub & Hannon (2006) provide potentially relevant evidence for the generalisation of the typological to acquisition in their discussion of simple and complex meter: "After 6-month-old infants listen to a folk tune with a simple or complex meter for 2 min, they prefer variations that disrupt the original meter to those that preserve it both for simple and for complex meters ... 6-month-olds' differential responsiveness to meter-preserving and meter-disrupting variations parallels the ratings of Bulgarian and Macedonian adults. By 12 months of age, however, infants respond differentially to meter-preserving and meter-disrupting variations in simple-meter contexts but not in complex-meter contexts" (Trehub & Hannon, 2006:88). That this could be PV is suggested by the difference they note between this and other abilities: "the early acquisition of adultlike biases in perceiving rhythm and meter contrasts with the protracted developmental course of sensitivity to hierarchical pitch structure" (Trehub & Hannon, 2006:89).

We are now in a position to give in (13) and (14) a summary list of Universal principles and putative cultural parameters together with, in Table 3, an indication of whether these satisfy the identity criteria for PV.

(13) Universal musical principles:
   a. Tonal encoding of pitch
   b. Tonic as centre; hierarchical organisation of cognitive pitch space distances; tonal relaxation and tension
   c. Exploitation of Scales
   d. Exploitation of Metres

(14) Cultural parameters of music
a. Choices among scales
b. Choices among metres
c. Choice of harmony

1. Antecedently known	Probably in part (e.g. scales).
2. Mentally represented	Yes
3. Systematic	In part: there is apparently systematic and non-systematic variation.
4. Dependent on the input	Yes
5. Deterministic	Probably not but ontogenetic evidence is lacking

Table 3 - Parametric Variation in music?

As before such phenomena as occurring in a critical period are compatible with PV, but are not evidence for it. As with moral judgement, the putative unification of typology and acquisition is moot, as acquisitional evidence for the latter is slim, with much of the literature on music restricting itself to typology. It is also relevant that with harmony (which is of recent development) the limitations of the human voice mean that it can be perceptually triggered but not produced by the learner.

4 Generalisation of Parametric Variation in the domain of animal cognition?

The attempt to generalise PV to other domains of human cognition such as music is moot. Is it more plausible in the domain of animal cognition, in particular birdsong in oscine birds? We begin as before by specifying some of the background differences between the two domains. It is necessary to differentiate birdsong, bird calls and bird mimicry, as PV is of potential relevance only to the first of these. The building blocks of birdsong are 'notes', 'syllables', 'phrases', 'motifs', 'types', and 'bouts', arranged in a hierarchical structure (Brenowitz et al, 1997). As Gardner et al (2005:1046) put it: "Canary song is hierarchically structured: short stereotyped syllables are repeated to form phrases, which in turn are arranged to form songs". The 'notes' on the lowest level of the hierarchy may be supplemented by whistles, buzzes and trills. Assuming that PV operates over elements of the appropriate domain, this difference in 'building blocks' is expected. A more salient difference is that there are about 9000 species of bird (including 4000 species of songbird) and "there is enormous between-species variation in song structure, as well as in the characteristics of song acquisition and production" (Bolhuis & Macphail, 2001:429; cf. Suthers, 1999). This is important as birdsong typology is frequently cross-species, but comparability with language demands intra-species treatment.

Humans are an isolated species and no-one would describe our differences from other primates in terms of parametric choices. The various species of birds are related but (generally) do not inter-breed, suggesting that PV is perhaps irrelevant, as 'setting parameters' would be impossible, especially as "each songbird species seems to go about the process of learning to sing in its own way" (Marler 1999:295). One basic difference is seen in the contrast between 'open-ended learners' (such as canaries) and 'age-limited learners' (such as zebra finches). This contrast could not usefully be viewed as parameter setting, but corresponds if anything to the possibility of learning new structures and/or lexical items throughout life.

As striking as the variety of birdsong is the absence in all species of syntax or semantics. Calls may have some minimal content – e.g. indicating predators, and song has territorial and mate-selection functions, but it mainly manifests variety for its own sake. The absence of semantics is reminiscent of music but the putative absence of syntax[10] is a more radical distinction. It is accordingly necessary to specify what is meant by the contentious observation that there are no 'sentences' in animal communication (Marler, 2000:31). It is true that birdsong shows the recombination of learned sequences in many different ways but we agree with Marler that this is not like human syntax and *a fortiori* could not exhibit recursion. (Marler, 2000:39). Finally, and again reminiscent of the observation about a musical lexicon above, it seems that birdsong repertoires are largely memorised rather than 'creative'.

Despite these obvious differences there are also striking commonalities as pointed out by e.g. Kuhl (1999) and Doupe & Kuhl (1999). Thus, birdsong is parallel to language in being internalist, unconscious, reliant on auditory input, acquired in a sensitive period, and universal (for particular species).[11] Like human infants, birds are sensitive to their own 'language': "young birds must hear the songs of their own species in order to learn them, but when faced with a potentially confusing array of songs, they are able to select the ones of their own species to serve as learning templates" (Whaling, 2000:69). This is comparable to the sensitivity to their own language from intra-uterine experience that new-born infants manifest but is presumably prior to any setting of parameters.

---

[10] When birdsong specialists talk of 'syntax' they are referring to what linguists would call 'phonotactics' – the syntax of phonology. This terminological confusion bears closer scrutiny. Marler makes a distinction between 'Phonocoding' - "the ability to create new sound patterns by recombination simply to generate signal diversity" and 'lexicoding' - "when meaningful elements are syntactically joined" (Marler 2000:31). The latter is supposedly unique to humans. The reference to 'meaning' suggests that what is intended is really semantics rather than syntax: birdsong might well have some syntax but with no semantic function. It is significant that when Suthers (1999:51) observes that brown thrashers probably have in excess of 1000 syllables in their repertoire, he refers to 'syllables' – a phonological construct rather than a syntactic one.

[11] Though often only the male sings.

In later development songbirds appear to go through comparable stages to language-learning infants. Thus 'template memorisation' and 'vocalisation matching' correspond to the child's acquisition of the phonological representations of lexical items of its language and the subsequent mastery of their production. Sub-song; plastic song, full (crystallised) song (cf. Bolhuis, 2005), correspond to babbling, and early and late phonological mastery. There are even closer parallels, of which we mention three.

First, Kuhl observes that in humans "language input alters the brain's processing of the signal, resulting in the creation of complex mental maps" (Kuhl, 1999:424). This is an example of the perceptual magnet effect and the native language magnet whereby infants lose some of their innate discriminatory abilities as a function of exposure to a specific ambient language. A parallel to this magnet effect is found in birdsong. Gardner et al (2005) report on canaries' ability to imitate ill-formed song when young, and how this ill-formed song is then "reprogrammed to form typical canary phrasing".

Second, is the issue of 'selection' versus 'instruction'. Just as with language acquisition there are "genetic contributions to the development of *learned* behaviors" (Marler, 1999:311) and oscine birds show interesting parallels with the 'selection' account of acquisition (Piattelli-Palmarini, 1989). "Songbirds actually inherit much of the information required to generate a normal species-specific vocal repertoire … as though memorization is based not on instruction … but on selective processing, imposed on a fund of *innate* knowledge that is to some degree unique to each species" (Marler, 1999:315).

Third is the putative ability of starlings to acquire recursive grammars. The claim (Gentner et al, 2006; Marcus, 2006; cf. Jackendoff et al, 2006) is that European starlings can be trained to acquire complex recursive grammars and that this "challenge[s] the recent claim that recursion forms the computational core of a uniquely human narrow faculty for language (FLN)" (Gentner et al, 2006:1206). We are convinced by Jackendoff et al's (2006) rebuttal of this claim, casting doubt on both the substance and the implications of the research.

A final parallel between birdsong and language is that birds manifest two kinds of variation: regional (dialectal) and individual (idiolectal); that is, there are birdsong dialects (see e.g. Catchpole, 1991). Searcy & Nowicki (1999) report that, as measured by courtship display to recordings, sparrows are sensitive to differences to the song of conspecifics from New York and Pennsylvania. They emphasise that "geographical variation is the product of chance historical factors" (ibid p.591), and that whereas individual variation is adaptive, geographical variation is not. Presumably the development of dialects is crucially dependent on the selection of particular patterns in the input in language as in birdsong.

The crucial consideration for us is whether any of these characteristics satisfy the criteria for PV. Many, such as occurring in a critical period, having hierarchical structure, illustrating 'magnet' effects, and so on are compatible with PV, but are

not evidence for it. Liu et al (2004) demonstrate individual differences in song learning by zebra finches, including cases where variation was shown among siblings who were "members of the same clutch, and they all imitated the same model" (2004:18178). Crucially, however, the juveniles gradually converge (2004:18180) on the 'same' adult song. This kind of individual variation looks comparable to the individual variation found in consonant harmony in the acquisition of phonology, and the selective choice characteristic of PV plays no role (cf. Doupe & Solis, 1999).

In Table 4 we set out the usual contrasts.

1. Antecedently known	Yes, for some species, but imitation + improvisation suffice without PV.
2. Mentally represented	Yes.
3. Systematicity	Mixed. Gardner et al (2005) report on the perceptual magnet effect and "reprogramming" by canaries. Importantly, however, the reprogramming "occurred in the absence of any exposure to normal canary song", leading to the conclusion that "inferred innate rules forced a complete reprogramming of the imitated song" (2005:1047). None of this seems to be comparable to PV in human learning of phonology.
4. Dependent on the input	We know of no cascade effects in the development of birdsong, but it is clear that many of the details are input-dependent. It is also clear that some individual variation in song production is not parametric but comparable to that found in consonant harmony. The identifiable juvenile nature of the imitations cited by Liu et al (2004) makes it reasonably clear that juvenile birdsong does not always correspond to the adult state.
5. Deterministic	Yes.

Table 4 – Parametric Variation in birdsong?

It is time to take stock. As before, the unification of typology and acquisition is moot, the abilities of starlings are impressive in terms of short-term memory, but the best case for PV may reside in the existence of regional dialects. These dialects appear to be learned rather than inherited (Catchpole, 1991:288; cf. Searcy & Nowicki, 1999) and are clearly functional. However, even here there is no evidence (that we know of) that the choices are antecedently given, and the

complexity of what is learned may not be sufficient to motivate the need for PV: there may be no avian equivalent of Plato's problem.

We began by expressing our scepticism towards Hauser et al's claim that what is unique to human language is recursion. Can we provide an alternative conclusion?

5 Conclusions

We started by envisaging a number of possibilities, summarised below:

(15) a. PV is unique to human language
b. PV is unique to humans but not just to language
c. PV is common to human language and birdsong, but not the rest of cognition
d. PV is common to everything – language, cognition, birdsong …
e. There is no coherent (or uniform) notion of PV.

(For historical reasons we excluded logical possibilities such as (f):

f. PV characterizes e.g. birdsong but **not** human language)

A definitive answer is too much to hope for. However certain tentative observations are in order. Critical periods and universals may both be necessary but are clearly not sufficient for PV. As far as typology is concerned, PV can be adduced harmlessly to describe the limits of variation in all the relevant domains: language, moral judgement, music and birdsong. We view this, however, as somewhat banal. The core interest of PV in linguistics lies in its solution of Plato's problem and the unification of typology and acquisition, so the crucial issue is whether the generalisation also applies in music, moral judgement and birdsong. We suspect that it applies in none – that is, (a) above is correct, but evidence in other domains is lacking. As regards music it is tempting to see a role for PV in the determination of the musical idiom internalised by individuals, but the infantile evidence is lacking and we are concerned that the parallels even then are forced. The reason is that in language (syntactic) PV is limited to the lexicon and there is no comparable construct for music.

So our conclusion is muted. We reject the pessimistic (e) above though on bad days it looks fairly persuasive, and is close to Haider's (1993:1) position that PV is epiphenomenal,[12] and settle on (a). But the reasoning is less decisive than we had hoped, leaving unanswered the question WHY should PV be unique to language? We suspect that the answer is going to be messy and complex. Human language is

[12] We disagree with Haider's claim that "UG cannot have any guiding effect on language acquisition" (1993:4); for discussion cf. Smith, 1990.

remarkably complicated and it is that complexity which makes Plato's problem so hard and so interesting. The other domains despite their interest and richness are not complex in the same way. So we envisage multiple answers: partly (but not exhaustively) recursion; partly, as Jackendoff et al (2006) put it: what is "[u]nique to language is a very large learned vocabulary consisting of long-term memory associations between meanings and structured pronunciations plus varied phrasal syntax". And partly PV, where crucially typology and acquisition are united and the central factor distinguishing language from all the rest is the antecedently available knowledge of the possible choices.

References

Ackema, P., P. Brandt, M. Schoorlemmer & F. Weerman (eds.) (2006), *Arguments and Agreement*. Oxford: Oxford University Press.

Baker, M. (1996) *The polysynthesis Parameter*. New York, OUP.

Bolhuis, J. (2005) "Development of behavior". In J. Bolhuis & L.-A. Giraldeau (eds) *The Behavior of Animals: Mechanisms, Function, and Evolution.* Oxford, Blackwell; pp. 119-145.

Bolhuis, J. & E. Macphail (2001) "A critique of the neuroecology of learning and memory". *Trends in Cognitive Sciences* 5:426-433.

Brenowitz, E., D. Margoliash & K. Nordeen (1997) „An introduction to birdsong and the avian song system". *Journal of Neurobiology* 33:495-500.

Brown, S., B. Merker & N. Wallin (2000) "An introduction to evolutionary musicology". In N. Wallin, B. Merker & S. Brown (eds) *The Origins of Music*. Cambridge MA, MIT Press; pp. 3-24.

Catchpole, C. (1991) "Songs". In M. Brooke & T. Birkhead (eds) *The Cambridge Encyclopaedia of Ornithology*; Cambridge, CUP; pp.283-291.

Chomsky, N. (1971) "On interpreting the world." *Cambridge Review* 92: 77–93.

Chomsky, N. (1975) *Reflections on Language*. New York, Pantheon.

Chomsky, N.(1980) *Rules and Representations.* Oxford, Blackwell

Chomsky, N. (1981a) *Lectures on Government and Binding.* Dordrecht, Foris.

Chomsky, N. (1981b) "Principles and parameters in syntactic theory." In N. Hornstein & D. Lightfoot (eds.), *Explanations in Linguistics.* London, Longman. 123–146.

Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin and Use*. New York, Praeger.

Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA, MIT Press.

Chomsky, N. (2002) *On Nature and Language*. Cambridge, CUP.

Chomsky, N. (2003) *Power and Terror: Post-9/11 Talks and Interviews*. New York, Seven Stories Press.

Chomsky, N. (2006) *Language and Mind*[3]. Cambridge, CUP.

Doupe, A. & P. Kuhl (1999) "Birdsong and speech: Common themes and mechanisms". *Annual Review of Neuroscience*, 22:567-631.

Doupe, A. & M. Solis (1999) "Song- and order-selective auditory responses emerge in neurons of the songbird anterior forebrain during vocal learning". In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.343-368.

Dresher, E. (1999) "Charting the learning path: Cues to parameter setting". *Linguistic Inquiry* 30:27-67.

Dresher, E. & J. Kaye (1990) "A computational learning model for metrical phonology". *Cognition* 34: 137-195.
Fanselow, G. (1993) "Instead of a preface: Some reflections on parameters"; in G. Fanselow (ed) *The Parametrization of Universal Grammar;* Amsterdam, John Benjamins; pp.vii-xvi.
Fodor, J. (1975) *The Language of Thought.* New York, Crowell.
Gardner,T., F. Naef & F. Nottebohm (2005) "Freedom and rules: The acquisition and reprogramming of a bird's learned song". *Science* 308:1046-1049.
Gentner, T., K. Fenn, D. Margoliash & H. Nusbaum (2006) "Recursive syntactic pattern learning by songbirds". *Nature* 440:1204-1207.
Haider, H. (1993) "Principled variability: Parametrization without parameter fixing". In G. Fanselow (ed) *The Parametrization of Universal Grammar;* Amsterdam, John Benjamins; pp.1-16.
Harris, P. (2000) *The Work of the Imagination.*
Hauser, M. (1997) "Minding the behaviour of deception". In A. Whiten & R. Byrne (eds) *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge, CUP; pp.112-143.
Hauser, M. (2006) *Moral Minds*. New York; HarperCollins.
Hauser, M., N. Chomsky & W. Tecumseh Fitch (2002) "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298:1569-1579 (November 22).
Hauser, M., F. Cushman, L. Young, R. K.-X. Jin & J. Mikhail (2007) "A dissociation between moral judgments and justifications". *Mind & Language* 22:1-21.
Hjelmslev, L. (1961) *Prolegomena to a Theory of Language.* Madison, University of Wisconsin Press.
Hockett, C. (1958) *A Course in Modern Linguistics.*
Hyams, N. (1987) "The theory of parameters and syntactic development". In T. Roeper & E. Williams (eds) *Parameter Setting*; Dordrecht, Reidel; pp.1-22.
Jackendoff, R. & F. Lerdahl (2006) "The capacity for music: What is it, and what's special about it?" *Cognition* 100:33-72.
Jackendoff, R., M. Liberman, G. Pullum & B. Scholz (2006) "Starling study: Recursion" http://linguistlist.org/issues/17/17-1528.html.
Jones, D. (2007) "The depths of disgust". *Nature* June 14, 2007; pp.768-771.
Joos, M. (ed.) (1957) *Readings in Linguistics.* Chicago, University of Chicago Press.
Jusczyk, P. (1997) *The Discovery of Spoken Language.* Cambridge MA, MIT Press.
Kayne, R.(2005) "Some notes on comparative syntax, with special reference to English and French". In G. Cinque & R. Kayne (eds) *The Oxford Handbook of Comparative Syntax,* Oxford, OUP; pp. 3-69.
Kuhl, P. (1999) "Speech, language and the brain: Innate preparations for learning". In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.419-450.
Lakoff, G. (1996) *Moral Politics*. Chicago; University of Chicago Press.
Liu, W., T. Gardner & F. Nottebohm (2004) "Juvenile zebra finches can use multiple strategies to learn the same song". *PNAS* 101:18177-18182.
Manzini, M.R. & L. Savoia (2007) *A Unification of Morphology and Syntax.* London, Routledge.
Marcus, G. (2006) "Startling starlings". *Nature* 440:1117-1118.
Marler, P.(1999) "On innateness: Are sparrow songs 'learned' or 'innate'?" In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.293-318.
Marler, P. (2000) "Origins of music and speech: Insights from animals". In N. Wallin, B. Merker & S. Brown (eds) *The Origins of Music*. Cambridge MA, MIT Press; pp. 31-48.
Mikhail, J. (2007) "Universal moral grammar: theory, evidence and the future". *Trends in Cognitive Sciences* 11:143-152.

Mundinger, P. (1999) "Genetics of canary song learning: Innate mechanisms and other neurobiological considerations". In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.369-389.
Muysken, P. (in press) *Functional Categories*. Cambridge, CUP.
Nettl, B. (2000) "An ethnomusicologist contemplates universals in musical sound and musical culture". In N. Wallin, B. Merker & S. Brown (eds) *The Origins of Music*. Cambridge MA, MIT Press; pp. 463-472.
Nettl, B. (2005) *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Urbana IL, University of Illinois Press.
Peretz, I. (2006) "The nature of music from a biological perspective." *Cognition* 100: 1-32.
Peretz, I. & M. Coltheart (2003) "Modularity of music processing". *Nature Neuroscience* 6:688-691.
Piattelli-Palmarini, M. (1989) "Evolution, selection and cognition: from learning to parameter setting in biology and in the study of language." *Cognition* 31: 1–44.
Pinker, S. (1984) *Language Learnability and Language Learning.* Cambridge MA Harvard University Press.
Postma, G. & J. Rooryck (2007) "Phase-recursion, restricted linguistic systems, and full language". MS Leiden.
Roberts, I. (1997) *Comparative Syntax*. London, Arnold.
Roberts, I. & A. Roussou (2003) *Syntactic Change: A Minimalist Approach to Grammaticalization.* Cambridge, CUP.
Safir, K. (1987) "Comments on Wexler & Manzini". In T. Roeper & E. Williams (eds) *Parameter Setting*; Dordrecht, Reidel; pp.77-89.
Saygin, A., F. Dick & E. Bates (2005) "An on-line task for contrasting auditory processing in the verbal and nonverbal domains and norms for younger and older adults". *Behavior Research Methods* 37:99-110.
Searcy, W. & S. Nowicki (1999) "Functions of song variation in song sparrows". In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.577-595.
Smith, N.V. (1973) *The Acquisition of Phonology: A Case Study.* Cambridge, CUP.
Smith, N.V. (1990) "Can pragmatic fix parameters?" In I. Roca (ed) *Logical Issues in Language Acquisition* Foris, pp. 277-289.
Smith, N.V. (1999) "The structure of noise". *Glot International* 4,5:9. (Reprinted in Smith, 2002).
Smith, N.V. (2002) *Language, Bananas and Bonobos: Linguistic Problems, Puzzles and Polemics.* Oxford, Blackwell.
Smith, N.V. (2004) *Chomsky: Ideas and Ideals* (2nd edition); Cambridge, CUP.
Smith, N.V. (2005) *Language, Frogs and Savants: More Linguistic Problems, Puzzles and Polemics.* Oxford, Blackwell.
Smith, N.V. (2007) "Unification and the unity of Chomsky's thought". *Chomskyan Studies* 2:1-21.
Smith, N.V. & A. Cormack (2002) "Parametric poverty", *Glot International* 6:285-287. Reprinted in Smith (2005).
Suthers, R. (1999) "The motor basis of vocal performance in songbirds". In M. Hauser & M. Konishi (eds) *The Design of Animal Communication*. Cambridge MA, MIT Press; pp.37-62.
Tomalin, M. (2007) "Reconsidering recursion in syntactic theory". *Lingua* 117:1784-1800.
Trehub, S. (2000) "Human processing predispositions and musical universals". In N. Wallin, B. Merker & S. Brown (eds) *The Origins of Music*. Cambridge MA, MIT Press; pp. 427-448.
Trehub, S. & E. Hannon (2006) "Infant music perception: Domain-general or domain-specific mechanisms?" *Cognition* 100:73-99.

Whaling, C. (2000) "What's behind a song? The neural basis of song learning in birds". In N. Wallin, B. Merker & S. Brown (eds) *The Origins of Music*. Cambridge MA, MIT Press; pp. 65-76.

Yip, M. (2003) "Casting doubt on the Onset-Rime distinction". *Lingua* 113:779-816.

Phonology

ATR allophones or undershoot in Kera?

MARY PEARCE

Abstract

Kera (a Chadic language) has 6 vowels, 3 of which have +/-ATR allophones. [+ATR] vowels appear in non-heads of feet and [-ATR] vowels in heads and elsewhere. This binary classification is sufficient until we examine the acoustic measurements of F1, F2 and duration in footed and non-footed syllables. These results suggest that the variation in quality relates to the duration of the vowel rather than directly to the foot structure. We will consider the evidence for claiming that there is a gradient relationship between the F1 value and the duration. The key data for this claim come from vowels in non-footed syllables at the right edges of phrases and vowel initial syllables. In non-footed syllables the duration of the vowel is longer than a non-head vowel, but shorter than a head vowel. The F1 value for these vowels is equally between the average head and non-head values. Gendrot and Adda-Decker (2006) have demonstrated similar patterns in other languages where a shorter duration means a more centralised vowel. Their results could lead us to suppose that the reason for this gradient is articulatory, due to the need of a certain amount of time for articulators to arrive at the target position, and that all languages may exhibit a similar phonetic pattern. A few counter-examples suggest that this pattern can be over-ridden by phonological factors. In the case of Kera we may well be seeing a process that began as a gradient phonetic change but which is now in the process of being phonologized. Therefore the use of the term 'allophone' correctly describes the phonology, but the phonetics also has a role to play in the quality of the vowel.

1 Introduction

Kera has been analysed in the literature (Ebert 1974, 1979, Pearce 2003) as having 6 vowels, 3 of which have +/-ATR allophones based on the position of the syllable in the iambic foot. [+ATR] vowels appear in non-heads of feet and [-ATR] vowels in heads and elsewhere. Up to now this binary classification has been generally accepted. However, a closer inspection using acoustic measurements of F1, F2 and duration reveals that the variation in quality may be due principally to duration rather than foot structure (although the foot structure affects duration). This would lead us to suppose that rather than a categorical distinction between the allophones associated with head and non-head syllables, we may have a gradient relationship between the F1 value and the duration. Both increase together until the target F1 value is reached, at which point a further increase in duration no longer affects the quality. The key data for this claim

come from vowels in non-footed syllables at the right edges of phrases and vowel initial syllables. In both of these cases, the duration of the vowel is longer than a non-head vowel, but shorter than a head vowel. The F1 value for these vowels is equally between the average head and non-head values. Neither of these cases fits neatly into a binary division of allophones.

A useful comparison can be made with French. Gendrot and Adda-Decker (2006) have measured F1, F2 and duration in corpuses from several languages including French, and conclude that in each language the F1 and F2 values appear to vary with duration in a gradient relationship, particularly in non-high vowels. They observe that the polygons made by the vowel space in an F1/F2 plot converge as the duration decreases towards a schwa like vowel. Kera shows a similar convergence, but towards a horizontal line rather than a point. For a full understanding of the Kera facts, we need to combine an undershoot account with a consideration of the effects of the metrical structure on duration and the contribution made by the rich vowel harmony system, which may be constraining the variation in F1 and F2.

This paper begins by looking at the case for allophones and the case for undershoot. We then move to consider other languages as mapped out by Gendrot and Adda-Decker, and we compare the Kera results with these languages. Finally, we will discuss whether the Kera facts are best considered as a categorical split between allophones or a gradient of qualities that vary with duration.

2 The case for allophones

Kera is typical of a number of Chadic languages in having a symmetrical system of vowels which can be paired into high and non-high vowels. Kera has 6 vowels, and the three high vowels do not differ much in quality regardless of the duration or position of the vowel in the foot. But the 3 non-high –ATR vowels have +ATR allophones in non-heads of iambic feet[1]. This binary analysis of these vowels gives us [+ATR] vowels in non-heads of feet and [-ATR] vowels in heads and elsewhere[2].

[1] More information on iambicity in Kera is available in Pearce (2006).

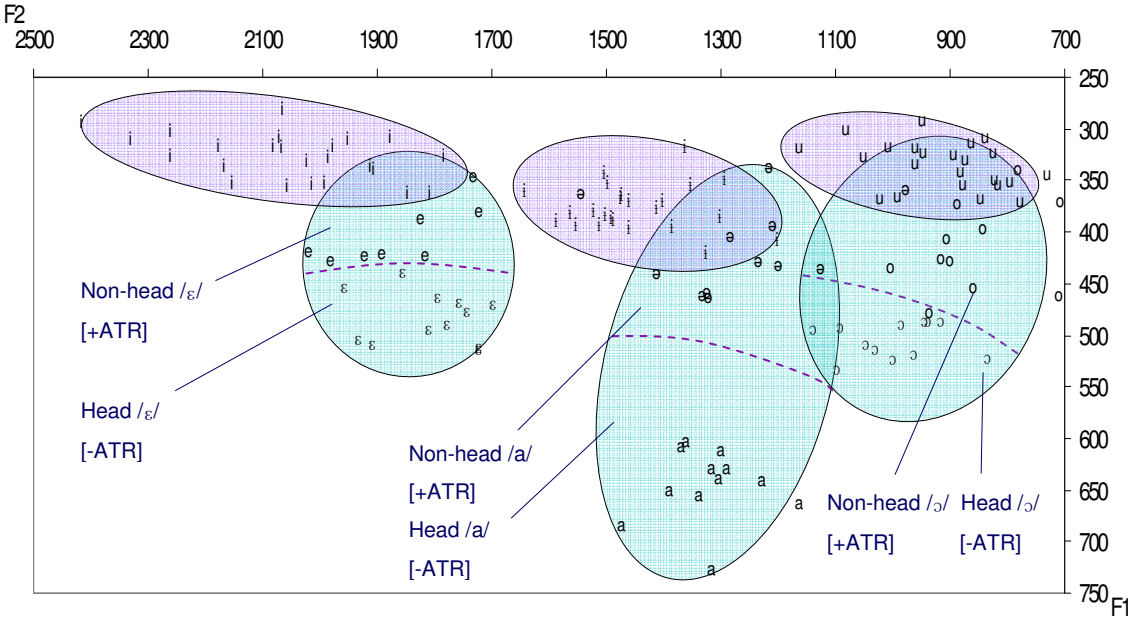
[2] The use of [ATR] rather than [tense/lax] is not meant to be significant. Either could be used in Kera. It is important however that this feature is differentiated from the high/non-high distinction which is contrastive, playing a role in height harmony. Casali (2001) supports the use of the [ATR] feature in this case even if the action of the tongue root is not proven, but it should be noted that in most African languages when [ATR] is involved, it is used to mark contrasts in an ATR harmony system.

(1) *Allophones in Kera vowels*

Phonemes:	/i/	/ɨ/	/u/	/ɛ/	/a/	/ɔ/
Head (-ATR)	[i]	[ɨ]	[u]	[ɛ]	[a]	[ɔ]
Non-head (+ATR)				[e]	[ə]	[o]

In (2), the three non-high vowels are split by a dotted line separating the allophones.

(2) *Means of footed Kera vowels for 12 speakers*



Because the ə/a alternation involves two allophones that are phonetically much further apart than the others, Ebert (1979) treated this alternation as a special case of a process of dissimilation which changed every other /a/ into [ə]. This alternating pattern was actually the result of the metrical structure which prefers disyllabic feet where non-heads and heads will be alternating. Unfortunately the apparent special case for the low vowel /a/ has led several linguists including Buckley (1997), Suzuki (1998), de Lacy (2004) and Archangeli and Pulleyblank (2007) to give this Kera example as support for theories of dissimilation processes involving low vowels, although all of these authors cite other languages as well as Kera. But this alternation behaves exactly like the alternation for o/ɔ and e/ɛ, and the three pairs should be treated in the same way.

Examples of the allophones can be seen in the following words where feet are indicated by parentheses and head vowels are underlined. In these examples, there is

total vowel harmony, so any change in quality is due to the choice of allophone for the position within the foot.

(3) *Allophones chosen by the position in the foot (head vowels underlined)*

(gədaa)(yaw)	'pots'
(dak)(tələaw)	'bird'
(sɛɛ)(renɛn)	'rescued me'
(gɔl)(donɔn)	'searched for me'

Kera is not alone in having such alternations between stressed and unstressed syllables. Among Chadic languages, Pearce (2007) notes that Hausa (Newman 2000), Sokoro, Goemai, Bade and Ngizim show a similar pattern. Beyond Chadic, there are languages such as Catalan (Harrison 1997) which has a 7 vowel system /i, e, ɛ, a, ɔ, o, u/ which reduces to 5 vowels. [ɛ] and [ɔ] ([-ATR]) appear only in stressed syllables.

Returning to Kera, as long as we consider only the vowels contained within feet, this analysis appears to be perfectly adequate and the case for a binary choice between allophones seems solid. However, if we look at vowels that are either unfooted or epenthetic, the case becomes less clear.

3 The case for gradiency

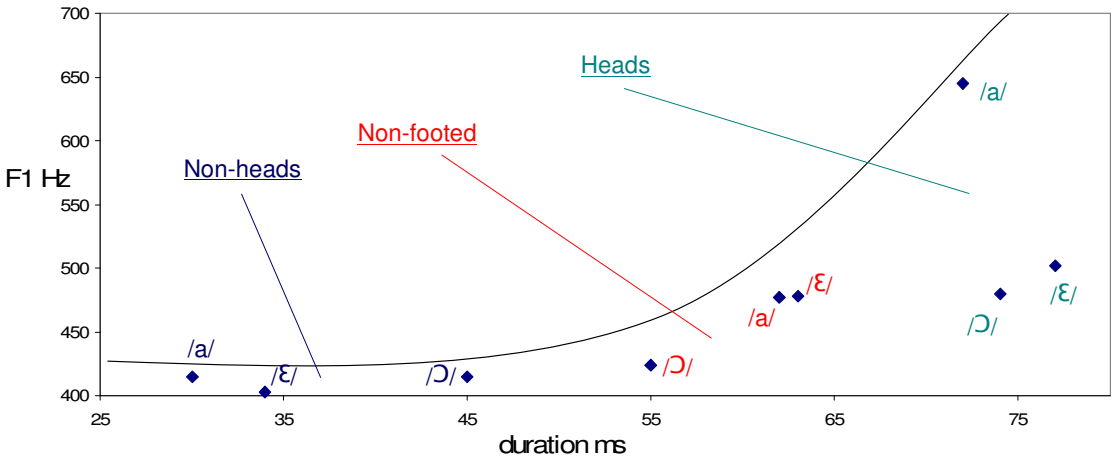
Our key evidence for gradiency is that non-footed vowels have a duration and F1 value between head and non-head vowels. They do not fit neatly into either of the two categories. Non-footed syllables are found at the right edges of phrases and in vowel initial words where the syllable is made up of just the vowel. These vowels do not behave like the vowels in footed syllables. The table in (4) demonstrates this, also including epenthetic vowels which equally do not fit well in a categorical system. The quality that we previously called +/-ATR does not seem to vary according to inclusion in a foot or headedness. It does however appear to relate to the duration of the vowel. The shortest vowels are closer to a [+ATR] category and the longest vowels are closer to a [-ATR] category. There is the temptation to argue that this is still a categorical distinction between short and long vowels, but all the vowels in (4) are phonologically short. Kera has a phonologically long vowel which has not been mentioned. This vowel typically has a duration of around 110 ms with a quality in keeping with the feature [+ATR]. But the difference in quality in (4) appears to occur at around 50 ms which is not the same place as the phonological boundary between short and long.

(4) *The correlation between quality, duration and position*

	Head V	Not a Head V		
	Head V (tar̲) 'run'	Non-head V (cə̲wa:) 'sun'	Epenthetic V (gɔl)do̲(to̲nɔn)	Non-footed V (baa)ŋa̲ 'elephant'
Footed	✓	✓	?	
Duration	70ms	30ms	30ms	50ms
+ATR		✓	✓	?

So instead of a categorical distinction between the allophones associated with head and non-head syllables, we might find that positing a gradient relationship will suit us better. As the phonological feature [ATR] is categorical, it does not lend itself to being treated as a gradient. Instead, we will consider the gradient in terms of the F1 value. In (5), the non-high vowels are plotted on a graph with F1 against duration. The mean values for vowels in each category are shown. Non-heads are short and the F1 value is also low. Heads on the other hand have a much greater duration and a much higher F1. The non-footed vowels have a duration between the footed vowels already considered, and likewise the F1 value is between that of the others.

(5) *A gradient change in duration and F1 (diagram in Pearce forthcoming)*



It seems that duration and the F1 value increase together until the target F1 value is reached, at which point a further increase in duration no longer affects the quality. The curved line in (5) is there as an indication of the gradient nature of the curve. These data are not enough to give any precise equations for this line, but the implication is that the vowel has to be of a certain duration before the F1 target can be reached and

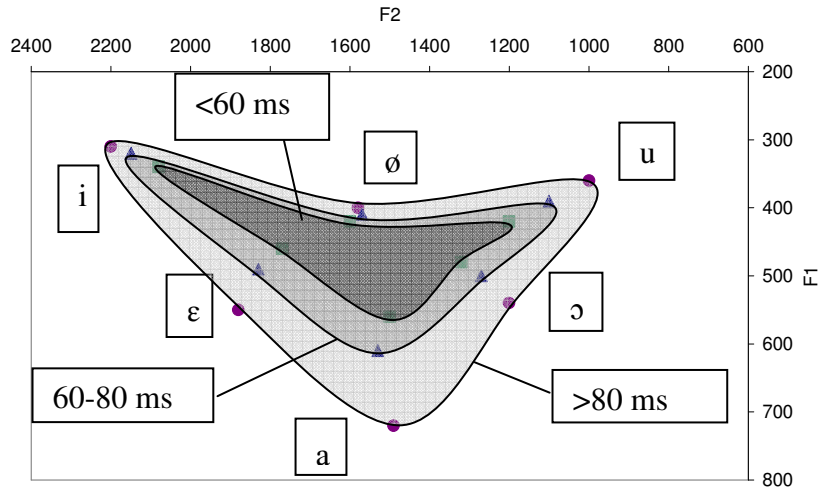
that if it is shorter than this, the vowel will have a reduced quality. The reason for this could well be that the articulators do not have enough time to reach the target.

So the analysis of a gradient curve rather than an allophonic split now seems the better option. We now consider other languages to see if there is evidence for a similar gradient curve there.

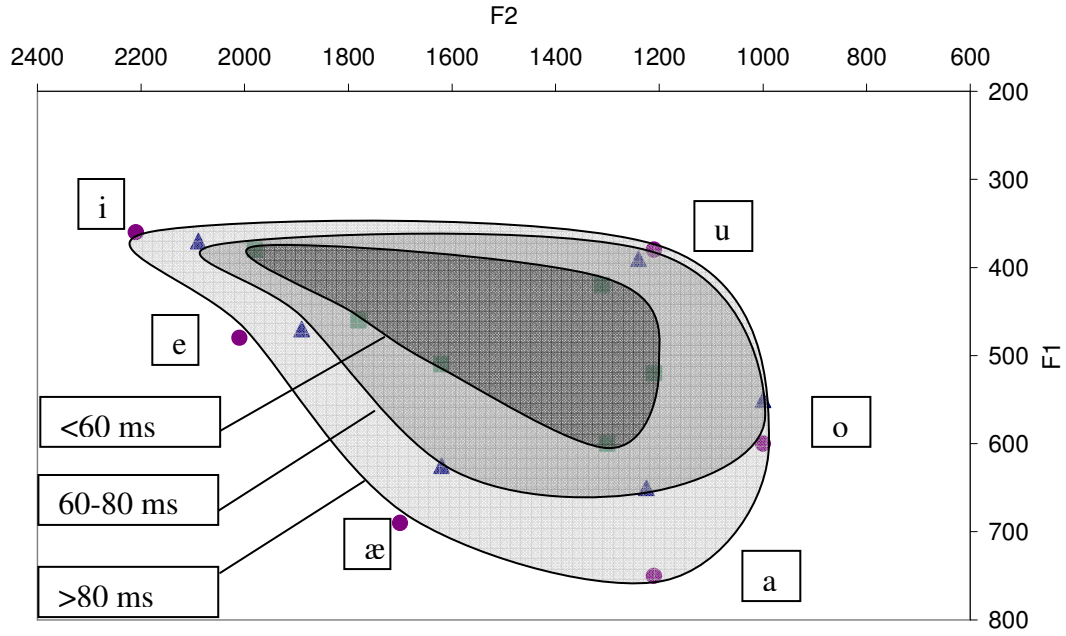
4 Convergence in other languages

For this section, we will make use of the work of Gendrot and Adda-Decker (2005, 2006). They have studied eight languages, using a large corpus for each, and they have found similar results in each language. In the diagrams below, I include only the vowels which bear some correspondence with the Kera vowels, but the results with all of the vowels in each language are available in the original papers. My purpose here is to demonstrate the trend that appears to be present in all of the languages studied. For each language, a plot of F1 and F2 is made with different polygons according to the duration of the vowel. The results show clearly that F1 and F2 values vary with duration in a gradient relationship, particularly in non-high vowels. The polygons made by the vowel space in an F1/F2 plot converge as the duration decreases. Gendrot and Adda-Decker suggest that the explanation for the convergence effect might be partly articulatory. This view would be supported if all languages converge in the same way. This is true in their data. Four language plots are given here.

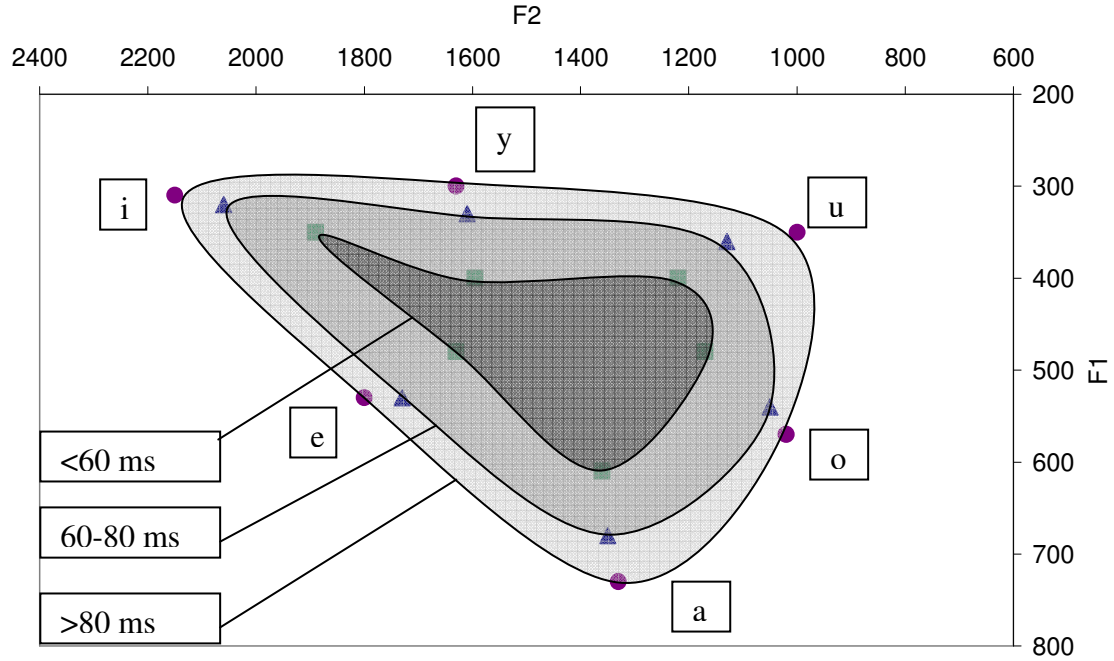
(6) *Measured mean average values of F1 and F2 for French vowels according to duration, data from Gendrot and Adda-Decker (2006), selected vowels only*



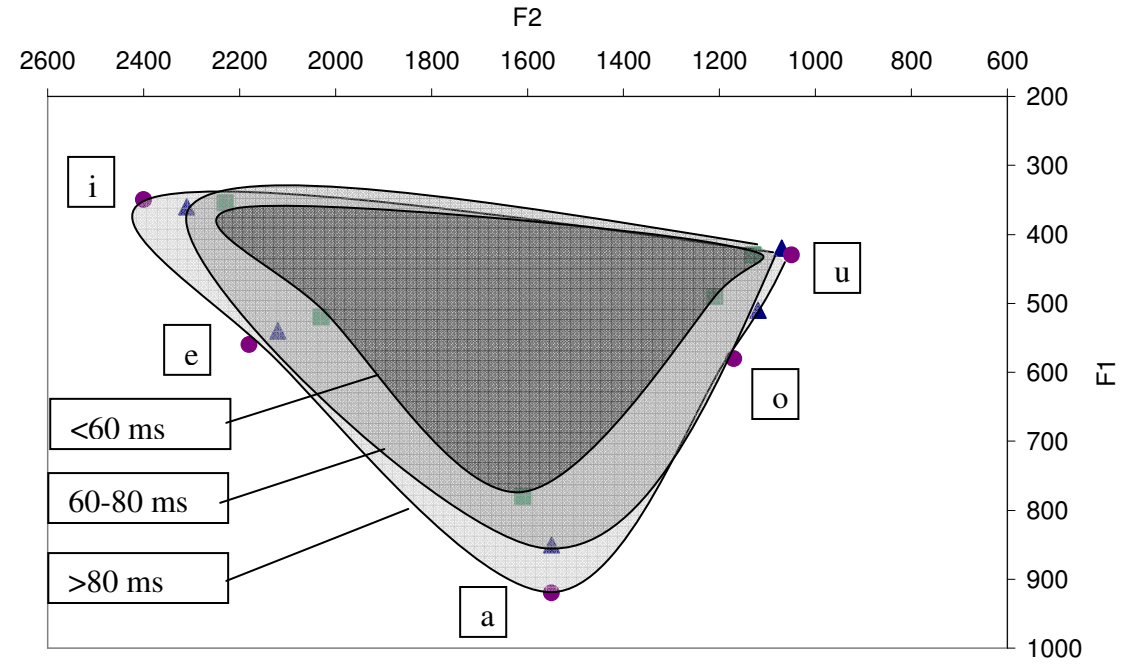
(7) *Measured mean average values of F1 and F2 for English vowels according to duration, data from Gendrot and Adda-Decker (2006), selected vowels only*



(8) *Measured mean average values of F1 and F2 for German vowels according to duration, data from Gendrot and Adda-Decker (2006), selected vowels only*



(9) *Measured mean average values of F1 and F2 for Mandarin vowels according to duration, data from Gendrot and Adda-Decker (2006)*

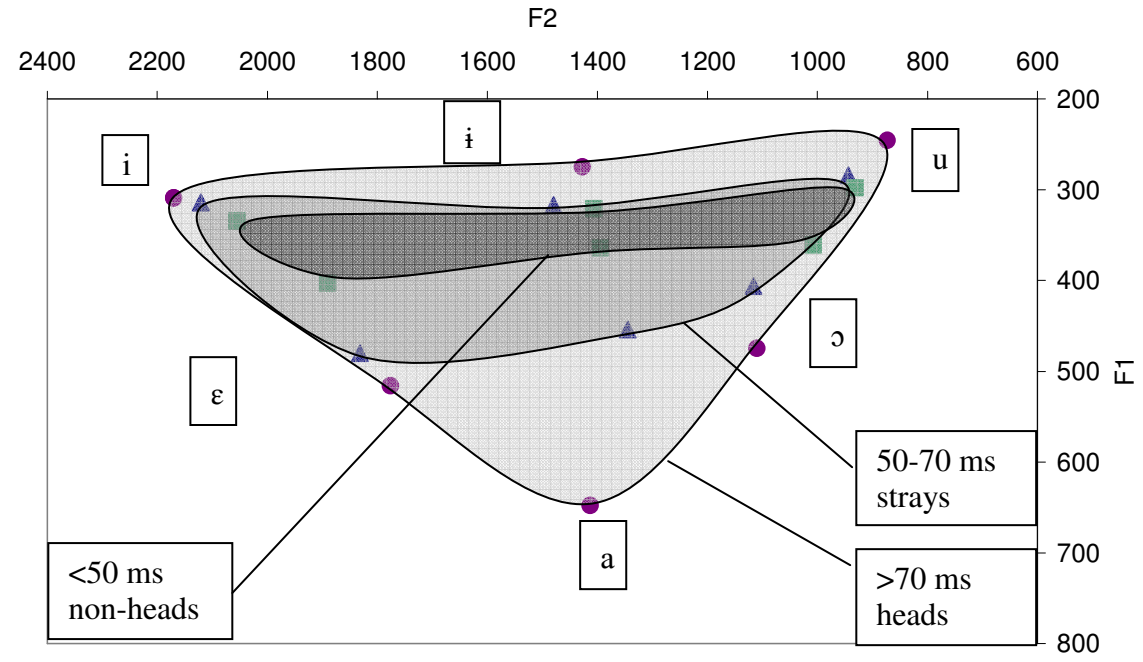


The other languages measured by Gendrot and Adda-Decker giving similar results were: Arabic, Spanish, Italian, and Portuguese. The question arises whether it is possible to claim the same kind of vowel reduction for all languages. At first glance there appear to be counter-examples in the work of Archambault and Maneva (1996) and Gussenhoven (2004). The first of these studies considers devoicing in post-vocalic Canadian-French obstruents. Other cues for devoicing were measured, including the vowel duration and quality. They make the observation that lax vowels, which appear before voiceless obstruents are shorter, while tense vowels, which appear before voiced obstruents are longer. This would seem to be the inverse of the diagrams above. However, as the sample included both high and non-high vowels, and as separate results are not given, we cannot compare this study with that of Gendrot and Adda-Decker. This study also refers only to vowels before obstruents where the voicing is known to affect a number of cues. So the results could be different if a more comprehensive study of vowels in all positions were made. A similar comment can be made for the Gussenhoven study on Limburgian dialects of Dutch and English coda obstruents. The results seem to be the inverse of what we are expecting, but this study is again looking at specific cases where a phonological contrast is being cued by the duration and quality of the vowel. Gussenhoven notes that high vowels are perceived as longer than non-high vowels, possibly compensating for the inverse relationship in

production. Maddieson (1997) and Catford (1977) claim that high vowels tend to be shorter than non-high vowels in production because the articulators are already in position making it easier to move on to the next consonant. These results show that the relationship between duration and quality can be affected by a number of factors. Clearly, we certainly cannot assume that every language behaves like the eight tested by Gendrot and Adda-Decker, and it seems that the trends can be reversed where phonological contrast plays a role, but their work merits further research, and a useful development would be to look at more non-Indo-European languages.

We now turn to consider what the plot would be for Kera. It is not possible to use the same size of corpus, but nevertheless, the following plot gives significant differences between the three polygons.

(10) *Measured mean average values of F1 and F2 for Kera vowels according to duration*



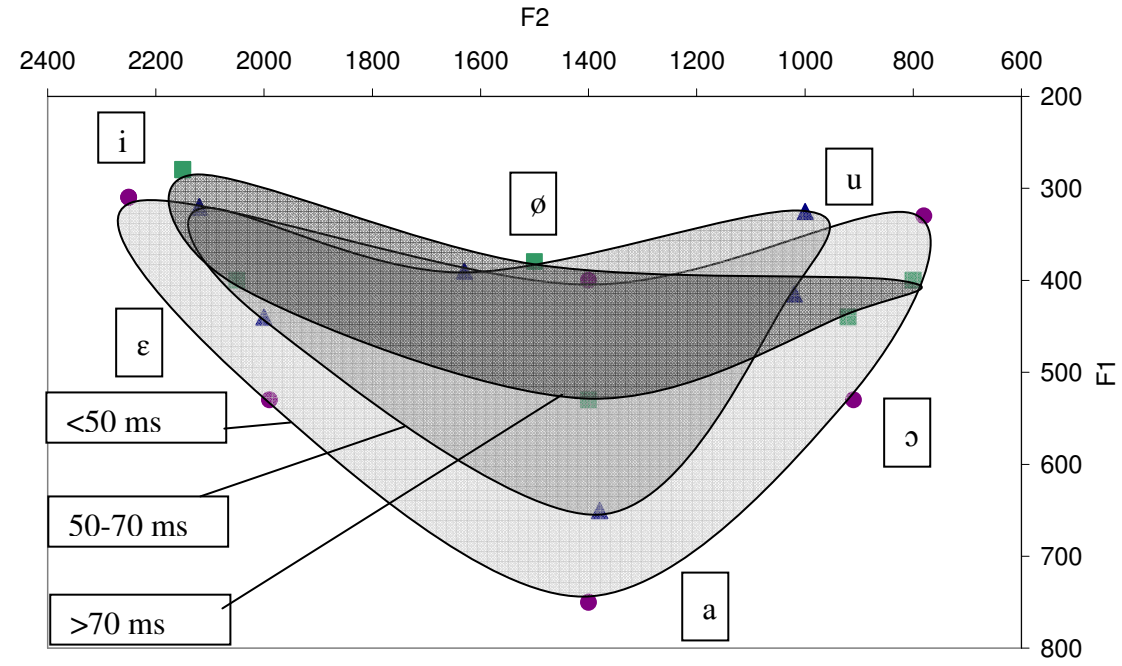
Kera shows a similar convergence to that found in the other languages, but only in one dimension. The convergence is towards a horizontal line rather than a point. The variation is most striking for the /a/ vowel. As with French, high vowels are relatively unaffected by duration. But unlike French, the F2 value is also unaffected by duration. The Kera result makes the articulatory explanation less appealing, but it is hard to find another explanation. A categorical phonological explanation does not appear to fit the

facts as there does seem to be a gradient. But the question remains open as to why only one dimension changes. Unlike the eight languages measured in the study of Gendrot and Adda-Decker, Kera has vowel harmony. It is possible that this fact has a role to play in the different pattern, but as Kera has fronting, rounding, height and total harmony over different domains, it is hard to see why some types of harmony affect this result while others don't. As the domain for height harmony is larger than the domain for other types of harmony, it is possible that some loss of height is tolerated without loss of overall information because other vowels in the word indicate whether the vowels are phonologically high or not. The only way to test this would be to measure several other languages with vowel harmony to examine if similar patterns emerge. Without this information, we have to stay tentative about why Kera differs from other languages. One such language would be Pasiego Spanish, with reduction and height harmony (McCarthy 1984, Harris and Lindsey 1995, Penny 2000, Walker 2005) At this point, there is little information available from other African languages, but one which appears to behave in a similar manner to Kera is Rangi (Stegen 2000). The number of vowels in Rangi is in dispute, ranging from 5 to 9, but focusing on the non-high front and rounded vowels, the move from a typical [+ATR] to a typical [–ATR] position appears to be gradient (both in terms of quality and how far the quality 'spreads'), suggesting that the differences may be caused by undershoot or some other phonetic cause rather than from an allophonic binary split.

5 Categorical or gradience, or both?

Before returning to our main question as to the phonological nature of the non-high vowels, we will consider one more plot of F1/F2 vowels. This time, the speaker is Kera, but speaking French. As this is the plot of only one speaker, we cannot base our conclusions mainly on this graph, but the result is interesting nonetheless, as there appears to be a combination of the two plots for French (6) and Kera (10). When a Kera man speaks French, it appears that the short vowels are like Kera, whereas the middle range and longer vowels show similar patterns in phonetic reduction to French. So this plot could lead us to conclude that for this speaker at least, the non-head vowels may be phonologically distinct from other vowels, but that there is also a phonetic reduction occurring at the same time. This applies to his French, but is less clear when he is speaking Kera.

(11) *Measured mean average values of F1 and F2 for the French vowels of one Kera speaker according to duration, selected vowels only*



A number of languages are known to have both gradient and categorical effects within the same vowel system, for example Brazilian Portuguese, Bulgarian and Russian (Barnes 2007). A profitable next step in this research would be to measure the vowels in these languages under similar conditions to the measurements referred to above.

5 Conclusion

We have considered arguments for both allophony based on weight and undershoot based on duration. Barnes (2006) and Crosswhite (2001) suggest that there are two common situations in languages. Either there is prominence-reducing vowel reduction, in which case the lack of stress means a reduction in quality in a desire to avoid effortful articulations, or there is vowel undershoot in which case there is insufficient time to produce the quality in non-high vowels. The first of these options tends towards being categorical and the second gradient. In the Kera case discussed here, undershoot seems more likely because of the gradient nature of the results and because only non-high vowels are affected.

We are still left with the question as to whether these 'allophones' should be dealt with by phonology or phonetics. It is possible that the answer is that both processes are

involved. It may be that the variation had a phonetic and gradient origin, based on the ability of the articulators to reach their target in time, but then synchronically, there could be a process of phonologization as subsequent generations observe that non-head vowels always have the [+ATR] allophone. The vowel harmony facts could also be playing a role in the exact positioning of the target for each vowel.

To test these ideas, there is a need for an investigation into other vowel harmony languages, and also languages where an allophonic variation of this kind has been attested. What is clear is that the Kera vowel system is not as clear cut as it seems at first, and it may be that what we are seeing here is a language in process of change where a process with a phonetic explanation is being phonologized. If so, it certainly merits further study. There is also potential in enlarging the number of languages that have been measured in the detail required to produce the F1/F2/duration plots, particularly for non-Indo-European languages.

References

Archangeli, Diana & Douglas Pulleyblank. 2007. Harmony. In (Ed.) Paul de Lacy. The Cambridge Handbook of Phonology, Cambridge. CUP.

Archambault, D. and B. Maneva. 1996. Devoicing in post-vocalic Canadian-French Obstruents. In ISCLP 1996. 1533-1536.

Barnes, Jonathan. 2006. Strength and Weakness at the Interface: Positional Neutralization at the Interface. Mouton de Gruyter.

Barnes, Jonathan. 2007. Ms. Draft submitted. Phonetics and Phonology in Russian unstressed vowel reduction: A study in hyperarticulation. http://www.bu.edu/linguistics/UG/barnes/

Buckley, Eugene. 1997. Against vowel length in Tigrina. Studies in African Linguistics 26.1: 63-102.

Casali, Roderic. 2001. Vowel Systems in African Languages. Paper given at SIL Central Africa Linguistics Consultants Seminar, April 2001.

Catford, J.C. 1997. Fundamental Problems in Phonetics. Edinburgh University Press.

Crosswhite, Katherine. 2001. Vowel reduction in Optimality Theory: access, mobility, and resilience among urban Latino youth. New York. Routledge.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. Phonology 21: 123-199.

Ebert, Karen. 1974. Partial vowel harmony in Kera. Studies in African Linguitics. Supplement 5: 75-80.

Ebert, Karen. 1979. Sprache und Tradition der Kera (Tschad). Teil III. Grammatik. Berlin. Reimer.

Gendrot, Cédric and Martine Adda-Decker. 2005. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. Paper given at Interspeech Sept. 4-8, 2005, Lisbon.

Gendrot Cédric and Martine Adda-Decker. 2006. Is there a universal impact of duration on formant frequency values of oral vowels? : An automated analysis of speech from eight languages. Poster presented at LabPhon 10, Paris, June 2006.

Gussenhoven, Carlos. 2004. A vowel height split explained: compensatory listening and speaker control. Paper presented at 9th conference on laboratory phonetics, June 24-26, 2004. Urbana Champaign, Illinois.

Harris, John and Geoff Lindsey. 1995. The elements of phonological representation. In (Eds.) Durand, Jacques and Francis Katamba, Frontiers of phonology: atoms, structures, derivations. 34-79. Harlow, Essex. Longman.

Harrison, Phil. 1997. The relative complexity of Catalan vowels and their perceptual correlates. UCL Working Papers in Linguistics 9. University College London.

Maddieson, Ian. 1997. Phonetic universals. In (Eds.) W.J.Hardcastle and J. Laver, The Handbook of Phonetic Sciences. Oxford. Blackwell. 619-639.

McCarthy, John. 1984. Theoretical consequences of Montañes vowel harmony. Linguistic Inquiry. 15. 291-318.

Newman, Paul. 2000. The Hausa Language: An Encyclopedic Reference Grammar. New Haven and London. Yale University Press.

Pearce, Mary. 2003. Vowel Harmony in Kera (Chadic). MA dissertation at University College London. [www.sil.org/africa/cameroon/index.html]

Pearce, Mary. 2006a. The Interaction between metrical structure and tone in Kera. In (Eds.) Bert Remijsen and Vincent J. van Heuven. Special Issue of Phonology: Between Stress and Tone. Phonology 23.2: 259-286

Pearce, Mary. 2007. Iambicity in Kera. Selected Proceedings of the 37th Annual Conference on African Linguistics, Cascadilla Press.

Pearce, Mary. Forthcoming. The interaction of tone with voicing and foot structure: Evidence from Kera phonetics and phonology. PhD dissertation at University College London.

Ralph, Penny. 2000. Variation and change in Spanish. CUP. Cambridge.

Stegen, Oliver. 2000. Rangi Vowel System: Five or Seven? Ms. Paper presented at the 30th Colloquium of African Languages and Linguistics at Leiden University, August 28-30, 2000.

Suzuki, Keiichiro. 1998. A typological investigation of dissimilation. PhD dissertation at the University of Arizona.

Walker, Rachel. 2005. Weak Triggers in Vowel Harmony. NLLT. 23.4.

*Consonant clusters in the acquisition of Greek: the beginning of the word**

EIRINI SANOUDAKI

Abstract

In this paper, I study the production of consonant clusters by Greek children and examine the consequences of the acquisition data for phonological theory, with particular emphasis on the word initial position. Using a non-word repetition test, I tested the order of acquisition of word initial and word medial sT, TT and TR clusters in 59 children. The results provide evidence against any analysis that assigns identical syllabic status to word initial sT and word initial TT, such as models of extrasyllabicity, and lend support to an alternative analysis of the beginning of the word, based on Lowenstamm's (1999) initial ON hypothesis.

1 Introduction

In the study of phonology, considerable insight can be gained from first language acquisition data. The order in which children acquire various phonological structures is of particular interest, as acquisition has been shown to be influenced by markedness. Specifically, various studies provide evidence that children master the production of unmarked sounds or structures before marked ones (Demuth 1996; Jakobson 1968; Stites, Demuth, & Kirk 2004).

An area that would greatly benefit from acquisition data is that of word initial consonant clusters. Though consonant clusters have been extensively studied by acquisitionists (Barlow 1997; Demuth & Kehoe 2006; Freitas 2003; Jongstra 2003; Kirk & Demuth 2005; Lleó & Prinz 1996; Pan 2005; Pan & Snyder 2004; Vanderweide 2005) the focus of the research on the word initial position has been on obstruent-sonorant clusters (TR) and s+consonant (sC) or s+obstruent (sT) clusters. Other word initial clusters, such as obstruent-obstruent clusters (TT) have been largely ignored. These clusters (for example *ft, xt,* which are attested in Greek) are problematic for phonological theory as they do not respect the regular

*I would like to express my deepest gratitude to the children who made this study possible, and to thank the teachers of the 2nd State Nursery, 2nd Municipal Nursery and Workers' Guild Nursery in Rethymno, and the nursery 'Baby Club' in Iraklio, for their cooperation. I would also like to thank my supervisors, John Harris and Neil Smith, as well as Rex Galbraith, Cecilia Kirk, Angelos Lengeris and Chloe Marshall for their help. This research was supported by a Grant from the Greek State Scholarship Foundation (IKY).

rising-sonority pattern associated with the beginning of a syllable, a fact which has led phonologists to the assumption that these clusters are extrasyllabic. Word initial TT is generally assumed to share the same structure as sT clusters, which are problematic not only in phonological theory, but also in the study of language acquisition. Consequently, studying the acquisition of TT clusters alongside sT clusters could help us understand the behaviour of sT clusters. More generally, studying the acquisition of different clusters, for example word initial clusters alongside their word medial counterparts, can be a lot more insightful than studying the acquisition of a cluster type in isolation.

Following this reasoning, in order to examine the phonology of the clusters in question, I test the production of consonant clusters by children acquiring Greek as their first language.

The paper proceeds as follows: Section 2 contains a short discussion on the word initial clusters in question. Section 3 deals with the data collection and general results, and in section 4 I proceed to the analysis; in section 4.1 some problems of the extrasyllabic theory are presented, and in section 4.2 I introduce an alternative proposal for the analysis of the data based on Lowenstam's initial ON hypothesis. A short conclusion follows.

2 Word initial consonant clusters

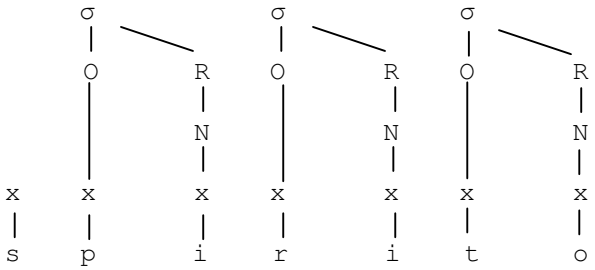
2.1 Word initial extrasyllabicity

Word initial sT clusters do not respect the Sonority Sequencing Generalisation (SSG, Clements (1990)) according to which sonority increases towards the syllable peak and decreases towards the edges. Initial sT breaks this generalisation, since the second member of the cluster has a lower (in the case of stops) or an equal (in the case of fricatives) sonority value when compared to the first member (*s*). This is the opposite of what the SSG dictates for onsets, namely that the second member of the cluster should be of higher sonority.

Faced with this inconsistency, several researchers have opted for a syllabification algorithm that leaves the *s* outside the onset: the *s* is extrasyllabic[1] (e.g. Halle & Vergnaud (1980), Levin (1985), Steriade (1982)). An example of such a structure is given in (1) below.

[1] Other attempts include analysis of sT as a contour-complex segment (Selkirk (1982), Weijer (1993) cf. Scobbie (1997)) and the abandoning of the SSG as a universal principle (Cairns 1988).

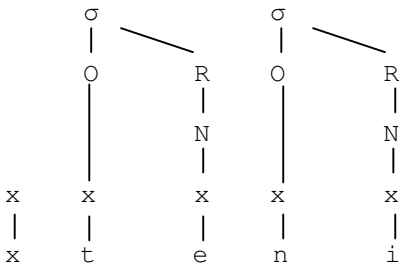
(1) sT extrasyllabicity: Italian *sp'irito* 'spirit'



Later in derivation the *s* may be linked to a constituent via some kind of adjunction rule. The desired effect is thus attained: at the first stage, the SSG is not violated, since the *s* is not linked to the onset, while at the same time eventual integration to the syllabic structure is achieved.

The same extrasyllabic structure has been proposed for word initial TT clusters (e.g. Rubach & Booij (1990), Steriade (1982)).

(2) TT extrasyllabicity: Greek *xt'eni* 'comb'



These clusters, too, violate the SSG, and an identical phonological analysis for both sT and TT, such as extrasyllabicity, seems to be a sensible move.

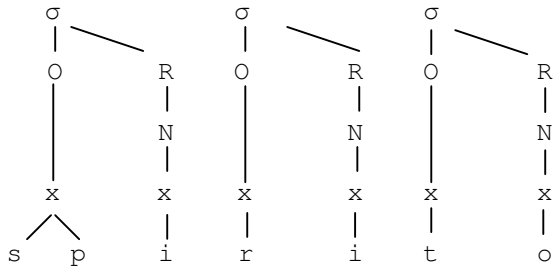
2.2 Order of acquisition

In first language acquisition, sT extrasyllabicity shows unusual behaviour: sT can be acquired after, but also before TR. Several studies have shown that children start producing initial sT clusters after TR clusters (e.g. Chin (1996), Smith (1973)). However, other studies (e.g. Barlow (1997), Gierut (1999)) found that some children produce initial sT clusters first.

The variation in the order of #sT-#TR acquisition has long puzzled researchers and there have been a number of proposals developed in order to tackle this problem. For example, it has been suggested that the explanation for these data lies in the possibility that some children acquire branching onset structures (TR) before extrasyllabicity, while others acquire extrasyllabic structures first (Fikkert 1994). This assumes that extrasyllabicity and branching onsets (TR) are different, but

equally marked structures, and the order of acquisition is therefore subject to variation. A different suggestion holds that, in acquisition, extrasyllabic clusters (and more generally consonantal sequences) may be structured like affricates (Barlow (1997), Lleó & Prinz (1997)). The relevant structure is shown below.

(3) sT as an affricate: Italian *sp'irito* 'spirit'



As seen in (3), sT clusters are represented as complex segments with a single timing slot. According to this approach, if a child does not structure sT like an affricate, s/he will acquire it after TR (i.e. as extrasyllabic, and therefore more marked). If, on the other hand, in a developing grammar, sT is structured like an affricate, it will be acquired before TR (on the assumption that complex segments are less marked than complex onsets). This optionality of structure, it is argued, can account for the variation in #sT-#TR acquisition. However, this approach does not seem to be particularly insightful, as it does not define what circumstances regulate whether a consonantal sequence will be structured as an affricate or as a cluster[2].

The acquisition of word initial TT clusters has not received much attention. Even though researchers have shown an increasing interest in the acquisition of Greek phonology, the majority of the studies are concerned with the acquisition of stress patterns (Kappa 2002b; Tzakosta 2003, 2004) or of different sounds (Kappa 2000; Nicolaidis et al. 2004; Tzakosta 2001b) in specific positions (e.g. word final consonants: Kappa, (2001). The studies that deal with consonant clusters are mostly concerned with what consonant children preserve when they simplify consonant clusters (Kappa 2002a; Tzakosta 2001a). While these studies provide some data on children's production of TR and TT clusters, they are typically isolated examples. There is some evidence regarding the acquisition of TR clusters as compared to e.g. other word medial clusters, but word initial TT clusters have largely been ignored (see, e.g., Kula & Tzakosta (2002)). Even when cluster types are examined separately, no distinction is made between word initial and word medial clusters (e.g. Papadopoulou 2000).

[2] See also Scobbie (1997) for a conceptually motivated criticism of the contour segment analysis.

From a theoretical point of view, an analysis that assumes extrasyllabicity of initial clusters of non-rising sonority will predict the same variation in TT versus TR acquisition as in sT versus TR acquisition. If TT is extrasyllabic like sT, and sT is acquired before or after TR, then TT is expected to be acquired before or after TR. Both analyses of sT versus TT acquisition outlined above (that extrasyllabicity can be acquired before or after TR, or that extrasyllabic clusters can be structured as complex segments in acquisition) would make the same prediction in this case. Moreover, word initial sT and TT are expected to be acquired at roughly the same time, under the assumption that they share the same (extrasyllabic) structure. Furthermore, a comparison of these clusters to their word medial counterparts can further test the theory. Though it is not clear from the theory of extrasyllabicity whether we should expect to find a difference between word initial and word medial sT, and if so, in what direction, whatever the relationship between initial and medial sT (i.e. whichever is acquired first), the same relationship should hold between initial and medial TT. Word initial and word medial TR, on the other hand, are expected to show no difference, since both positions involve the same structure (namely complex onset).

3 The experiment

3.1 Goal

The purpose of this experiment is to test the role of markedness on Greek children's production of consonant clusters. Different clusters in different positions will be tested and the results compared. The cluster pairs we are interested in are the following:

(4) #sT versus #TR
#TT versus #TR
#sT versus #TT
#sT versus -TT
#TR versus -TR
#TT versus -TT

word initial
- word medial

3.2 Methods and materials

3.2.1 Subjects. Fifty-nine monolingual Greek children were tested (21 boys and 38 girls). Nine more children were excluded from the study, since they refused to cooperate or did not manage to complete the task. The age range was from 2;03 to 5;00, mean age 3;08. The experiments took place in four different nurseries in

Crete (three in Rethymno and one in Iraklio) and, in the case of one child only, in a relative's house.

The children were selected according to linguistic and general developmental criteria. The developmental criteria required normal development, i.e. no background of cognitive, behavioural, hearing or physical impairment. I asked the nursery staff whether the child had any relevant problems. All fifty-nine children participating in this study were reported by staff as being healthy. The linguistic criteria required that i) the child's native language be Greek, ii) the child be raised in a monolingual environment iii) the child have a normal linguistic development iv) the child be able to produce at least some consonant clusters. Finally, the children had to be willing to participate in a non-word repetition task.

3.2.2 Methodology. A non word repetition task was used. Children were asked to repeat novel, made-up words that had the desired structures. The task was chosen for its effectiveness in producing a large amount of relevant data, compared to spontaneous production. Also, novel words allowed me to control for familiarity effects, which would be present in imitation tasks containing existing words. Furthermore, using nonsense words allowed me to control the phonological environment of the clusters across conditions.

Non-word repetition has been used mainly as a test of working memory (e.g. Gathercole (1995), Gathercole et al. (1994), Laws (1998), cf. van der Lely and Howard (1993)) and has been proposed as a screening measure for language impairment (e.g. Dollaghan and Campbell (1998), Weismer et al (2000)), but it is also used in studies examining young children's acquisition of phonology (e.g. Kirk and Demuth (2006), Zamuner and Gerken (1998), Zamuner, Gerken, and Hammond (2004)). Kirk and Demuth (2006), for example, used a non-word repetition task in order to examine English children's production of coda consonants. Although it has been suggested that imitative speech may not tap into the child's phonological system in the same way as spontaneous speech, there are results showing that the patterns found in imitation tasks are similar to those found in spontaneous speech. For instance, a production study by Kehoe and Stoel-Gammon (2001) showed no difference in the accuracy of children's imitated and spontaneous productions.

Extra care was taken to ensure the naturalness of the task. Firstly, the words were paired with pictures of novel animals, so that the words would have a referent; I thus made sure that the task is a linguistic one (rather than a general non-linguistic sound-production task). Secondly, the children did not hear the stimuli from a recording, but from a person (the experimenter), something that is more likely to occur in everyday life. Later evaluation of the spoken stimuli words by the experimenter showed consistent use of appropriate stress and segmental content. Thirdly, the task was not presented to the children as a request to repeat words, but

as a game in which they were taking active part. The game was designed in a way that reflected real life interactions (see procedure, section 3.2.4).

I have good reasons to believe that I have succeeded in making the task natural and linguistic. Apart from the reassuring fact that children were enjoying the 'game' and some were asking for more, they were making comments that indicated that they were in an everyday situation, one that could have taken place in their classroom, and not just in an artificial experimental environment; for example: 'Will my sister meet these animals, too?' (Argiro 4;01).

Moreover, some children formed diminutives out of some words, in the regular way for Greek nouns. In the case of neuter nouns this is done by adding -aki to the stem of the noun, after removing the inflectional ending. So, for example, an animal called kixrˈo became kixrˈaki.

(5) *to* *mikrˈo* *kixrˈaki*
the.N.SG little.N.SG kixro.N.SG.DIM
'the little kixro'

This involved recognising the word as a neuter singular noun by the ending *-o,* removing the ending and adding the diminutive suffix. This was a linguistic operation that could not be carried out unless the child was involved in a linguistic task.

3.2.3 Materials. The experiment consisted of six conditions: the first three conditions involved words with sT, TR and TT clusters in word initial position, and the remaining three conditions contained words with the same clusters in word medial position. Specifically, the following combinations of consonants were tested:

(6) sT *sp, st, sk, sf, sx*
TR *tr, kl, fl, xr, vr*
TT *ft, xt, vð, ɣð, vɣ*

The construction of the nonwords used in the experiment followed the phonotactics of Greek. The words were either feminine or neuter nouns, with inflectional endings *-a* (feminine), *-i* (feminine or neuter), or *-o* (neuter). No masculine endings were used, because they involve (in the nominative) a word final consonant (*-s*), and that would increase the structural complexity of these trials. All words were bisyllabic, with a voiceless stop (*p, t* or *k*) as an onset for the non target syllable. There were five stimuli in each condition. The stimuli of the word initial conditions were the following:

(7) sT *sp'oki*[3], *st'ipo, sk'api, sf'ito, sx'ika*
TR *tr'ika, kl'ito*[4], *fl'api, xr'oki, vr'ipo*
TT *ft'ipo, xt'ika, vð'ito, ɣð'oki, vɣ'api*

The stimuli used in the word medial conditions were formed by reversing the syllable order. The stimuli were the following:

(8) sT *kisp'o, post'i, pisk'a, tosf'i, kasx'i*
TR *katr'i, tokl'i, pifl'a, kixr'o, povr'i*
TT *poft'i, kaxt'i, tovð'i, kiɣð'o, pivɣ'a*

For uniformity, the target cluster always preceded the stressed vowel. This creates pairs such as *sp'oki – kisp'o*. Note that both members of these pairs are well-formed in Greek, which is characterised by a lexical accent system, restricted by the trisyllabic window (i.e. stress must fall in one of the last three syllables of the word).[5]

3.2.4 Procedure. I first spent some time with the children in the classroom, taking part in their activities, so that I would become familiar to the children. After selecting children according to the linguistic and general developmental criteria discussed above, I tested each of the selected children individually in a separate room. Each session lasted about half an hour.

The test items were arranged in three different pseudo-random[6] orders so as to avoid sequence effects, and each of these orders was followed for a third of the children tested. There were four warm-up items without any clusters.

---

[3] Notice that *k* in Greek (and all the other velar consonants) becomes palatal before a front vowel. For example, *ɣð'oki* would be pronounced [*ɣð'oci*]. ın Cretan dialects, the velar might undergo even further fronting (Newton 1972). Indeed, all children exhibited some degree of fronting, the extent of which depended on the child's background. However, that does not affect our experiment in any crucial way. The stimulus producer's dialect has moderate fronting, typical of Cretan urban areas.

[4] One of the nurseries was in an area (Iraklio) where *l* tends to be is palatalised before *i.* For example, *tokl'i* would be pronounced [*tokʎ'i*]. Indeed, some children exhibited palatalisation of *l*. However, that does not affect our experiment in any crucial way.

[5] For analyses of the Greek stress system see Arvaniti (1991), Drachman and Malikouti-Drachman (1999), Malikouti-Drachman (1989), Philippaki-Warburton (1976), Ralli (1988), Revithiadou (1999) amongst others. For the acquisition of stress in Greek see Tzakosta (2004).

[6] Items were put in a random order, and then sequences consisting of three or more items belonging to the same category were broken up.

Pictures of novel animals were put inside a Russian doll representing a wizard. The child was told that the wizard had eaten some strange animals, and that he/she could free them by calling each animal with their name. The child was then invited to open the wizard, take out the animals one by one, and say their name. If after two attempts the child was not replying, we would move on to the next animal/ word, and the word would be added to the end of the list as the name of some other animal. The same (that is repetition of the word at the end) was done for words that were obscured by background noise. Designing the session in a way that involves an active task ensured that children's interest was kept throughout the experimental session.

Moreover, in order to vary the task, not all the pictures were inside the wizard-doll. Some were 'sleeping' inside a fairy's dress and the child was asked to wake them up, others were hiding inside a box with a small opening, through which only the child's hand could go, some others were absorbed in reading a book and got lost in its pages, some were in the belly of a smaller Russian doll representing a girl, where they went to keep warm, and, finally, some were hiding inside a pair of trousers, and the child was asked to find them so that I could put on my trousers. This way, the children's attention was constantly renewed and sessions were enjoyable for both the children and the experimenter.

During the session, there were spontaneous conversations between the child and the experimenter before, during, and after the task with the intention of giving the child and the experimenter some rest and keeping the child's attention. From these conversations (all DAT-recorded) information on the child's production of singletons was extracted.

3.2.5 Transcription and coding. The responses were transcribed on-line by the experimenter. The transcription was done in a fairly broad way, using the International Phonetic Alphabet. The sessions were also DAT recorded. The original transcriptions were then checked and amended off-line by the experimenter, with the aid of spectrographic analysis when necessary. Spectrographic analysis was used when a response was not entirely clear, and there was doubt as to the identity of the relevant consonants. Responses that were inaudible or covered by background noise were excluded[7].

An independent transcription was made by a second transcriber, who is a Greek native speaker and is well-trained in doing transcriptions. Ten percent of the data were cross-checked. In particular, one-tenth of the responses of each child were

---

[7] In all tests such cases were between 0 and 0.7 percent of total responses.

transcribed. The consistency rate between the two transcriptions, focusing on the cluster data, was 96 percent.

Moreover, notes where taken during the experiment and during the analysis of the recordings regarding any peculiarities of the child's speech. Specifically, care was taken to note any consistent substitutions that the child was making (in single consonant production). One such substitution was the substitution of *l* for *r* (9), and another common substitution was that of *θ* for *s* (10).

(9) *l* for *r* substitution (Emanouela 4;11,21)

a. Single consonant production	b. Cluster production
or'ea → *ol'ea* 'pretty'N.PL	*kart'i* → *kalt'i*
xor'ai → *xol'ai* 'fit'3RD SG	*kixr'o* → *kixl'o*

(10) *θ* for *s* substitution (Kali 3;00,03)

a. Single consonant production	b. Cluster production
pol'es → *pol'eθ* 'many'F.PL	*st'ipo* → *θt'ipo*
	sf'ito → *θf'ito*

Responses that involved one of these two substitutions were coded as correct.

During the coding, only changes in the consonant cluster were considered. Changes of any other consonant, any vowel or stress were ignored. Vowels were seldom changed, and neither was the stress pattern[8].

3.3 Results

Figure 1 below contains the percentage of correct responses for each of the clusters in word initial and word medial position. Percentages were calculated on the basis of conflated raw figures. This method of calculation was possible because of the structure of the data: there was an equal amount and type of data for each child.

[8] Coding was also performed using a set of alternative criteria, whereby any responses that involve a cluster belonging to the same category as the target cluster are coded as correct, even if the cluster is not the target one. The reason for implementing this coding criterion is that such responses may be taken as an indication that the child can produce the relevant structure, even if s/he is unable to produce the segmental content of the specific cluster. The use of these criteria did not alter the findings (for more details see Sanoudaki (submitted)).

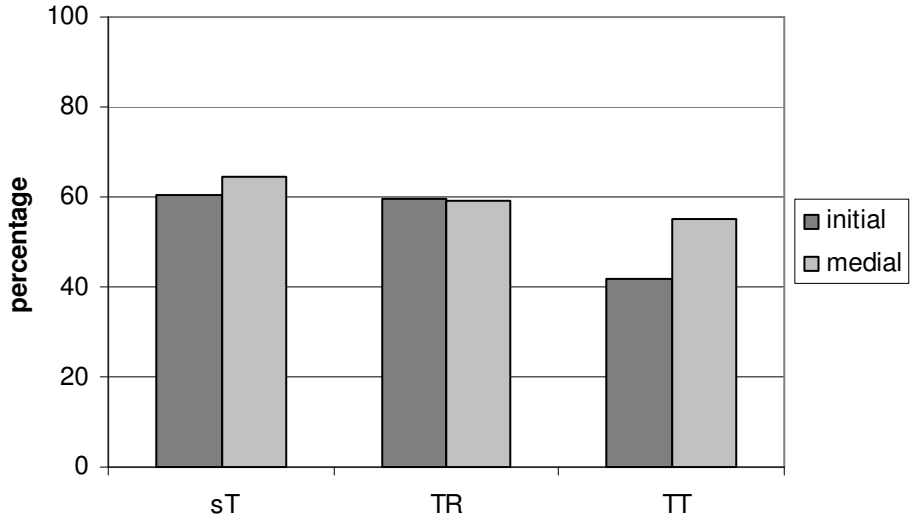


Fig. 1. Percentage of correct responses for word initial sT, TR and TT clusters in word initial and word medial position for all children combined

A visual examination of the figure indicates that only word initial TT is different. Detailed comparisons will be now presented for the cluster pairs that interest us, starting with word initial sT versus word initial TR.

The results for #sT and #TR are very similar, and no statistically significant difference was found (χ^2=0.034, p=0.859, DF=1). In addition, the table showing the number of correct responses for each child for the two conditions (#sT and #TR) is shown below (table 1). This organisation of the data allows us to look at the overall results in conjunction with the results of each individual child.

In table 1 the vertical dimension represents the number of correct responses in the #sT condition (from zero to five), while the horizontal dimension corresponds to the number of correct responses in the #TR condition (again from zero to five). One can therefore read out of the table the number of correct responses each child gave in the two conditions. For example, nine children (in the first row) gave no correct responses in the sT condition. Of these children, four (in the first cell starting from the left hand side) gave no correct responses in the TR position either, two (in the second cell) gave one correct response, two (third cell) two correct responses and so on. Children are divided into two groups, represented by the two sectors, divided by the diagonal: the top right sector contains children that performed better at TR, while the bottom left sector consists of children that performed better at sT. Children that fall on the diagonal performed the same in both conditions.

		#TR					
		0	**1**	**2**	**3**	**4**	**5**
#sT	**0**	////	//	//		/	
	1	//			//	/	
	2	/		//		/	
	3	/	/	/	//	///	/
	4	/	/	///	////	//// /	////// ///
	5				//	///	////

Table 1: Number of correct responses for # sT and #TR for each child

A visual examination of the table shows that the top right and the bottom left sector are equally populated. A one-variable chi-square test that was carried out to test the difference between the two sectors had a χ^2 value of 0.095, with an associated probability value of p=0.758, DF=1. The test found no statistically significant difference between the two sectors.

Moreover, the tally marks representing the children are scattered all over the table, showing that there is wide variation in performance. This includes children that performed almost adult-like in #sT but badly at #TR, and vice-versa, as well as children that were equally advanced in the two cluster types. Some examples of children, characteristic of the diversity, are given below. Kostantinos (11a) performed very badly at #sT and very well at #TR, while Fanouris (11b) showed the opposite pattern. Aglaia (11c) had roughly the same performance for the two cluster types, being only slightly better at #sT (3 correct responses out of 5 as opposed to 2 out of 5 for #TR).

(11) a. Kostantinos (2;11,17)

#sT: 1 out of 5 target	#TR: 4 out of 5 target
skˈapi → *θkˈapi*	*klˈito* → *klˈito*
spˈoki → *pˈoki*	*vrˈipo* → *vrˈipo*
stˈipo → *θˈipo*	*xrˈoki* → *xrˈoki*
sfˈito → *fˈito*	*trˈika* → *trˈika*
sxˈika → *çˈiçka*	*flˈapi* → *xlˈapi*

b. Fanouris (3;04,15)

#sT: 4 out of 5 target	#TR: 1 out of 5 target
sp'oki→ sp'oxi	*fl'api → fl'a*
st'ipo → st'ipo	*kl'ito → pl'ito*
sk'api → sk'api	*vr'ipo → l'ipo*
sf'ito → sf'ito	*xr'oki → l'oki*
sx'ika→ x'ika	*tr'ika → t'ika*

c. Aglaia (3;03)

#sT: 3 out of 5 target	#TR: 2 out of 5 target
sk'api → sk'api	*kl'ito → kl'ito*
sp'oki→ sp'oki	*fl'api → fl'api*
st'ipo → st'ipo	*vr'ipo → pt'ipo*
sf'ito → f'ito	*xr'oki → k'oki*
sx'ika→ x'ika	*tr'ika → t'ika*

In a #TT versus #TR comparison, figure 1 shows a considerable difference in the percentage of correct responses. Children performed better at the #TR condition, and the difference is statistically significant (χ^2=18.337, p<0.001, DF=1).

As before, the table containing the number of correct responses for each child for both conditions (#TT versus #TR) was drawn.

		#TR					
		0	**1**	**2**	**3**	**4**	**5**
#TT	**0**	**////** **////**	**//**	**/**	**//**	**/**	**/**
	1		**/**	**////**	**/**	**//**	**//**
	2	**/**		**//**	**//**	**///**	**/**
	3			**/**	**/**	**///**	**////**
	4		**/**		**///**	**////**	**///**
	5				**/**	**/**	**///**

Table 2: Number of correct responses for #TT and #TR for each child

The vertical dimension represents the number of correct responses in the #TT condition (from zero to five), while the horizontal dimension shows the number of correct responses in the #TR condition (again from zero to five). A visual examination of the table shows that the top right sector, corresponding to children that performed better at #TR, is much more populated than the bottom left sector, which includes children that performed better at #TT. The difference is statistically significant (χ^2=14.400, p<0.001, DF=1). Several children performed well at #TR and badly at #TT, while the reverse pattern was uncommon[9]. In (12) I give some examples of individual children's performance. Kostantinos (12a) and Agelos (12b) performed very well at #TR and badly at #TT. Zoi (12c) performed better than the two previous children at #TT and adult-like at #TR.

(12) a. Kostantinos (2;11,17)

#TR: 4 out of 5 target	#TT: 1 out of 5 target
trˈika → *trˈika*	*ftˈipo* → *ftˈipo*
klˈito → *klˈito*	*vðˈito* → *vˈito*
xrˈoki → *xrˈoki*	*ɣðˈoki* → *ðˈoki*
vrˈipo → *vrˈipo*	*vɣˈapi* → *ɣˈapi*
flˈapi → *xlˈapi*	*xtˈika* → *tˈixa*

b. Agelos (3;04,12)

#TR: 5 out of 5 target	#TT: 0 out of 5 target
trˈika → *trˈika*	*xtˈika* → *ftˈika*
klˈito → *klˈito*	*ftˈipo* → *stˈipo*
xrˈoki → *xrˈoki*	*vðˈito* → *vˈito*
vrˈipo → *vrˈipo*	*ɣðˈoki* → *xrˈoki*
flˈapi → *flˈapi*	*vɣˈapi* → *ɣˈapi*

c. Zoi (4;02,17)

#TR: 5 out of 5 target	#TT: 3 out of 5 target
trˈika → *trˈika*	*vɣˈapi* → *vɣˈapi*
klˈito → *klˈito*	*xtˈika* → *xtˈika*
xrˈoki → *xrˈoki*	*ɣðˈoki* → *ɣðˈoki*
vrˈipo → *vrˈipo*	*vðˈito* → *ðˈito*
flˈapi → *flˈapi*	*ftˈipo* → *xtˈipo*

[9] In the cases of children that performed better at TT, the difference between TT and TR is small: specifically, there were only cases of one response difference (3-2, 4-3, 5-4), two responses difference (2-0, 5-3) and one case of three responses difference (4-1) (see table 2).

The results for initial sT and initial TT also differ significantly (χ^2= 19.866, p<0.001, DF=1), with children performing better in the sT condition.

The table showing the number of correct responses for each child for both conditions (#sT and #TT) is given below.

		#sT					
		0	**1**	**2**	**3**	**4**	**5**
#TT	**0**	//// ///	///	/	//	//	
	1	//	/	/	/	//// /	
	2		/	//	///	///	
	3				/	//// /	///
	4				//	//// ///	//
	5					/	///

Table 3: Number of correct responses for #sT and #TT for each child

More children performed better at sT than at TT (χ^2=16.9, DF=1, p=0.001). The top right sector is much more populated than the bottom left one. There is a large number of children that performed well at sT and badly at TT, while few children performed better at TT than at sT[10]. Examples of individual children's performance are given below. Agelos (13a) and Maro (13b) performed very well at sT and very badly at TT. Zoi's (13c) performance at TT was better than that of the two previous children, and at sT her performance was adult-like.

[10] In the cases of children that performed better at TT, the difference between TT and sT is consistently small: specifically, there were only cases of one response difference (1-0, 2-1, 4-3, 5-4) (see table 3).

(13) a. Agelos (3;04,12)

#sT: 4 out of 5 target	#TT: 0 out of 5 target
skˈapi → *skˈapi*	*xtˈika* → *ftˈika*
spˈoki→*spˈoki*	*ftˈipo* → *stˈipo*
stˈipo → *stˈipo*	*vðˈito* → *vˈito*
sfˈito → *sfˈito*	*ɣðˈoki* → *xrˈoki*
sxˈika→ *skˈika*	*vɣˈapi* → *ɣˈapi*

b. Maro (3;09,23)

#sT: 4 out of 5 target	#TT: 1 out of 5 target
spˈoki→ *spˈoki*	*ftˈipo* → *ftˈipo*
stˈipo → *stˈipo*	*xtˈika* → *tˈika*
skˈapi → *skˈapi*	*vðˈito* → *ðˈito*
sxˈika→ *sxˈika*	*ɣðˈoki* → *ðiˈoki*
sfˈito → *fˈisto*	*vɣˈapi* → *vgˈapi*

c. Zoi (4;02,17)

#sT: 5 out of 5 target	#TT: 3 out of 5 target
spˈoki→ *spˈoki*	*vɣˈapi* → *vɣˈapi*
stˈipo → *stˈipo*	*xtˈika* → *xtˈika*
skˈapi → *skˈapi*	*ɣðˈoki* → *ɣðˈoki*
sxˈika→ *sxˈika*	*vðˈito* → *ðˈito*
sfˈito → *sfˈito*	*ftˈipo* → *xtˈipo*

Having examined the results in the word initial conditions, I now compare the results in the word initial position with those in the word medial position, starting with sT clusters. The percentage of target responses in word initial and word medial sT does not differ significantly (χ^2=1.225, p=0.268, DF=1).

Moreover, a table containing the number of correct responses for each child in both conditions is constructed.

		-sT					
		0	**1**	**2**	**3**	**4**	**5**
#sT	**0**	//// //	/	/		/	
	1	//		/	/	/	
	2			/	/	//	
	3			/	///	//// /	
	4			/	////// ///	//// ///	//// //
	5					//	//// ///

Table 4: Number of correct responses for #sT and -sT for each child

There is no statistically significant difference between the two sectors (χ^2=0.714, p=0.398, DF=1). Some examples of individual children's performance are listed in (14a-c) below. Emanouela (14a) performed adult-like in both conditions, while Epistimi (14b) did not give any correct responses in either condition. Finally, Maraki's performance (14c) was between that of the two previous children, with two correct responses (out of five trials) in each condition.

(14) a. Emanouela (4;11,21)

#sT: 5 out of 5 target	-sT: 5 out of 5 target
skˈapi → *skˈapi*	*piskˈa* → *piskˈa*
spˈoki→*spˈoki*	*kispˈo*→ *kispˈo*
stˈipo → *stˈipo*	*postˈi* → *postˈi*
sfˈito → *sfˈito*	*tosfˈi* → *tosfˈi*
sxˈika→ *sxˈika*	*kasxˈi*→ *kasxˈi*

b. Epistimi(2;03,08)

#sT: 0 out of 5 target	-sT: 0 out of 5 target
spˈoki→ *pˈoki*	*piskˈa* → *pijkˈa*
stˈipo → *sˈipo*	*kispˈo*→ *kipsˈo*
skˈapi → *kˈapi*	*postˈi* → *kutˈi*
sxˈika→ *sˈika*	*tosfˈi* → *tofˈi*
sfˈito → *pʰˈipo*	*kasxˈi*→ *kakˈi*

c. Maraki (3;05,03)

#sT: 2 out of 5 target	-sT: 2 out of 5 target
spˈoki→ spˈoki	*piskˈa → piskˈa*
skˈapi → skˈapi	*postˈi → postˈi*
stˈipo → tˈipo	*kispˈo→ kipˈo*
sxˈika→ sˈika	*tosfˈi → sofˈi*
sfˈito → θˈito	*kasxˈi→ kaskˈi*

A further examination of table 4 reveals an imbalance, which might have affected the result of the statistical test. Specifically, there is a high concentration of tally marks around the bottom right corner, indicating that our sample includes a high number of children that were very advanced in sT cluster production. This concentration of children that are advanced in both conditions may have overshadowed the results coming from children at earlier stages of sT acquisition, and given overall results of no difference between the two conditions, while in fact there exists one in earlier stages of acquisition.

In order to control for this, I divide the children into three age groups so that the performance of younger children can be examined separately. Group 1 contains the youngest children (covering one-year age difference starting with the youngest one 2;03-3;05 n=24), group 3 the oldest children (one-year age difference 4;00-5;00 n=17) and group 2 the children between the two other groups (3;06-3;11 n=18). Figure 2 contains the percentage of correct responses for the two positions by age group.

The assumption behind this decision is that older children perform better than younger ones. Indeed, there is a statistically significant difference between age groups in word initial position (χ^2=26.488, p<0.001, DF=2) as well as in word medial position (χ^2=20.360, p<0.001, DF=2). A look at the results, figure 2, shows that the difference follows the expected direction; in both word initial and word medial position, performance improves with age. The results of group 3, the oldest group, are better than the results of group 2, which, in turn, are better than the results of group 1, the youngest group, in both positions.

After having checked that the assumption that older children perform better is supported, we can proceed to test whether there is a difference between children's performance in the word initial and the word medial position in each age group. No statistically significant difference was found for group 1 (age 2;03-3;05) (χ^2=1.082, p=0.298, DF=1), for group 2 (age 3;06-3;11) (χ^2=0.423, p=0.515, DF=1) or for group 3 (age 4;00-5;00) (χ^2=0.033, p=0.855, DF=1).

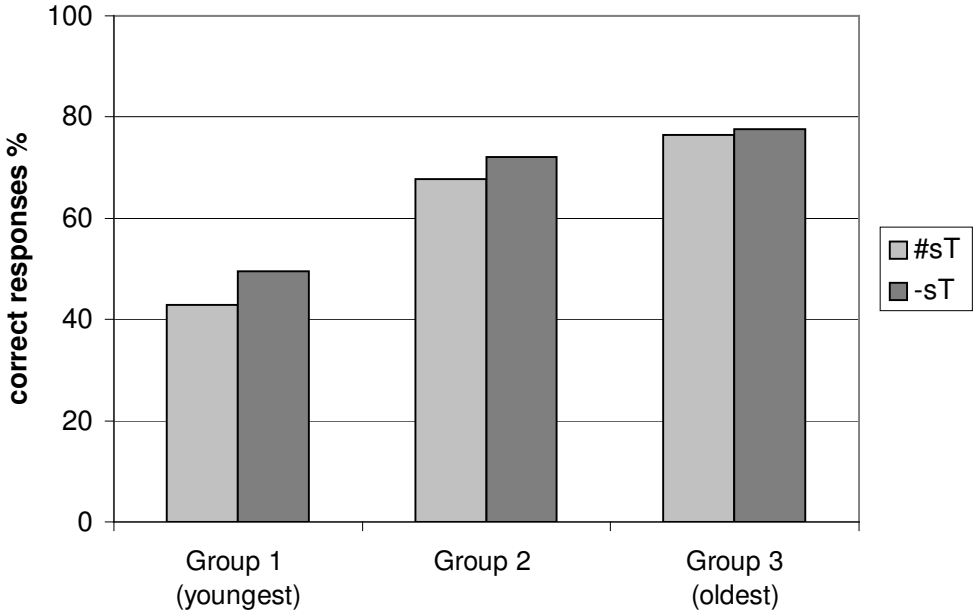


Figure 2. Percentage of correct responses for word initial versus word medial sT by age group

In a comparison of children's responses in the word initial TR and word medial TR condition, no significant difference is found (χ^2=0.007, p<0.933, DF=1).

In addition, the table showing the number of correct responses for each child for both conditions (#TR and -TR) is given below (table 5).

The top right and the bottom left sector (divided by the diagonal line), corresponding to children that performed better at #TR and -TR respectively, are equally populated (χ^2=0.111, p=0.739, DF=1).

There is a concentration of tally marks around the diagonal, indicating that children tended to perform equally well in both conditions. In (15) below, I give some examples of children's performance, illustrating this tendency. Stavros (15a) performed adult-like in both conditions, while Lena (15b) performed poorly in both conditions. Finally, Mario's performance (15c) was better than Lena's, but the child was still having problems with TR clusters in both positions.

(15) a. Stavros (3;11,24)

-TR: 5 out of 5 target	#TR: 5 out of 5 target
katrˈi → *katrˈi*	*trˈika* → *trˈika*
toklˈi → *toklˈi*	*klˈito* → *klˈito*
piflˈa → *piflˈa*	*flˈapi* → *flˈapi*

kixrˈo → kixrˈo *xrˈoki → xrˈoki*
povrˈi → povrˈi *vrˈipo → vrˈipo*

b.Lena (2;10,28)

-TR: 1 out of 5 target	#TR: 0 out of 5 target
toklˈi → toklˈi	*trˈika → tˈika*
katrˈi → katˈi	*klˈito → cˈito*
piflˈa → piçfˈa	*flˈapi → fˈapi*
kixrˈo → kixˈo	*xrˈoki → xˈoki*
povrˈi → povˈi	*vrˈipo → vˈipo*

c. Mario (3;03,01)

-TR: 3 out of 5 target	#TR: 2 out of 5 target
katrˈi → katrˈi	*trˈika → trˈika*
toklˈi → klˈi	*klˈito → klˈito*
piflˈa → iflˈa	*flˈapi → vlˈapi*
kixrˈo → ixkˈo	*xrˈoki → ˈoti*
povrˈi → tovˈi	*vrˈipo → vˈipo*

-TR \ #TR	**0**	**1**	**2**	**3**	**4**	**5**
0	//// //					
1	//	///		//		/
2	/		//	///	/	/
3			//	/	//// /	/
4	/		///	//	//// /	//// /
5			/	//	///	//// //

Table 5: Correct responses for initial and medial TR for each child

Finally, in a comparison of children's performance in word initial TT and word medial TT, a significant difference appears (χ^2=10.319, p<0.001, DF=1). As seen in figure 1, children's performance was better word medially.

Furthermore, I present below a table containing the number of correct responses for each child in the word initial and word medial condition.

		-TT					
		0	**1**	**2**	**3**	**4**	**5**
#TT	**0**	//// //	//// /	//	/		
	1	/	/	///	/	///	/
	2	//	/	/	//	//	/
	3			//	//// /	//	//
	4				//	///	//// /
	5					/	////

Table 6: Correct responses for initial and medial TT for each child

A visual examination of the table shows that most children performed better at medial TT than at initial TT. The top right sector is populated much more than the bottom left one, and the difference in statistically significant (χ^2=11.3, p=0.001, DF=1). There is a large number of children that performed well at -TT and badly at #TT, while there were few children that performed better at #TT than at -TT[11]. Some examples of individual children's performance are given below. Kostantinos (136a) performed adult-like in the word medial condition, while his performance in the word initial condition was very poor. Manouela's performance (136c) was very poor in both conditions, while Mario (136b) gave correct responses about half of the time in both conditions.

(16) a. Kostantinos (2;11,17)

-TT: 5 out of 5 target	#TT: 1 out of 5 target
poftˈi → *poftˈi*	*ftˈipo* → *ftˈipo*
kaxtˈi → *kaxtˈi*	*xtˈika* → *tˈixa*
tovðˈi → *tovðˈi*	*vðˈito* → *vˈito*

[11] In the cases of children that performed better at initial TT, the difference between the word initial condition and the word medial condition is consistently small: either one response difference (1-0, 2-1, 3-2, 4-3, 5-4) or maximum two responses difference (2-0) (see table 6).

kiɣðˈo → *kiɣðˈo*	*ɣðˈoki* → *ðˈoki*
pivɣˈa → *pivɣˈa*	*vɣˈapi* → *ɣˈapi*

b. Mario (3;03,01)

-TT: 3 out of 5 target	#TT: 2 out of 5 target
poftˈi → *poftˈi*	*ftˈipo* → *ftˈipo*
kaxtˈi → *taxtˈi*	*xtˈika* → *xtˈipa*
tovðˈi → *oðˈi*	*vðˈito* → *vˈito*
pivɣˈa → *tivɣˈa*	*ɣðˈoki* → *vðˈoki*
kiɣðˈo → *iɣˈo*	*vɣˈapi* → *xˈapi*

c. Manouela (2;11,19)

-TT: 1 out of 5 target	#TT: 0 out of 5 target
kaxtˈi → *ixtˈi*	*ftˈipo* → *pˈipo*
poftˈi → *xutˈi*	*xtˈika* → *pˈika*
tovðˈi → *tuvlˈi*	*vðˈito* → *θˈito*
kiɣðˈo → *tliɣˈo*	*ɣðˈoki* →*klˈoki*
pivɣˈa → *fijˈa*	*vɣˈapi* → *kˈapi*

To sum up, some children performed better at word initial sT than TR, while others performed better at TR than at sT, creating a balance in the overall results. Children's performance at word initial TT was systematically worse than at word initial sT and word initial TR. In a comparison with the word medial position, no difference was found in children's performance at sT (initial versus medial) and TR (initial versus medial), while there was a difference between word initial and word medial TT.

4 Analysis

4.1 Extrasyllabicity-problems

The results regarding initial sT versus initial TR were as expected. No overall difference was found between #sT and #TR clusters. These results were representative of the paradox that is found in the acquisition literature in other languages, with some of the children acquiring sT before TR and some following the opposite path.

These results, combined with the results for #TT and #TR are particularly problematic for the extrasyllabic analysis of sT and TT. Extrasyllabicity would only be able to account for one set of data: either the TT versus TR, or sT versus TR. The TT versus TR results could be explained by an extrasyllabicity model

according to which extrasyllabic structures are more marked than regular branching onsets and are therefore expected to be acquired later. This model would explain late acquisition of TT when compared to TR, but would fail to tackle the paradox of sT versus TR variation. On the other hand, the TT versus TR data would not be covered by the amended extrasyllabicity proposals discussed in section 2.2 – namely a) that extrasyllabicity can be acquired before or after branching onsets, and b) that in some children's grammar, extrasyllabic clusters are structured like affricates, and are therefore acquired before branching onsets. Any such proposal would cover the data it was designed to explain (sT-TR variation), but would have to answer the question of why the same variation is not found in TT versus TR acquisition. Either way, the results are problematic for the extrasyllabic approach. Evidence for the different nature of sT and TT was also found in the comparison with their word medial counterparts. Although word initial TT was acquired later than word medial TT, such imbalance was not found with sT clusters.

In order to account for the data, it would be possible to add an auxiliary hypothesis that assumes two different kinds of extrasyllabicity, one for #sT and one for #TT. However, this would not be enough: we would further have to stipulate the order of acquisition of these different structures. Specifically, we would have to stipulate that sT-type extrasyllabicity is acquired before TT-type extrasyllabicity. If TT was found to be acquired before sT, extrasyllabicity could simply stipulate that it is TT that is acquired before sT, and thus be made consistent with the opposite reality.

To make matters even more complicated, in languages that have both, sT and TT show identical behaviour in some adult language phenomena. Specifically, there is evidence from languages that have both #sT and #TT that the two are syllabically the same (Seigneur-Froli 2006; Steriade 1982). A well-known example is attic Greek reduplication: past forms of roots beginning with sT and TT follow the same pattern, in contrast to verbs that begin with TR. The imperfective past forms of roots commencing with a single consonant (including *s*) are formed by reduplication; an initial syllable consisting of the first consonant followed by *e* is added (17a). In the case of roots commencing with TR clusters, reduplication also takes place (the initial syllable consists of the obstruent plus *e*) (17b). In contrast, in the case of roots commencing with TT no reduplication takes place: the vowel *e* is added word initially (17c). As for the past forms of roots starting with sT, these are formed in the same way as TT initial roots (17d).

(17) a. CV

Present	Past	
paide'u-oo	*pep'aideuka*	'bring up'
l'uoo	*l'eluka*	'loosen'
sale'uoo	*ses'aleumai*	'cause to rock'

b. TR

kr'inoo	*k'ekrika*	'pick out'
kl'inoo	*k'eklika*	'make to bend'
pl'eoo	*p'epleuka*	'sail'

c. TT

pt'aioo	*'eptaika*	'make to stumble'
kt'einoo	*'ektaamai*	'kill'
p^ht^h'anoo	*'ep^ht^haka*	'come first'

d. sT

sp'aoo	*'espaka*	'draw' (a sword)
st'elloo	*'estalka*	'arrange'
ski'azoo	*'eskiasmai*	'shade'

Such behaviour has led phonologists to conclude that sT and TT in (ancient) Greek have the same structure (Seigneur-Froli 2006; Steriade 1982). On the other hand, our experimental results demonstrate that sT and TT are different in some way crucial to first language acquisition. The question arises, if sT and TT have the same structure, why are they not acquired together?

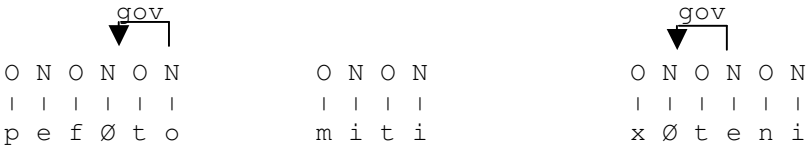
4.2 Towards an analysis

If we try to categorise word initial clusters based on the acquisition data, the division appears to be between TR and sT on one side and TT on the other side. The acquisition of TT clusters requires an extra step when compared to the rest of the word initial clusters.

Interestingly, this descriptive division corresponds to a theoretical division that has been suggested on entirely different grounds, based on adult language phenomena. Scheer (2000, 2004), divides (adult) languages into those that allow word initial TT clusters and those that do not. The theoretical distinction he proposes is the absence versus presence of an onset nucleus pair at the left margin of the word. The theoretical proposal is part of a system that defines structure according to relationships segments establish along the syntagmatic dimension i.e.

governing and licensing relations with what follows and what precedes, thus eliminating vertical-branching structure.

(18) Syntagmatic representations: *p'efto* '(I) fall', *m'iti* 'nose', *xt'eni* 'comb'

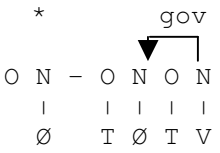


The examples in (18) show the representations of three Greek words. As may be seen, the skeleton is a sequence of onsets (consonantal positions) and nuclei (vocalic positions), which may or may not enjoy segmental instantiation. An empty nucleus Ø is allowed to exist if it is followed by a filled nucleus, which can govern the empty position.

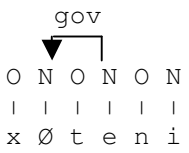
Based on Lowenstamm's (1999) proposal that the left margin of the word (traditionally noted as #), corresponds to an onset nucleus pair without any segmental content (ON), Scheer proposes a parameterisation of the initial ON. The existence of an initial ON pair in a language creates a ban on word initial TT clusters. This is because the empty nucleus of the initial ON would fail to be governed, since the following nucleus (within the TT cluster) is itself empty (19a). Absence of the initial ON in a language makes the existence of initial TT clusters possible (19b).

(19) Parameterisation of initial ON and typology of #TT clusters

a. ON present (English) *#TT



b. ON absent (Greek) #TT



On the other hand, the presence or absence of the initial ON pair does not affect the existence of word initial TR and sT clusters, which have alternative ways of governing the empty nucleus of the initial ON (for TR see Scheer (2004), for sT see Sanoudaki (submitted)). The proposal finds independent support in diachronic lenition and fortition phenomena (Seigneur-Froli 2003; 2006).

Extending this to first language acquisition, the presence of an initial ON pair in a developing grammar creates a ban on word initial TT clusters. Word initial TT clusters only appear when the initial ON pair has disappeared from the child's

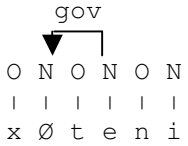
grammar[12]. This can explain the later acquisition of initial TT when compared to other word initial clusters.

(20) Acquisition stage n-1: ON present: No TT clusters
Acquisition stage n: ON absent: TT clusters

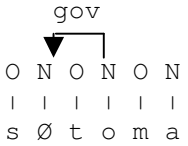
Moreover, once the initial ON pair has disappeared (in other words, when initial TT is acquired) initial sT and initial TT have the same structure, as that is defined by the governing and licensing relations in their environment[13].

(21) Word initial TT (*xtˈeni* 'comb') and sT (*stˈoma* 'mouth') in Greek

a. #TT



b. #sT



Thus, the discrepancy between adult language and first language acquisition whereby sT is acquired earlier, while in adult language sT and TT behave identically, is predicted.

The remaining findings are also consistent with this model. Word initial TT is acquired later than its word medial counterpart because initial TT, unlike medial TT, has the extra requirement that the ON pair be absent. There is no difference in the acquisition of sT in initial versus medial position, since no such extra requirement is involved. The same holds for initial versus medial TR. Finally, the optionality in the acquisition of sT versus TR can be attributed to optionality in the mastering of the relevant structure: along the lines of Fikkert's (1994) suggestion (see section 2.2), some children master the sT structure first and others the TR structure first. Fikkert's suggestion involves extrasyllabicity versus branching onsets, while in a syntagmatic view of phonology different structures would be involved, but the suggestion is of the same nature. For details of a model of consonant cluster acquisition based on this view on phonology, the reader is referred to Sanoudaki (submitted).

---

[12] For theoretical motivation of the acquisition stages based on learnability issues see Sanoudaki (submitted).

[13] The representations in (18), (19) and (21) are simplified for expository reasons. For complete representations see work mentioned above.

5 Conclusion

Despite what most phonologists would think, sT and TT in word initial position are different. The existence of the difference would not have been discovered without the help of developmental data, which show that Greek children acquire TT later than sT. The nature of the difference was further examined by comparing children's production of different clusters in different positions. While word initial sT is acquired before TR by some children and after TR by others, TT is systematically acquired later than TR. Moreover, initial TT was acquired later than its word medial counterpart, while no such difference was found for word initial versus word medial sT. These findings point against existing extrasyllabic analyses of these clusters and indicate a division between initial sT and TR on the one hand and initial TT on the other hand, which is best captured by Lowenstamm's initial ON hypothesis.

References

Arvaniti, A. (1991). *The phonetics of Modern Greek and its phonological implications*. Ph.D. Dissertation. University of Cambridge.

Barlow, J. (1997). *A constraint-based account of syllable onsets: evidence from developing systems.* Ph.D. Dissertation. Indiana University.

Cairns, C. E. (1988). Phonotactics, markedness and lexical representations. *Phonology*. vol. 5. pp. 209-236.

Chin, S. B. (1996). The role of the sonority hierarchy in delayed phonological systems. In *Pathologies of speech and language: contributions of clinical phonetics and linguistics*, T. Powell, ed., International Clinical Phonetics and Linguistics Association, New Orleans, pp. 109-117.

Clements, G. (1990). The role of sonority cycle in core syllabification. In *Papers in Laboratory phonology I*, J. Kingston & M. Beckmann, eds., Cambridge University Press, Cambridge, pp. 283-333.

Demuth, K. (1996). Stages in the acquisition of prosodic structure. In *Proceedings of the 27th child language research forum*, E. V. Clark, ed., CSLI, Stanford University.

Demuth, K. & M. M. Kehoe (2006). The acquisition of word-final clusters in French. *Catalan Journal of Linguistics*. vol. 5. pp. 59-81.

Dollaghan, C. & T. F. Campbell (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*. vol. 41. pp. 1136-1146.

Drachman, G. & A. Malikouti-Drachman (1999). Greek word accent. In *Word prosodic systems in the languages of Europe*, H. van der Hulst, ed., Mouton de Gruyter, Berlin, pp. 897-945.

Fikkert, P. (1994). *On the acquisition of prosodic structure.* HIL Dissertations, Dordrecht.

Freitas, M. J. (2003). The acquisition of onset clusters in European Portuguese. *Probus*. vol. 15. pp. 27-46.

Gathercole, S. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*. vol. 23. pp. 83-94.

Gathercole, S. E., C. Willis, A. D. Baddeley, & H. Emslie (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory*. vol. 2. pp. 103-127.

Gierut, J. (1999). Syllable onsets: clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research*. vol. 42. pp. 708-726.

Halle, M. & J.-R. Vergnaud (1980). Three-dimensional phonology. *Journal of Linguistic Research*. vol. 1. pp. 83-105.

Jakobson, R. (1968). *Child language, aphasia and phonological universals.* Mouton, Hague and Paris.

Jongstra, W. (2003). *Variation in reduction strategies of Dutch word-initial consonant clusters*. Ph.D. Dissertation. University of Toronto.

Kappa, I. (2000). On the acquisition of fricatives in Modern Greek. In *Proccedings of the 13th international Symposium on Theoretical and Applied Linguistics*, K. Nicolaidis & M. Mattheoudaki, eds., Aristotle University of Thessaloniki, Thessaloniki, pp. 197-205.

Kappa, I. (2001). Final coda licensing in Greek. In *Proceedings of the Gala 2001 conference on language acquisition*, J. Costa & M. J. Freitas, eds., pp. 320-326.

Kappa, I. (2002a). On the acquisition of syllabic structure in Greek. *Journal of Greek Linguistics*. vol. 3. pp. 1-52.

Kappa, I. (2002b). On the phonological development of the early words: a case study. In *Selected papers on Theoretical and Applied Linguistics*, M. Makri-Tsilipakou, ed., Aristotle University of Thessaloniki, Thessaloniki, pp. 124-133.

Kehoe, M. M. & C. Stoel-Gammon (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of child language*. vol. 28. pp. 393-432.

Kirk, C. & K. Demuth (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of child language*. vol. 32. pp. 709-734.

Kirk, C. & K. Demuth (2006). Accounting for variability in 2-year-olds' production of coda consonants. *Language learning and development*. vol. 2. pp. 97-118.

Kula, N. C. & M. Tzakosta (2002). *Acquisition of complex structures in Greek: a GP and an OT approach.* Ms. Leiden University.

Laws, G. (1998). The use of nonword repetition as a test of phonological memory in children with Down syndrome. *Journal of Child Psychology and Psychiatry*. vol. 39. pp. 1119-1130.

Levin, J. (1985). *A metrical theory of syllabicity*. Ph.D. Dissertation. MIT.

Lleó, C. & M. Prinz (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment. *J.Child Lang.* vol. 23. pp. 31-56.

Lleó, C. & M. Prinz (1997). Syllable structure parameters and the acquisition of affricates. In *Focus on phonological acquisition*, M. Young-Scholten & S. J. Hannas, eds., John Benjamins, Philadelphia, pp. 143-164.

Lowenstamm, J. (1999). The beginning of the word. In *Phonologica 1996*, J. Rennison & K. Kühnhammer, eds., Holland Academic Graphics, La Hague, pp. 153-166.

Malikouti-Drachman, A. & G. Drachman (1989). [Tonismos sta Ellinika] Stress in Greek. In *Studies in Greek linguistics*, University of Thessaloniki, Thessaloniki, pp. 127-143.

Newton, B. (1972). *The generative interpretation of dialect: a study of modern Greek phonology.* Cambridge University Press, London.

Nicolaidis, K., J. Edwards, M. Beckman, & G. Tserdanelis (2004). Acquisition of lingual obstruents in Greek. In *Proceedings of the 6th international conference of Greek linguistics*, University of Crete, Rethymno.

Pan, N. (2005). *A Government Phonology approach to the acquisition of syllable structure*. Ph.D. Dissertation. University of Louisiana at Lafayette.

Pan, N. & W. Snyder (2004). Acquisition of /s/-initial clusters: a parametric approach. In *Proceedings of the 28th annual Boston University conference on language development*, A. Brugos, L. Micciulla, & C. E. Smith, eds., Cascadilla press, Somerville,MA, pp. 436-446.

Papadopoulou, K. (2000). *Phonological acquisition of Modern Greek.* Unpublished BsC Honours Dissertation. University of Newcastle upon Tyne, UK.

Philippaki-Warburton, I. (1976). On the boundaries of morphology and phonology: a case study from Greek. *Journal of Greek Linguistics*. vol. 12. pp. 259-278.

Ralli, A. (1988). *Éléments de morphologies du Grec Moderne*. Ph.D. Dissertation. Université de Montreal.

Revithiadou, A. (1999). *Headmost accent wins*. Ph.D. Dissertation. LOT.

Rubach, J. & G. Booij (1990). Edge of constituent effects in Polish. *Natural language and linguistic theory*. vol. 8. pp. 427-463.

Sanoudaki, E. (submitted). *A CVCV model of consonant cluster acquisition: evidence from Greek.*

Scheer, T. (2000). *De la localité, de la morphologie et de la phonologie en phonologie.* Habilitation thesis. University of Nice.

Scheer, T. (2004). *A lateral theory of phonology. Volume 1: What is CVCV, and why should it be?* Mouton de Gruyter, Berlin.

Scobbie, J. M. (1997). *Autosegmental representation in a Declarative Constraint-based framework.* Garland Press, New York.

Seigneur-Froli, D. (2003). Diachronic consonantal lenition & exotic word-initial clusters in Greek: a unified account. In *Studies in Greek linguistics 23. Proceedings of the 23rd annual meeting of the department of linguistics of AUTH*, M. Stavrou-Sifaki & A. Fliatouras, eds., Faculty of Philosophy, Aristotle University of Thessaloniki, Thessaloniki, pp. 345-357.

Seigneur-Froli, D. (2006). *Statut phonologique du début de mot grec*. Ph.D. Dissertation. Université de Nice Sofia-Antipolis.

Selkirk, E. (1982). The syllable. In *The structure of phonological representations, Part 2*, H. van der Hulst & N. Smith, eds., Foris, Dordrecht, pp. 337-383.

Smith, N. V. (1973). *The acquisition of phonology, a case study.* Cambridge University Press, Cambridge.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. Ph.D. Dissertation. Garland Press.

Stites, J., K. Demuth, & C. Kirk (2004). Markedness vs. frequency effects in coda acquisition. In *Proceedings of the 28th annual Boston University Conference*.

Tzakosta, M. (2001a). I pragmatwsi twn sumfwnikwn sumblegmatwn ston paidiko logo. In *Proceedings of the 21st annual meeting of Greek Linguistics*, Faculty of Philosophy, Aristotle University of Thessaloniki, Thessaloniki, pp. 736-746.

Tzakosta, M. (2001b). The acquisition of manner features in Modern Greek. In *Proceedings of HILP5*, C. A. Féry, D. Green, & R. van de Vijver, eds., University of Potsdam, Potsdam, pp. 178-193.

Tzakosta, M. (2003). Four of a kind: stress patterns in child language. In *Proceedings of the 23rd annual meeting of Greek linguistics*, Faculty of philosophy, Aristotle University of Thessaloniki, Thessaloniki, pp. 467-478.

Tzakosta, M. (2004). *Multiple parallel grammars in the acquisition of stress in Greek L1*. Ph.D. Dissertation. LOT.

van der Lely, H. K. J. & D. Howard (1993). Children with specific language impairment: linguistic impairment or short-term memory deficit? *Journal of speech and hearing research*. vol. 36. pp. 1193-1207.

Vanderweide, T. (2005). *Cue-based learning and the acquisition of pre-vocalic clusters*. Ph.D. Dissertation. University of Calgary.

Weijer, J. v. d. (1993). The manner-place dependency in complex segments. *Linguistics*. vol. 31. pp. 87-110.

Weismer, S. E., J. B. Tomblin, X. Zhang, P. Buchwalter, J. G. Chynoweth, & M. Jones (2000). Nonword repetition in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*. vol. 43. pp. 865-878.

Zamuner, T. & L. A. Gerken (1998). Young children's production of coda consonants in different prosodic environments. In *The proceedings of the twenty-ninth annual child language research forum*, E. V. Clark, ed., pp. 121-128.

Zamuner, T., L. A. Gerken, & M. Hammond (2004). Phonotactic probabilities in young children's speech production. *Journal of child language*. vol. 31. pp. 515-536.

Syntax

*Verb movement and VSO-VOS alternations**

DIRK BURY

Abstract

This paper discusses an approach to syntax that offers a new perspective on verb-initial languages that allow both VSO and VOS orders. A central assumption is that word order is not rigidly determined by structure. The predictions of this model are discussed in the context of St'át'imcets, Chamorro and Tongan. A comparison with several alternative analyses of VSO-VOS alternations shows that this model is the only one that predicts that a language should display such a word order alternation. In the alternative models, one of the two orders is basic and the other is derived by a special rule.

1 Introduction

This paper discusses an approach to syntax that may shed new light on the structure of certain verb-initial languages that allow clauses to alternate more or less freely between VSO and VOS order. A central feature of this approach is the assumption that word order is not rigidly determined by phrase structure and that instead a single structure may be linearised in different, well-defined ways. After introducing this model, its predictions are discussed in the context of three languages that display VSO-VOS alternations, St'át'imcets, Chamorro and Tongan. A comparison of the model with several alternative analyses of VSO-VOS alternations shows that one feature sets it apart from these alternatives: this model is the only one that predicts that a language should display such a word order alternation. In the alternative models discussed here one of the two orders is basic and the other order is derived by a special rule.

Section 2 provides the theoretical background. This paper loosely follows up on an earlier proposal (cf. Bury 2005) and this section reviews the relevant features of that proposal. Section 3 explain why this model could be of interest for the analysis of languages with a VSO-VOS alternation and explores how well this model's predictions are realised in St'át'imcets, Chamorro and Tongan, three languages that

* Address for correspondence: School of Linguistics and English Language, University of Wales, Bangor LL57 2DG. Part of this material was presented in Vancouver (UBC), Leiden, Bangor, and at Gregynog. Many thanks to the audiences for their comments and questions. This work was supported by a British Academy small research grant (SG-43015).

display a VSO-VOS alternation. Section 4 summarises three alternative analyses of VSO-VOS alternations and contrasts these analyses with the one proposed here. Section 5 concludes the paper with some comments on the status of verb movement in the analysis of verb-initial languages and a summary.

2 Theoretical background

2.1 Structure and order

The model proposed in Bury 2005 assumes that word order is not fully determined by phrase structure. Linearisation is only constrained by constituent boundaries. The same structure can then be linearised in multiple, but restricted, ways. For example, an unordered structure like *[Subj [Obj V]]* can be linearised in four different way, namely *Subj Obj V*, *Subj V Obj*, *Obj V Subj*, and *V Obj Subj*. The orders **V Subj Obj* and **Obj Subj V* are excluded because here the constituent containing *Obj* and *V* would be interrupted by *Subj*. This approach to linearisation is familiar, among others, from standard versions of X-bar theory from the mid-1980s onwards and from more recent work that does not adopt Kayne's antisymmetry hypothesis.[1]

For the discussion of verb-initial languages, two points should be noted. First, VOS order is compatible with any proposal that assumes that the verb forms a constituent with the object that excludes the subject. Second, VSO order can only be derived from such an underlying structure if something moves; a common analysis of VSO order involves movement of the verb to the left of the subject (see section 4 for a discussion of various alternatives).

A further important consequence of this approach to linearisation is that linear order does not necessarily affect the c-command relations in a structure. Thus, the different linearisations of *[Subj [Obj V]]* given above all correspond to the same structure in which *Subj* c-commands *Obj* – regardless of the order in which they are pronounced.

---

[1] Following Brody 2000, Bury 2005 assumes that there is no categorial projection. Consequently linearisation and other details differ somewhat from conventional approaches to constituent structure. See Bury 2003, Bury & Uchida 2007 for more discussion.

2.2 Verb movement and preverbal particles[2]

Bury 2005 assumes that clause structure is not universal and that as a consequence learnability imposes strict limits on the availability of empty clausal heads (cf. Iatridou 1990, among many others). While this does not mean that there can be no empty heads, it does imply that clausal heads will typically be associated with overt lexical material. For example, English modal verbs can be argued to form a paradigm that licenses an empty T head in clauses without modals (Koeneman 2000); in contrast, the complementizer *that* does not have an empty counterpart, i.e. English finite complement clauses without *that* are not CPs headed by an empty C (cf. Bury 2003).

Following Brody 2000, Bury 2005 assumes that head movement involves the pronounciation of a head's phonological matrix in the PF position of a different head. However, unlike Brody, Bury 2005 assumes that (leftward) head movement of H_1 to H_2, where both heads contain phonological material, is (usually) linearised as $H_2\ H_1$, not as $H_1\ H_2$, as it would on a standard head adjunction view.

Since, as discussed, VSO order cannot be generated without movement, Bury 2005 argues that these assumptions about clause structure and verb movement can be used to derive the generalisation that verb-initial languages typically have preverbal particles, at least for those languages in which the verb (head-)moves to the left: If clausal heads typically contain overt material, a head that is pronounced in the position of a different head (i.e. a head that is moved) will usually be adjacent to the phonological material of the other head. Thus, a moved verb will usually be pronounced next to the phonological material of the head targeted by the moved verb. Such a derivation of a VSO clause will then involve at least a constituent that contains subject, verb and object, a head H that occurs to the left of this constituent, and verb movement, i.e. the pronunciation of V in the position of H, and adjacent to any phonological material already contained in H:

(1) a. Structure (not linearised): [H [S [V O]]]
b. Pronunciation of (1a) H+V S O

[2] The assumptions discussed here are to some extent independent of the assumptions about linearisation just discussed. This means that the main arguments made below about the analysis of VSO-VOS alternations are also compatible with theories that make different assumptions about clause structure and verb movement. However, the connection between preverbal particles and verb-initial orders discussed here and below would likely be lost.

3 A possible analysis of VSO-VOS alternations

The previous section outlined a proposal that derives the generalisation that verb-initial languages (with VSO order) tend to have preverbal particles. This section suggests that the same model could also be useful for the analysis of languages in which both VSO and VOS clauses are found.

3.1 The structure of VSO and VOS

Recall the assumption that phrase structure does not determine word order. From this perspective, it is clear that the structure in (1a) above is not only compatible with VSO order but also with VOS. As noted earlier, a structure containing a subject and a constituent with a verb and an object can be linearised with the subject to the left or to the right of the verb-object constituent. There is no reason to assume that this freedom of linearisation should be affected by movement of the verb out of the subject-verb-object constituent. Consequently, (1a) could also be pronounced as (2):

(2) Alternative pronunciation of (1a): H+V O S

The combination of verb movement and flexible linearisation then offer a straightforward derivation for both VSO and VOS orders. Moreover, since both orders are linearisations of the same structure, this proposal implies that, unless there are independent constraints on word order in a given language, both VSO and VOS clauses should be possible in the same language. For ease of reference, I call this approach a FLEXIBLE LINEARISATION (FL) APPROACH.

3.2 When does word order flexibility surface?

Before considering the further implications of this type of analysis, it is worth to point out that constituent structure is not the only source for constraints on word order. Instead, we can think of the structural constraint on linearisation as imposing an upper limit on the possible linearisations of a given structure. Additional constraints may then impose further restrictions. For example, it has been argued that constraints on parsing explain the overwhelming preference for leftward movement across languages (cf. Abels and Neeleman 2006). If a moved constituent must precede its trace, the presence of a movement chain will further restrict the linearisation options for a structure. Moreover word order is often used, alongside other devices like agreement morphology and case marking, to mark grammatical functions. Thus, Welsh allows only VSO orders because the subject is marked under adjacency with the initial verb (cf. Sproat 1985). The more a language relies

on word order to mark grammatical functions, the less likely it will display effects of flexible linearisation.

3.3 Patterns predicted by the flexible linearisation model

We saw that the FL model can account for an alternation between VSO and VOS order in a simple way. Indeed, if there are no further word order constraints in a language, this pattern is expected to be the default in verb-initial languages with verb movement. Note that beside the VSO-VOS alternation and verb movement, a language that fits such an analysis is predicted to have three additional properties. First, since VSO and VOS are linearisations of the same structure, it is predicted that in either order, the subject c-commands the object. Second, since the analysis relies on verb movement, the language is predicted to have preverbal particles that could provide a landing site for verb movement. Finally, since flexible linearisation effects are expected to surface only in the absence of other word order restrictions, word order is expected not to play a major role in the marking of grammatical functions; therefore a rich agreement and / or case system is likely.

We'll now consider relevant data from St'át'imcets, Chamorro, and Tongan to see to what extent they are compatible with these predictions.

3.3.1 St'át'imcets (Lillooet Salish). St'át'imcets is a Salish language spoken in Southwest Interior British Columbia. Davis 2005 reports that St'át'imcets has two dialects, whose basic orders are VOS and VSO respectively, but that in both dialects the alternative order is also available. This pattern of flexible verb-initiality appears to be common among the Salish languages (cf. Kroeber 1999: 36–41) although there are exceptions like Bella Coola (H. Davis, p. c., Beck 2000). The following examples illustrate VOS and VSO clauses from the Lower Dialect of St'át'imcets.

(3) (Examples from Davis 2005: 36)

a. ts‘aw‘-an(-Ø)-as=ha ti=snúk‘wa7-sw=a
wash-DIR(3OBJ)-3ERG=ynq DET=friend-2SG.POSS=EXIS
ti=káoh-sw=a
DET=car-2SG.POSS=EXIS
‘Did your friend wash your car?’ (preferred)

b. ts‘aw‘-an(-Ø)-as=ha ti=káoh-sw=a
wash-DIR(3OBJ)-3ERG=YNQ DET=car-2SG.POSS=EXIS
ti=snúk‘wa7-sw=a
DET=friend-2SG.POSS=EXIS
‘Did your friend wash your car?’ (possible)

Moreover, Davis describes some binding facts that demonstrate that the subject c-commands the object in both VSO and VOS clauses:

(4) (Examples from Davis 2005: 40)

a. wa7 xwey-s-twitas i=kwekw7-i=ha
IMPF dear-CAUS-3PL.ERG PL.DET=grandmother-3PL.POSS=EXIS
takem i=sqaycw=a
all PL.DET=man=EXIS
(i) 'All the men$_i$ love their$_{i/j}$ grandmothers.'
(ii) 'Their$_{*i/j}$ grandmothers love all the men$_i$.'

b. wa7 xwey-s-twitas takem i=sqaycw=a
IMPF dear-CAUS-3PL.ERG all PL.DET=man=EXIS
i=kwekw7-i=ha
PL.DET=grandmother-3PL.POSS=EXIS
(i) 'All the men$_i$ love their$_{i/j}$ grandmothers.'
(ii) 'Their$_{*i/j}$ grandmothers love all the men$_i$.'

St'át'imcets has a range of aspectual markers (cf. Davis 2004) like the imperfective particle above that could possibly be argued to function as landing sites for verb movement in terms of the analysis suggested here –but see the comments on verb movement at the end of this subsection.

Finally, Davis (2005: 33) describes St'át'imcets as a 'radically head-marking' language: 'Arguments of a predicate (subject and primary object) are obligatorily marked on the head by agreement morphology, in the form of either clitics or affixes, sometimes null.'

Thus, St'át'imcets has most of the properties that the model outlined above predicts. However, the situation is less clear with regard to verb movement in the Salish languages. While Wiltschko (2003: 678) argues that 'there is significant evidence that verbs in Salish undergo movement to a functional head position which is at least higher than *v*P,' Davis 2004 argues on the basis of VP ellipsis data that there is no verb movement in St'at'imcets.

3.3.2 Chamorro. Chamorro is an Austronesian language spoken in the Mariana Islands in the Western Pacific (Chung 1998). Like St'át'imcets, Chamorro allows both VSO and VOS clauses although 'when the clause contains a transitive verb followed by two noun phrases, either of which would be syntactically licensed and pragmatically plausible as the subject, then for most speakers the VSO interpretation is forced' (Chung 1998: 22).

(5) (Examples from Chung 1998: 150f)

a. Ha-pula' i näna$_i$ i patgon-ň *pro*$_i$.
AGR-undress the mother the child-AGR
'The mother$_i$ undressed her$_i$ child.'
b. Ha-pula' i patgon-ň *pro*$_i$ i näna$_i$.
AGR-undress the child-AGR the mother
'The mother$_i$ undressed her$_i$ child.'

These examples and the following ones also illustrate that the subject c-commands the object in either order.

(6) (Examples from Chung 1998: 150f)

a. Mämaigu' käda patgun$_i$ gi mismu kattre-ň *pro*$_i$.
AGR.sleep.PROG each child LOC same bed-AGR
'Each child$_i$ is sleeping in his$_i$ own bed.'
b. ?Mämaigu' gi mismu kattre-ň *pro*$_i$ käda patgun$_i$.
AGR.sleep.PROG LOC same bed-AGR each child
'Each child$_i$ is sleeping in his$_i$ own bed.'

Chung (1998: 151f) reports that in (6b) the VOS order is only accepted by some speakers and judged marginal or ungrammatical by others. She explains this contrast in terms of a constraint that requires a quantified noun phrase to precede a pronoun that it binds.

Chamorro has a complex agreement system and makes extensive use of null anaphora (Chung 1998: 26–32). Moreover, Chamorro has a class of elements expressing finitess and tense-mood-aspect that Chung (1998: 25) argues realise I(nflection). At first, it appears as if these elements could serve as the possible landing sites for verb movement. However, these inflectional elements do not have to be adjacent to the verb. This is illustrated below where the future marker *pära* is separated from the verb by the adverb *käna* 'almost':

(7) (Example from Chung 1998: 131)

Pära käna ha‘ u-matmus, maolik na hu-li‘I‘.
FUT almost EMP AGR-drown good COMP AGR-see
‘He'd been about to almost drown; good thing I saw him.’

It is not clear how this pattern could be understood in terms of the model proposed above.

While some of the other arguments that Chung advances against verb movement in Chamorro do not apply to the model outlined here because they only apply to proposals that assume a fixed underlying SVO order, there does not appear to be any obvious independent support for verb movement here.

3.3.3 Tongan. Tongan is a Polynesian of the South Pacific with an ergative case system for full noun phrases and an accusative system for pronouns. According to Otsuka (2005: 73f), the unmarked word order in Tongan is VSO but VOS order is also possible, and ‘native speakers generally do not recognize any semantic difference between the two when used in isolation.’ The two orders are illustrated here:

(8) (Examples from Otsuka 2005: 73)

a. Na‘e kai ‘a e ika ‘e Sione.
PST eat ABS the fish ERG Sione
‘Sione are the fish.’
b. Na‘e fili ‘a Pila ‘e Sione.
PST choose ABS Pila ERG Sione
‘Sione chose Pila.’

Otsuka 2006 argues that the alternation is a result of scrambling of the object from a (derived) VSO structure. She proposes that movement of the object to the specifier of TP is triggered by an EPP feature and an information focus feature and provides a number of arguments to show that VOS structures have properties of A-movement rather than A-bar movement structures (see also section 4). However, it seems that the data she discusses are also compatible with a FL analysis. For example, while Otsuka suggests that the lack of weak crossover effects in VOS structures shows that the order is not derived by A-bar movement, the grammaticality of the following example is also compatible with an analysis in which the object has not undergone movement:

(9) (All following examples are from Otsuka 2006: 250ff.)

Na‘e fili ‘a e taha kotoa$_i$ ‘e he‘ene$_i$ tamai.
PST choose ABS DEF one every ERG his father
‘His$_i$ father chose everyone$_i$.’

Tongan doesn't have a specific set of reflexive pronouns but the same form can be used as either pronominal or anaphoric. For example, here the ‘the third person singular pronoun *ia* can either be coreferential with or disjoint from the subject *Sione*’(Otsuka 2006: 251).

(10) Na‘e fili ‘e Sione$_i$ ‘a ia$_{i/j}$ pē.
PST choose ERG Sione ABS 3.S only
‘Sione chose him/himself.’

However the interpretation of pronouns is not free. Otsuka provides the following examples:

(11) a. Na‘e fili ‘a ia$_i$ pē ‘e Sione $_{*i/j}$ t$_i$.
PST choose ABS 3.S only ERG Sione
‘Sione chose him/*himself.’
b. Na‘e fili ‘a Sione$_i$ ‘e ia$_{*i/j}$ pē t$_i$.
PST choose ABS Sione ERG 3.S only
‘He/*himself chose Sione.’

(11a) shows that a pronominal object cannot be corefential with the subject in a VOS clause. (11b) shows that a pronominal subject cannot be coreferential with a full noun phrase object in a VOS clause. As Otsuka notes, this second fact is surprising if VOS clauses involve A-movement of the object because after this movement the object should c-command the subject. These facts could be captured if the ungrammaticality of the bound interpretation in (11b) is not due to a c-command problem but rather due to a precedence constraint on bound pronouns of the type proposed by Chung for Chamorro (see above). Given such a constraint, the pattern illustrated in (11) would seem compatible both with Otsuka’s A-scrambling analysis as well as with a FL analysis.

In Tongan, ‘tense and aspect are indicated by independent lexical items, which immediately precede the verb’ (Otsuka 2000: 49). These tense and aspect particles

look as if they could be used as landing sites for verb movement.[3] Finally, Otsuka assumes that there is verb movement from V-to-T-to-C in Togan. However, since her arguments for verb movement are based on the assumption of an underlying SVO order, it is not clear to what extent they carry over to an FL approach.

3.3.4 Summary. This brief survey indicates an interesting match between the predictions of the FL model and the properties of the languages discussed –leaving aside for the moment the issue of verb movement (but see section 5). Thus, St'át'imcets, Chamorro and Tongan all allow both VSO and VOS orders. Binding patterns indicate that the subject c-commands the object in both orders in St'át'imcets and Chamorro, and the Tongan data in (11) seems to be at least compatible with such an analysis as well. The three languages all have some type of preverbal particles that look like potential landing sites for verb movement and they also have some rich system of agreement or case morphology which means that the marking of grammatical functions need not rely on word order.

While I hope that this discussion shows that the FL model may at least have the potential to make a contribution to the analysis of VSO-VOS languages, the selected data discussed here should of course be seen in the broader context that led Davis, Chung, and Otsuka to develop rather different analyses of St'át'imcets, Chamorro and Tongan, respectively. A detailed discussion of these different proposals is beyond the scope of this paper, but the next section provides a (necessarily brief) look at them and argues that they all differ from the FL proposal in one significant way.

4 Alternative analyses of VSO-VOS alternations

The previous section discussed the language profile predicted by a FL analysis of VSO-VOS alternations and explored to what extent the predictions materialise in St'át'imcets, Chamorro and Tongan. In this section, I briefly summarise the analyses of these languages provided by the authors whose data I used. We will see that Davis, Chung and Otsuka provide very different analyses for St'át'imcets, Chamorro, and Tongan respectively. However, these alternative analyses all differ from the FL approach outlined here in one significant: they all assume that one of the two orders (VSO and VOS) has a more basic derivation or structure, and that the other order is derived through a special rule.

Davis 2004, 2005. Davis argues that the 'postpredicative word order alternations in St'át'imcets show the same profile as "extraposition" dependencies in more

[3] While clitics can appear between the tense/aspect particle and the verb, 'a clitic pronoun and a tense marker form a single phonological unit' (2005:72); so it is not clear whether this would be a problem for the type of analysis proposed here.

familiar languages like Dutch and English' (2005: 55). Moreover, he provides evidence, such as the binding facts discussed above, for the existence of a VP constituent that contains the verb and its internal argument but that excludes the subject and he shows that the alternation is sensitive to prosodic weight. Davis (2005: 56) concludes that the alternation must be dealt with by the PF side rather then the LF side of the grammar but he notes that there currently is no theory on the market that captures the St'át'imcets pattern, or certain generalisations about right-peripheral linearisation more generally. Davis 2005 doesn't provide any structures to illustrate his assumptions about the St'át'imcets VSO-VOS alternation but he seems to assume that in the syntax, there is a constituent that contains the verb and the object and that the subject sits in a rightward specifier position (cf. Davis 2004). Thus, in his analysis St'át'imcets is VOS in (pre-PF) syntax and VSO order is derived by some form of PF extraposition of the object.

(12) *Davis on St'át'imcets* VOS order: [[V O] S]
VSO order: [[V ~~O~~] S] O (extraposition at PF)

While both Davis' analysis and the proposal made here invoke the syntax-PF interface, they make use of rather different processes. In the flexible linearisation proposal, syntax underspecifies word order and VSO and VOS are both orders compatible with the constituency defined by the same structure; in contrast, Davis' extraposition leads to a discontinuous linearisation of the VP constituent, which is of course similar to what may happen with extraposition in English and elsewhere.

Chung 1998, 2006. Like Davis for St'át'imcets, Chung assumes that Chamorro has a basic VOS structure with the subject sitting in the highest specifier (and that this is a rightward specifier). However she provides evidence for an analysis in which VSO order is derived by (syntactic) lowering of the subject into the constituent that contains the verb and the object. In these structures, the lowered subject is coindexed with a null pronominal in the original subject position (1998: 169):

(13) *Chung on Chamorro* VOS order: [[V O] S]
VSO order: [[V S_i O] pro_i] (syntactic lowering)

Chung 2006 notes that many of the patterns she discusses in the earlier work 'could equally be produced via linearisation' (2006: 714).

Otsuka 2005, 2006. Finally, Otsuka argues that Tongan VSO order is derived by leftward (head-)movement of the verb from an underlying SVO constituent. She assumes that in VSO clauses the subject occupies a derived position, the specifier

of TP. In VOS clauses, the subject remains in its base position and instead the object A-scrambles into the specifier of TP, triggered by an information focus feature. (This option becomes available because, following verb movement to T, the base positions of the subject and the object are equidistant from the specifier of TP (Otsuka 2006: 254).)

(14) *Otsuka on Tongan*

VSO order:	V_j [S_i [t_i t_j O]]	
VOS order:	V_j [O_i [S t_j t_i]]	(scrambling)

As we can see, these are very different ways of deriving an alternation between VSO and VOS orders. However, what unites them is that they all assume that one order is basic –VOS for Chung and Davis, VSO for Otsuka– and that the other order is derived by a special rule.[4] This is a significant property of these analyses since it implies that the grammar of a language that allows both VSO and VOS orders is more complex than the grammar of a language that allows only one of the two orders. In other words, the syntax of a language with a rigid VSO or VOS order would require one rule or one derivational step less. In this respect, these analyses collectively differ from the FL proposal, where an alternation between VSO and VOS orders is expected in a verb-initial language, unless word order, more specifically the position of the subject and/or the object, has some grammatical role.[5] Arguably then, the FL approach (with verb movement) and approaches that assume a linearised syntax make different predictions with regard to the expected relative frequencies of verb-initial languages with a fixed VSO or VOS system as opposed to languages with VSO-VOS alternation.[6]

---

[4] A further possibility is offered by VP-preposing analyses (cf. Massam 2000, Rackowski and Travis 2000 and the critical overview in Chung 2006): VOS order can be derived from an underlying SVO constituent through preposing of the VP, which contains the verb and the object. VSO order can be derived in a similar way, if the object leaves VP before VP fronts. The additional step of object shift in the VSO derivation also constitutes such a special rule.

Davis 2005 suggests that the St'át'imcets pattern may fall out from a general theory of right-peripheral linearisation. Given such a theory, it could be argued that object extraposition in St'át'imcets does not require a special rule. However, as Davis notes, currently there is no such theory.

[5] Recall that St'át'imcets has two dialects that differ with respect to which verb-initial order is the preferred order. Since in the FL proposal VSO and VOS clauses have the same structure but in Davis' analysis VSO order involves extraposition, this dialect split may arguably be less surprising given an FL analysis.

[6] However see Newmeyer 2005 who argues that such statistical generalisations should not be explained in terms of competence theories.

5 Conclusion

5.1 Verb movement in verb-initial languages

Before concluding this paper, I outline a number of possible directions of future work that relate to the status of verb movement in the analysis of verb-initial languages. First, the structure proposed in section 3 that may derive some VSO-VOS alternations crucially assumes that the initial verb is moved to a position outside of the constituent that contains the subject and the object (and the trace position of the verb). As the discussion above showed, it is not so easy to find clear evidence for verb movement in VSO-VOS languages (cf. Gärtner et al (2006: 9) on Austronesian, a language family with a significant number of VSO-VOS languages). Here I'd just like to suggest that there may be a principled reason for why verb movement in these languages is difficult to spot. Recall that, assuming flexible linearisation, VOS order can be derived without any movements. This means that VSO-VOS languages could have some verb-initial structures, namely some VOS orders, that do not involve verb movement. This would then make verb movement a string-vacuous process and consequently hard to detect. The problem is comparable to the status of rightward verb movement in verb-final languages and it may be that the arguments and diagnostics that have been proposed there can lead to some progress in the verb-initial context as well (cf. Vermeulen 2006 on verb movement in Japanese).

Second, as noted earlier, there is a class of derivations of verb-initial clauses that assumes neither head movement of the verb nor base-generated verb-initial orders. Verb-initial order can also be derived through the fronting of a phrasal constituent from which the subject and possibly the object have been removed. As Chung (2006) argues, this type of analysis may have a strong motivation for some verb-initial languages and consequently a discussion of such VP-preposing languages in the context of flexible linearisation is of great interest.

Finally, there has been no discussion here of languages with rigid VOS order (some of which may be VP-preposing languages). Since VOS order can in principle be derived without any verb movement, a detailed comparison of VOS languages with languages that allow both VSO and VOS could potentially bring to light asymmetries between the two that could bear on the question of verb movement in these languages.

5.2 Summary

Section 1 of this paper outlined an approach to syntax (in a broad sense) in which word order is not fully determined by structural hierarchies. Section 2 reviewed the relevant features of Bury 2005 and showed how this model can derive the generalisation that verb-initial languages tend to have preverbal particles. Section 3

argued that this model provides a simple analysis for a word pattern in which VSO and VOS orders alternate, a pattern that is found in many verb-initial languages. This analysis also predicts such languages to have certain properties and a brief look at St'át'imcets, Chamorro and Tongan suggests that these VSO-VOS languages seem largely compatible with at least the predictions discussed here. Section 4 summarised a number of earlier analyses of languages with a VSO-VOS alternation and concluded that the FL model proposed here is the only one that doesn't assume a special rule to derive the alternation.

References

Abels, K. & A. Neeleman. 2006. Universal 20 without the LCA. Ms., University of Tromsø and University College London.

Beck, D. 2000. Grammatical convergence and the genesis of diversity in the Northwest Coast *Sprachbund*. *Anthropological Linguistics* 42, 1–67.

Brody, M. 2000. Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry* 31, 29–56.

Bury, D. 2003. *Phrase structure and derived heads*. PhD dissertation, UCL.

Bury, D. 2005. Preverbal particles in verb-initial languages. In Carnie, Harley & Dooley (eds.), 135-154.

Bury, D. & H. Uchida. 2007. Constituent structure sets. Ms., University of Wales, Bangor and University College London.

Carnie, A. & E. Guilfoyle (eds). 2000. *The syntax of verb-initial languages*. Oxford: Oxford University Press.

Carnie, A, H. Harley & S. Dooley (eds.). 2005. *Verb first: On the syntax of verb-initial languages*. Amsterdam: John Benjamins.

Chung, S. 1998. *The design of agreement: Evidence from Chamorro*. Chicago and London: University of Chicago Press.

Chung, S. 2006. Properties VOS languages. In Everaert, van Riemsdijk, Goedemans & Hollebrandse (eds.), 685-720.

Davis, H. 2004. VP-ellipsis in St'a´t'imcets and its implications. *Papers for the 39th International Conference on Salish and Neighbouring Languages*, 117–140. Vancouver: UPBWPL.

Davis, H. 2005. Coordination and constituency in St'át'imcets (Lillooet Salish). In Carnie, Harley & Dooley (eds.), 31–64.

Everaert, H. van Riemsdijk, R. Goedemans & B. Hollebrandse. 2006. *The Blackwell companion to syntax*. Oxford: Blackwell.

Gärtner, H. P. Law & J. Sabel (eds.). 2006. *Clause structure and adjuncts in Austronesian languages*. Mouton de Gruyter.

Iatridou, S. 1990. About AgrP. *Linguistic Inquiry* 21, 551–577.

Koeneman, O. 2000. *The flexible nature of verb movement*. PhD dissertation, Utrecht University.

Kroeber, P. 1999. *The Salish language family: Reconstructing syntax*. Lincoln: University of Nebraska Press.

Massam, D. 2000. VSO and VOS: Aspects of Niuean word order. In Carnie & E. Guilfoyle (eds), 97–116.

Newmeyer, F. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Cambridge, MA: MIT Press.
Otsuka, Y. 2006. Scrambling and information focus: VSO-VOS alternation in Tongan. In Sabel and Saito (eds.), 243-279.
Otsuka, Y. 2005. Two derivations of VSO: A comparative study of Niuean and Tongan. In Carnie, Dooley & Harley (eds.), 65-90.
Rackowski, A. and L. Travis. 2000. V-initial languages: X or XP movement and adverb placement. In Carnie & Guilfoyle (eds.), 117–141.
Sabel, J. & M. Saito (eds.). *The free word order phenomenon: Its syntactic sources and diversity*. Berlin: Mouton de Gruyter.
Sproat, R. 1985. Welsh syntax and VSO structure. *Natural Language and Linguistic Theory* 3, 173–216.
Vermeulen, R. 2006. Case and coordination in Japanese. *West Coast Conference on Formal Linguistics* (WCCFL) 25, 417-425.
Wiltschko, M. 2003. On the interpretability of Tense on D and its consequences for Case Theory *Lingua* 113, 659–696.

*On the prosody and syntax of DPs: Evidence from Italian noun adjective sequences**

NICOLE DEHÉ AND VIERI SAMEK-LODOVICI

Abstract

The opposition between N- and NP-raising is central to the debate about the representation of DPs, yet it often eludes syntactic testing. The two hypotheses are however distinguished by the prosodic phrasing they predict. This paper presents the results of an experiment designed to test the prosodic phrasing of Italian N-A and A-N sequences as signaled by the lengthening effects induced by prosodic boundaries. We show that A and N share the same phonological phrase and that under all models of syntax prosody mapping the attested phrasing requires N-raising. Finally, we propose an analysis reconciling N-raising with Cinque's recent evidence for DP-internal phrasal movement.

1 Introduction

The last fifteen years have seen a still on-going debate about the internal syntactic representation of DPs with a particular focus on Romance N-A sequences. Since Cinque's seminal 1994 study identifying A-N as the base generated order for both Romance and Germanic languages, two main hypotheses have emerged to explain the mirror N-A order found in Romance. The N-raising hypothesis maintains that N raises as a head to the left of a preceding AP projection (see among others Sproat and Shih 1990; Crisma 1991; Valois 1991; Bernstein 1991, 1993; Cinque 1994; Zamparelli 1995; Longobardi 1994, 2001, 2005; Kishimoto 2000; Rutkowski and Progovac 2006; Willis 2006; Pereltsvaig 2006; Airtiagoitia 2006). The NP-raising hypothesis instead proposes that what moves is the entire NP (e.g., Bhattacharya 1998; Laenzlinger 2000; Alexiadiaou 2001; Shlonsky 2004; Knittel 2005; Cinque 2005, 2006). Within a structure à la Cinque (2005) where AP occurs in the specifier

* An earlier version of this paper was presented at the Prosody-Syntax Interface Workshop at University College London in October 2006 and we would like to thank all the participants for the lively discussion. We are also particularly grateful to Guglielmo Cinque, Laura Downing, Brechtje Post and Lisa Selkirk for discussion of and comments on aspects of the work presented here. Special thanks go to Sam Hellmuth and Malte Zimmermann for their efficient help with finding informants in the Berlin/Potsdam area, and to the participants of the experimental study.

of a functional projection FP dominated by an agreement phrase as shown in (1), the N-raising analysis requires N moving into Agr° through F° as shown in (2), whereas the NP-raising alternative moves NP into SpecAgrP as shown in (3).

(1) Base-generated structure: $[_{AgrP}\ Agr^{o}\ [_{FP}\ AP\ F^{o}\ [_{NP}\ N\ ]]]$

(2) N-raising: $[_{AgrP}\ N_i\ [_{FP}\ AP\ t_i\ [_{NP}\ t_i\ ]]]$

(3) NP-raising: $[_{AgrP}\ NP_i\ Agr^{o}\ [_{FP}\ AP\ F^{o}\ t_i\ ]]$

The two alternative hypotheses are notoriously difficult to tell apart by syntactic tests alone. They are however neatly distinguished by the prosodic phrasing they predict. As we will discuss in detail, all current major models of the syntax-prosody mapping –e.g. Nespor and Vogel (1986), Selkirk (1986, 2000), Ghini (1993), Truckenbrodt (1995, 1999)– predict that an N-A sequence will be parsed into a single shared phonological phrase like (4a) under N-raising, whereas it would require two distinct phrases as shown in (4b) under NP-raising.

(4) a. $(\ N_i\ AP\ )_{pp}$

b. $(\ NP_i\ )_{pp}\ (\ AP\ )_{pp}$

In this paper we present the results of an experimental reading study testing the prosodic phrasing of these sequences, thus aiming at resolving the opposition in (4). As a robust prosodic cue, the experiment tested syllabic and word lengthening induced by phonological phrase boundaries (Nespor and Vogel 1986; Beckman and Edwards 1987, 1990, 1991; Hayes 1988; Wightman et al. 1992; Ghini 1993; Cruttenden 1997; Post 2000; Vaissière 1983; among others). It uncovered a statistically highly significant lengthening of the second word of A-N and N-A sequences in its entirety, and of its final and lexically stressed syllables (shown in bold in (5) below). These results show that adjective and noun are wrapped in a single phonological phrase (pp) independently of their order, as symbolized by the round parentheses in the examples in (5).

(5) a. (pre.la.to po.**ten**.**te**)$_{pp}$
prelate powerful

b. (po.ten.te pre.**la**.**to**)$_{pp}$
powerful prelate

Our experimental results provide strong independent support for the N-raising hypothesis and the corresponding syntactic representation in (2) above. Yet this outcome may at first appear surprising, since it is at odds with convincing arguments against generalized N-raising and in support of DP-internal phrasal movement in Cinque (2005, 2006). As we show in the second part of this study, closer examination of the syntax of N-raising reveals that it is limited to the closest AgrP and that any further raising of the noun requires pipe-piping of the entire AgrP consistently with Cinque's empirical and theoretical results. Head and phrase raising co-exist, but target distinct syntactic categories within DP.

The implications of the prosodic phrasing attested in our experiment also extend to the analysis of Italian overt subjects, which will be shown to occur in a higher projection than the one hosting raised finite verbs, thus converging with the results in Barbosa (1995), Alexiadou and Anagnostopoulou (1998), Rizzi (2004), Cardinaletti (2004), Frascarelli (to appear). On the prosodic side, we will show that only Truckenbrodt's (1995, 1999) model can consistently extend the syntax-prosody mapping found for N-A sequences to A-N sequences as well.

We start in section 2 with a description of the experiment and its results relative to the prosodic phrasing of A-N and N-A sequences. The syntactic implications of this prosodic phrasing are discussed in section 3, where we show how current models of syntax-prosody mapping require an N-raising representation of A-N and N-A sequences as well as a higher syntactic position than normally assumed for Italian overt subjects. Finally in section 4 we examine Cinque's arguments against generalized N-raising, first showing that they do not exclude local N-raising of the kind advocated in this paper and then arguing for an analysis where N-raising and phrasal raising of functional projection à la Cinque co-exist, with their application governed by the principles responsible for movement locality proposed in Cinque (2005, 2006).

2 The prosodic phrasing of Italian A-N and N-A sequences: an experimental study

The experiment described in this section was designed to test the prosodic phrasing of Italian N-A and A-N sequences. To this end, we investigated domain-final lengthening effects, which have been established as a robust cue to prosodic boundaries in much previous research. As Vaissière (1983, p. 61) points out, there are four kinds of lengthening phenomena that mark the right boundary of a word or phrase: (i) lengthening of the very last syllable of a word or phrase, (ii) lengthening of the last stressed syllable in a phrase, (iii) lengthening of the entire last word in a phrase, and (iv) lengthening of the last sentence in a read paragraph. The first three kinds of lengthening are relevant to our study and were tested for Italian A-N/N-A

sequences. As for the first (lengthening of the very last syllable of a prosodic domain), Cruttenden (1997, p. 33) points out that it is a productive prosodic process attested in most languages and likely to constitute a universal property of prosodic phrasing (see also Hayes 1988; Wightman et al. 1992; Beckman and Edwards 1987, 1991; Post 2000). The syllable immediately preceding a prosodic boundary is lengthened regardless of whether it carries the lexical stress of the word (Cruttenden 1997, p. 33; Beckman and Edwards 1987, 1990, 1991). This is illustrated in (6) where the final syllable of a hypothetical three-syllabic word is lengthened due to the phonological-phrase boundary immediately following it; the affected syllable is shown in bold.

(6) Pre-boundary lengthening: … σ.σ.σ́.σ σ.σ́.**σ:**)$_{pp}$ (σ.σ́.σ …

The second lengthening process affects the syllable carrying lexical stress in the word immediately preceding the prosodic boundary. For example, a three-syllable word with penultimate lexical stress like the one in (7) below would have its penultimate syllable (in bold) further lengthened due to the following boundary. Lengthening in this case is arguably caused by the prosodic head of the pp, which in Italian always falls on the rightmost word of the pp, thus adding additional prosodic prominence to its stressed syllable (Selkirk 1984, 1986, 1995, 2000, 2004; Hammond 1984; Halle and Vergnaud 1987; Edwards and Beckman 1988; Hayes 1995; Truckenbrodt 1995). An Italian example from Ghini (1993) showing vocalic lengthening in the stressed syllables of the pp-final words is provided in (8).

(7) Stressed-syllable lengthening: … σ.σ.σ́.σ σ.**σ́:**.σ)$_{pp}$ (σ.σ́.σ …

(8) (i ca.ri.bú n[a:].ni)$_{pp}$ (sono es.t[i:]n.ti)$_{pp}$
the caribous dwarf are extinct
'Dwarf caribou have been extinguished'

Finally, the last word of a prosodic domain has also been observed to lengthen (Umeda and Quinn 1981; Hellmuth, to appear). For Italian, Nespor and Vogel (1986, p. 176) observe that pp-boundaries induce lengthening of the domain-final word, an effect possibly simply measuring the cumulative effect of the two lengthening processes introduced above. For example, they note how the word 'pasticcini' is longer when it occurs pp-finally in (9b) than when occurring pp-internally as in (9a).

(9) a. Ho mangiato (dei pasticcini ripieni)$_{pp}$
(I) have eaten some donuts filled
'I have eaten some filled donuts'

b. Ho mangiato (dei pasticcini)$_{pp}$ (ripieni)$_{pp}$ (di cioccolata)$_{pp}$
(I) have eaten some donuts filled of chocolate
'I have eaten some chocolate-filled donuts'

Together, the above lengthening processes provide a powerful tool for determining the prosodic phrasing of Italian N-A and A-N sequences. Consider, for example, the sequences in (10) below, where lexical stress falls on the penultimate syllable in both words (the stressed syllable is shown in bold).

(10) A-N: po.**ten**.te pre.**la**.to
powerful prelate
N-A: pre.**la**.to po.**ten**.te

If both sequences are wrapped into a single pp ending at the right edge of the sequence as in (11), then lengthening will only affect the final word (in bold) and its last two syllables (in capitals). Crucially, the duration of the affected syllables and the word itself is predicted to change according to its position in the sequence, with increased length expected when the word occurs second, i.e. immediately preceding the pp boundary. For example, under the phrasing shown below, the noun 'prelato' and the syllables 'la' and 'to' in it are predicted to be longer under the A-N order than the N-A one. Due to the inherent symmetry of the example, the same holds for the adjective and its final syllables, which are predicted to be longer in N-A sequences.

(11) A-N: ... po. ten.te **pre.LA.TO**)$_{pp}$ (…
N-A: ... pre.la.to **po.TEN.TE**)$_{pp}$ (…

The distribution of word and syllabic lengthening just illustrated is distinctive of the above prosodic phrasing. Let us call it its 'length signature'. Any other conceivable prosodic phrasing has a different length signature. Consider for example the three symmetric phrasings in (12)-(14) below, which together with (11) above exhaust all possible symmetric phrasings. If the sequences are parsed into a single pp but with no pp-boundary following the last word, as in (12), then no lengthening occurs and the relative length of the relevant syllables and words remains constant independently of sequence order; for example, the syllables 'la' and 'to' of the noun 'prelato' would remain equally long across the A-N and N-A orders.

If on the other hand A and N are phrased into distinct pps each preceding a pp-boundary as in (13) then the relevant syllables and word are lengthened across the board, again predicting equal length independently of the sequence order. Finally,

if only the first word immediately precedes a pp-boundary, as in (14), then lengthening only occurs in sequence-initial position. None of these cases thus predicts the increased length in sequence final position associated with (11) above.

(12) A-N: (... po.ten.te pre.la.to ...)$_{pp}$
N-A: (... pre.la.to po.ten.te ...)$_{pp}$

(13) A-N: ... **po.TEN.TE**)$_{pp}$ (**pre.LA.TO**)$_{pp}$
N-A: ... **pre.LA.TO**)$_{pp}$ (**po.TEN.TE**)$_{pp}$

(14) A-N: ... **po.TEN.TE**)$_{pp}$ (pre.la.to ...
N-A: ... **pre.LA.TO**)$_{pp}$ (po.ten.te ...

Many additional possible phrasings are conceivable once we allow for asymmetric phrasing across the two sequences. As the reader may easily verify none of them matches the length signature of phrasing (11) since all mixed cases necessarily involve one of the phrasing described in (12)-(14) for at least one of the A-N and N-A sequences. Consequently, they all predict an asymmetric lengthening distribution once again distinct from the length signature of (11).

In our experiment, we measured the length signature of a set of A-N and N-A sequences embedded in carrier sentences. The results closely match the prosodic phrasing in (11). The details of the experiment are described below.

2.1 Materials

The experimental materials involved 5 adjective-noun pairs, presented under A-N and N-A order. The 5 pairs were contained in 20 carrier sentences, 10 of which were constructed such that the target sequence was the subject of the sentence (cf. (15) and (16)), while the other 10 were constructed such that the target sentence was the object of the sentence (cf. (17) and (18)). For each subject and object position, 5 sentences represented the A-N order (exemplified in (15) and (17)), while the other 5 represented the N-A order (as in (16) and (18) below). Examples (15)-(18) show the four sentence types for one adjective-noun pair. The full experimental materials are provided in the appendix at the end of the article.

(15) Un **po.ten.te pre.la.to** può imporre il suo punto di vista anche al papa
A powerful prelate can impose the his point of view even to-the pope

(16) Un **pre.la.to po.ten.te** può imporre il suo punto di vista anche al papa
A prelate powerful can impose the his point of view even to-the pope

(17) Abbiamo contattato un **po.ten.te pre.la.to** con il permesso del papa
(We) have contacted a powerful prelate with the permission of-the pope

(18) Abbiamo contattato un **pre.la.to po.ten.te** con il permesso del papa
(We) have contacted a powerful prelate with the permission of-the pope

In addition to these items, 35 sentences were part of the materials which were originally designed to test the prosodic phrasing of adverbs in VPs. They were part of another study and thus irrelevant to the present discussion. Moreover, the materials contained 46 filler sentences, adding up to 101 items overall. In the experimental design, these 101 items were pseudo-randomized under the usual restrictions.

2.2 Participant, apparatus and procedure

The experiment was carried out with 12 untrained native speakers of Italian (8 female, 4 male). At the time of the testing, they were unaware of the aim of the study. The target utterances were shown individually on a computer screen, using Microsoft PowerPoint. The participants were instructed to familiarize themselves with each sentence, read it out loud, and then move on to the next sentence. They were asked to produce each sentence as naturally as possible at a normal speech rate. The list of target items was preceded by five practice items to familiarize the participants with the procedure. All utterances were recorded to a Samsung laptop computer using an AKG C444 headset microphone with AKG B29L battery power supply and *Cool Edit*^{TM}*96* software. The recordings were later digitized into individual sound files using the same software. The individual sound files were analyzed in PRAAT (Boersma 2001).

2.3 Data treatment

Overall, the 12 speakers produced 240 target sentences: 60 sentences in each of the four data sets exemplified in (15) to (18) above. Of the 240 sentences, 15 contained speech errors and were discarded from the analysis. For each item that entered the analysis, the length of the two target words (A and N), and the length of the stressed and final syllables were measured, and the mean values calculated. The results were coded with respect to a) the order of A and N, and b) the syntactic function of the relevant constituent (subject vs. object).

2.4 Results

As outlined above, three kinds of final lengthening phenomena were addressed. Overall, the length of all target elements (final syllable, stressed syllable, full word) was clearly affected by the position of the respective word. They were longer when the respective word was second in its target sequence. Specifically, in both subject and object position, the following effects were recorded:

(i) The final syllable of N, the stressed syllable of N, and N itself were longer when N followed A than when N preceded A (cf. Figure 1).

(ii) The final syllable of A, the stressed syllable of A, and A itself were longer when A followed N than when A preceded N (cf. Figure 2).

The results for N and its stressed and final syllables are illustrated by the three panels in Figure 1 below. The first panel shows how the final syllable of N increases in length when N is final in either subject (first row) or object position (second row) than when N precedes A (third and fourth rows). The next two panels in figure 1 show the corresponding lengthening effects for the stressed syllable of N and for N itself according to the same layout. Figure 2 shows the corresponding lengthening effects for A in N-A sequences.

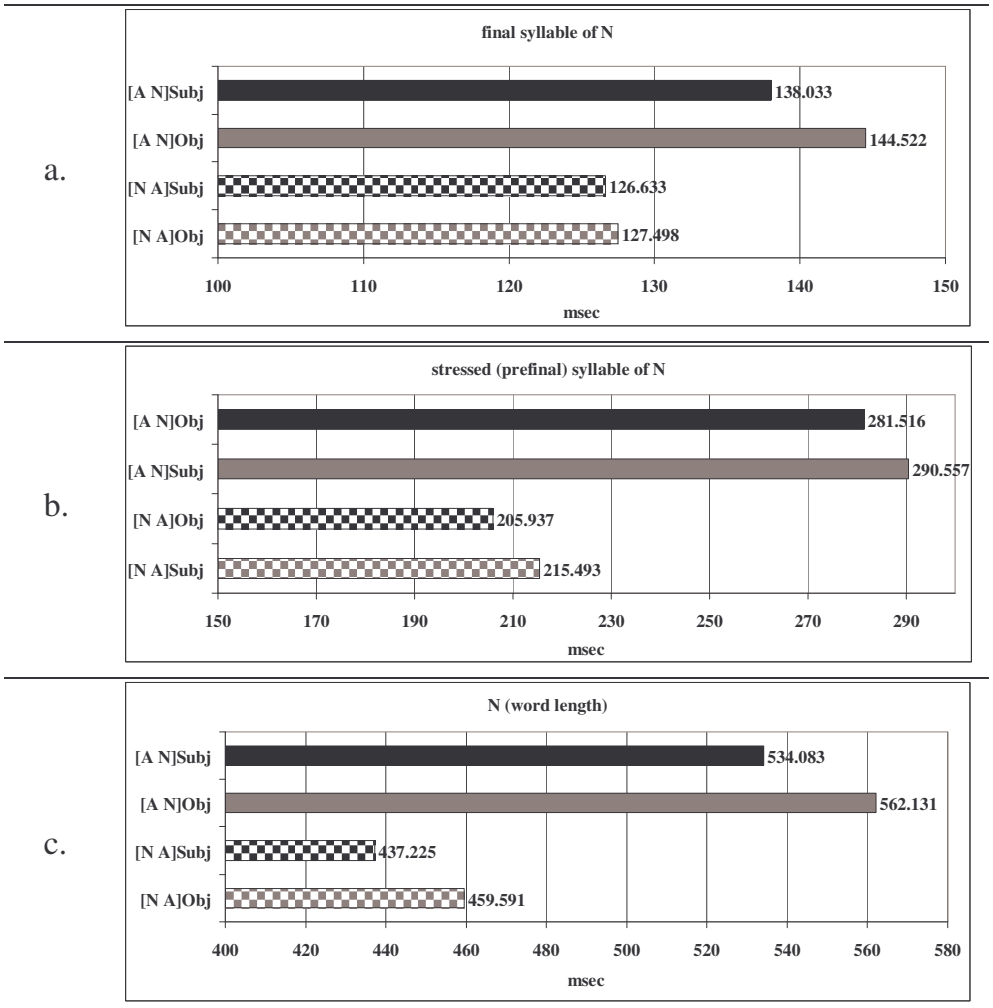


Figure 1 Length of the final syllable of N (panel a), the stressed syllable of N (panel b) and the full N (panel c) in the two word orders, in subject and object position

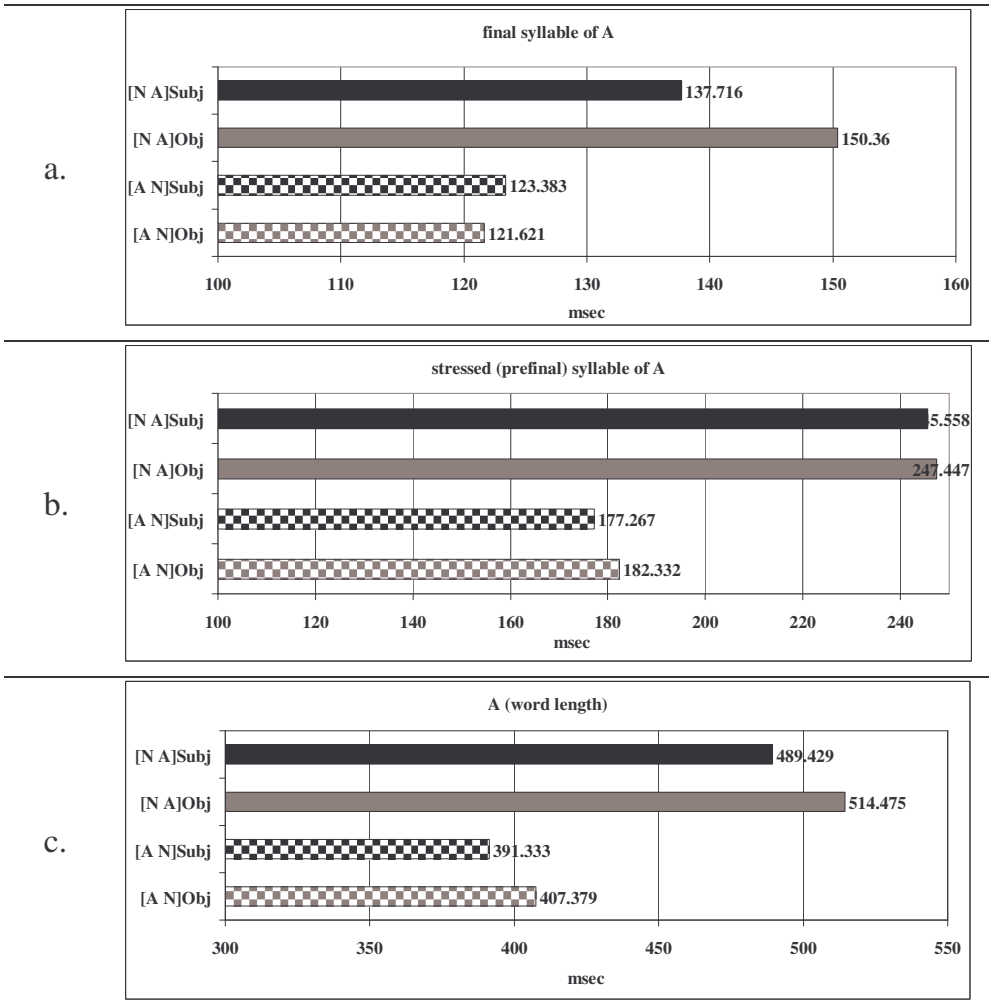


Figure 2 Length of the final syllable of A (panel a), the stressed syllable of A (panel b) and the full A (panel c) in the two word orders, in subject and object position

The duration values were submitted to an analysis of variance with the factors POSITION (2 levels: A-N/N-A) and SYNTACTIC FUNCTION (2 levels: subject vs. object). The scores were pooled over both speakers and items, and the α level determining the significance threshold was set at 0.05 (significant)/ 0.02 (highly significant). The most relevant results for the present study concern the factor POSITION. The variance analysis showed that the lengthening of A and N in their entirety was highly significant in both the item and speaker analyses. POSITION was also highly significant across item and speaker analyses for the stressed syllable of both A and N. As for the final syllable, POSITION was highly significant in the speaker analysis for A and in the item analysis for N.

The SYNTACTIC FUNCTION factor –less relevant to the goals of this paper– was found significant only in the subject analyses and only for full A and N and for the stressed syllable of N, showing that these elements may be longer in object position than they are in subject position. At present, we can only speculate that these results might be related to the temporal organization of the sentence such that initial parts of the sentence are said more quickly than later parts since more material is yet to be produced, but nothing hinges on this in this paper. More importantly, no interaction was found between the factors POSITION and SYNTACTIC FUNCTION in any of the subparts of the analysis, confirming that POSITION affects the length of the target element in subject and object constituents alike.

In conclusion, the statistical analysis confirms the lengthening effects described above which in turn entail the following two conclusions:

(i) There is a pp-boundary after the target sequence in both word orders and regardless of whether the sequence functions as subject or object of the carrier sentence.

(ii) There is no pp-boundary separating A and N in either order, neither in subject nor in object position.

Needless to say, these are exactly the properties uniquely characterizing the prosodic phrasing in (11) above. The next section will examine what underlying syntactic representation may determine the attested prosodic phrasing.

3 The syntactic representation of Italian N-A and A-N sequences

Since the seminal research by Bernstein (1991) and Cinque (1994), a rich variety of studies have shown that Romance and Germanic nominal expressions share an identical underlying hierarchical structure where adjectives are generated to the left of nouns (but see Larson and Marušič 2004 for a different position). Cinque (2005) is particularly convincing in this respect as he shows that given some general necessary restrictions on the possible movement types, only the universal base-generated order <D # A N> (where '#' stands for 'numeral') provides a suitable basis for the derivation of the crosslinguistically attested orders of D, #, A, and N among the 24 that are logically possible. We therefore assume the order <A N> as the base-generated order from which A-N and N-A sequences are derived.

As for the specific structural position of A, while some studies follow Abney (1987) in allowing it to be part of the main spine of the structure as in (19a) below –e.g. Artiagoitia (2006)– we follow the majority of scholars and assume that A is generated in the specifier of a corresponding functional projection as in (19b); see

for example Crisma (1991), Zamparelli (1995), Laenzlinger (2000), Shlonsky (2004), Knittel (2005), and Cinque (2005, 2006).

(19) a. … $[_{AP1}\ A_1\ [_{AP2}\ A_2\ [\ \ldots\ [_{NP}\ N\ ]$

b. … $[_{FP}\ AP_1\ F\ [_{FP}\ AP_2\ F\ [\ \ldots\ [_{NP}\ N\ ]$ (assumed in this study)

N-A sequences do not match the above structure and must therefore arise from movement of the noun to a functional projection above the adjective. For the sake of concreteness we identify this projection as Cinque's (2005) AgrP projection although the exact nature of this projection is irrelevant to the argument developed in this work. More relevant to our goals is the disagreement concerning whether the N-A order is obtained via N- or NP-raising to the left of the adjective. Under the above assumptions the N-raising hypothesis yields the structure in (20a), whereas the NP-raising hypothesis yields the structure in (20b).

(20) a. N-raising: $[_{AgrP}\ \_\_\ \ N_i\ [_{FP}\ AP\ \ t_i\ \ [_{NP}\ t_i\ ]]]$

b. NP-raising: $[_{AgrP}\ NP_i\ \ Agr\ \ [_{FP}\ AP\ \ F\ \ t_i\ \ ]]$

The main syntactic arguments supporting these opposite claims are examined in section 4; in this section we focus instead on the prosodic phrasing predicted by each representation under current models of the syntax-prosody mapping, which we then compare against the prosodic phrasing attested in our experiment. We start with N-A sequences and then move to A-N sequences in section 3.2.

3.1 Syntactic structure and predictions made by current prosodic theory

All major models of the syntax-prosody mapping map the N-raising structure (20a) into a single pp, as illustrated in (21a) below, and the NP-raising structure structure (20b) into separate pps as shown in (21b) (when matched against our experimental results NP and AP should of course be conceived as respectively containing a single noun and adjective and no additional modifiers). This identifies the N-raising structure in (20a) as the underlying syntactic representation of the N-A sequences examined in our experiment.

(21) a. (… N AP)

b. (… NP)(AP)

Below we examine how each model reaches the above prediction. While some details may differ, crucially all models dictate that a maximal projection such as NP

in the NP-raising structure is necessarily followed by a pp-boundary, thus determining the distinct predictions in (21).

Nespor and Vogel (1986) – Consider first the detailed study of Italian prosodic phrasing in Nespor and Vogel (1986). Their model requires lexical heads to form their own pps, with an additional optional restructuring rule extending these same pps to the first complement of the head (Nespor and Vogel 1986, p. 168,173)[1]. Under this model N and A would therefore form their own pps in both structures (20a) and (20b). The restructuring rule would then successfully apply to the N-raising structure where AP occurs within the complement of the raised N and place N and AP into a single shared pp. The same rule however would not apply to the NP-raising structure because AP is not part of the complement of N, hence leaving NP and AP each in a pp of its own against the phrasing attested in the above experiment.

Selkirk (1986) – The same prediction is made by the influential model proposed by Selkirk (1986) requiring the right edge of a maximal projection to be always followed by a corresponding pp-boundary. A pp-boundary must then follow the raised NP of (20b) and produce two distinct pps for NP and AP. The raised N of (20a) on the other hand escapes the requirement by being non-maximal, thus letting N and AP share the same pp.

Ghini (1993) – Closely following Selkirk, Ghini too provides a model for the prosodic phrasing of Italian that explicitly assumes a pp-boundary after the right edge of maximal projections. Ghini also examines additional principles of rhythmic organization that at first may appear to enable parsing the NP and AP in structure (20b) as a single pp. Ghini however states very clearly that these rhythmic principles never apply across the right edge of a maximal projection, thus subordinating them to Selkirk's right-edge rule. As Ghini points out, this restriction is necessary to prevent incorrectly phrasing a subject with a following verb, or an object with a following indirect object, postverbal subject, or higher adjunct. In all these cases the first item is a maximal projection separated from the following items by a pp-boundary. Subordinating Ghini's principles to Selkirk's right edge rule ensures that the attested separate phrasing remains unaltered.[2] It follows that

---

[1] Nespor and Vogel's restructuring rule joins together the pp of the selecting head with the pp of the complement provided the latter is a non-branching complement. This latter condition ensures that the pp contains at most one 'clitic group', i.e. no more than a single lexical item. The N-A sequences examined in our experiment satisfy this condition because AP contains a single complement-less modifier-free adjective. Indeed Nespor and Vogel (1986:172) provide the example '(caribú nani)' –meaning 'caribou dwarf'– as an N-A sequence sharing a single pp due to the restructuring rule.

[2] Some examples are provided in (22). Unless they are blocked from applying across the right edge of maximal projections, Ghini's principles of *average weight*, *symmetry*, and *increasing*

Ghini's model too predicts a pp-boundary after the raised NP of structure (20b). In fact, since Ghini's rhythmic principles are subordinated to Selkirk's right-edge rule, any analysis of the '(N A)' phrasing attested in our experiment in terms of Ghini's principles necessarily presupposes the lack of an NP-boundary and therefore the N-raising structure in (20a).

Truckenbrodt (1995, 1999) – A post-NP pp-boundary is also predicted under Truckenbrodt's model (1995, 1999, see also the similar model in Selkirk 2000). Truckenbrodt's model is based on Optimality Theory (Prince and Smolensky 1993/2004) and aims at capturing the effects of focus on prosodic phrasing across distinct languages. Since focus and stress do not play a role in our study we can limit our discussion to the two constraints governing the mapping between lexical projections and pp-boundaries. The first constraint, AlignXP(XP,Right,pp,Right), recasts Selkirk's right-edge rule in terms of McCarthy and Prince's (1993) theory of generalized alignment by requiring the edge of every lexical maximal projection XP to be aligned with a pp's right boundary. The effect is to introduce a pp-boundary immediately after every lexical XP, as in Selkirk (1986). The second constraint, Wrap-XP (or 'Wrap' for short), ensures that all the syntactic material dominated by the maximal projection of a lexical item be wrapped into a single pp. Wrap may for example force all the items dominated by a VP-node to be parsed into a single pp.

Together, AlignXP and Wrap predict separate pps for the NP-raising structure (20b). AlignXP requires a pp-boundary to follow the right edge of NP while Wrap is satisfied by the wrapping of NP and AP each on a pp of its own. The opposite holds for the N-raising structure: assuming that raised lexical heads make the functional projections hosting them prosodically lexical (Samek-Lodovici 2005), Wrap requires all material dominated by AgrP, namely N and AP, to be wrapped into a single pp, while AlignXP remains satisfied because the right edge of all available maximal projections, i.e. those of AgrP and AP, are properly followed by a pp-boundary. Crucially, N itself is not maximal and therefore not subject to AlignXP.

Truckenbrodt (1995) – The last model is a variant of the model just examined where AlignXP is replaced by the interaction of two constraints. The first one, StressXP, requires lexical XPs to express the prosodic peak of a pp (this peak, the pp's prosodic head, provides XP with pp-level stress, hence the name of the constraint). The second constraint, Align-pp(pp,Right,Head(pp),Right), ensures that the pp's prosodic head is aligned with the pp's right boundary. Together these two constraints once again ensure that lexical XPs are always immediately followed by

---

units, would incorrectly predict a shared single pp for (22a) rather than the attested two. They would also predict the phrasing '(V)(Obj XP)' in (22b) and (22c) rather than the attested '(V Obj)(XP)'.

a pp-boundary. On the one hand StressXP forces an XP to express the head of a pp, on the other hand Align-pp ensures that the same head is immediately followed by a pp-boundary. The overall effect is a pp-boundary after the XP. It follows that under this model too, the NP-raising structure projects two distinct pps to ensure that the NP and AP are each assigned their own right-aligned pp-head. The N-raising structure, on the other hand, allows for a single pp headed on the AP. This phrasing satisfies both constraints because the pp-head on AP is aligned with the pp-boundary and since the AP is in the complement of N the pp-head on AP also lies within the projection headed by N –i.e. AgrP– as required by StressXP (for a detail discussion of how StressXP can be satisfied with respect to a head H by placing a pp-head on its complement see Truckenbrodt 1995).

The convergence of all above models in predicting a pp-boundary after the raised NP of structure (20b) is not accidental. It is dictated by the need to capture the generalization that lexical maximal projections are always followed by a pp-boundary at their right edge (Nespor and Vogel 1986; Selkirk 1986, 2000; Ghini 1993; Truckenbrodt 1995, 1999). This generalization holds for Italian too and is well illustrated by the examples in (22) from Ghini (1993) and Frascarelli (2000) showing maximal projections in different clausal positions all trigerring a pp-boundary at their right edge. Specifically, we have a subject DP in (22a); an object DP followed by an indirect object in (22b) and a postverbal subject in (22c); and finally a left-peripheral topic followed by a subject in (22d) (further examples are available in Nespor and Vogel 1986 and Frascarelli 2000). In so far we expect N-A sequences to follow the generalization illustrated in these examples and accounted for by the above models, the NP-raising structure in (20b) cannot constitute the syntactic representation underlying the single pp experimentally attested for N-A sequences.

(22) a. (La veritá) (vínce)
the truth wins.3sg
'Truth wins'

b. (Daró líbri) (a Gianni)
(I) will-give.1sg books to John
'I will give books to John'

c. (Esamineránno il cáso) (gli espérti)
Will-examine.3pl the case the experts
'The experts will examine the case'

d. (a Gegé) (Páola) (gli parlerá) (dománi)
To Gegé Paola to-him will-speak.3sg tomorrow
'As for Gegé, Paula will talk to him tomorrow'

In contrast, joint prosodic phrasing of a head and its complement of the sort predicted to occur between N and AP under the N-raising structure (20a) is well attested and illustrated by the examples in (23) below[3] (see also Nespor and Vogel 1986; Ghini 1993; Frascarelli 2000). In (23a), a finite verb in T is phrased together with its object, while in (23b) a finite verb in T is phrased with a postverbal subject arguably stranded in specVP position.

The joint phrasing is confirmed by three phonological processes known to be impossible across pp-boundaries. The first, known as 'raddoppiamento sintattico', lengthens a word's consonantal onset whenever the previous word ends in a stressed syllable. Its effects are visible on the lengthened [p:] and [d:] onsets of the nouns in (23a) and attest the absence of pp-boundaries before those nouns. The second process lengthens the syllable receiving the pp's prosodic peak, here determining the lengthened [e:] and [u:] in the nouns in (23a). The third process solves potential stress clashes by shifting the first stress leftwards whenever a word with word-final stress is followed by a word with word-initial stress. In the examples below, the final stress of the verb shifts to its first syllable, shown in bold.

Sentence (23b) is particularly revealing because its underlying structure, shown in (23c), parallels the N-raising structure in (20a), repeated in (23d). The DP occurs in the specifier of the complement projection of the raised V much like AP does with respect to the raised N. This structural parallelism guarantees that any model blind to categorial labeling that accounts for the shared pp in sentences like (23b) will also necessarily predict a shared pp for N-A sequences derived via N-raising.

(23) a. (**Sof**frirá [p:][é:]ne) (incredibilménte d[ú:]re)
(s/he) will-suffer.3sg afflictions incredibly severe
'S/he will suffer extremely severe afflictions'

b. (**Nuo**terá Giánni)
Will-swim.3sg John
'John will swim'

c. $[_{TP}\ V_i\ [_{VP}$ DP $t_i\ ]]$

d. $[_{AgrP}\ N_i\ [_{FP}$ AP $t_i\ [_{NP}\ t_i\ ]]]$

The above discussion shows that the attested joint phrasing of N-A sequences requires the N-raising structure (20a) as their underlying representation.

[3] The examples also confirm our earlier assumption that raising lexical heads make their target functional projections lexical as far as prosody is concerned. If this were not the case, none of the analyses of syntax-prosody mapping discussed above would account for the attested phrasing, since the finite verb raised to T would no longer count as lexical.

Nevertheless, as a further test of the strength of this conclusion we may consider whether the above prosodic models possibly overlooked some crucial constraint or device that once taken into account would actually predict a single pp even under the alternative NP-raising hypothesis.

For example, extending Truckenbrodt's analysis one could hypothesize the existence of a constraint Wrap-FP requiring all material dominated by a functional projection to be parsed into a single pp. Under the NP-raising structure, repeated in (24) below, Wrap-FP would phrase NP and AP together because they are both dominated by the functional projection AgrP (or any other equivalent functional projection). In Truckenbrodt's analysis this constraint would conflict with AlignXP, but the ranking Wrap-FP>>AlignXP would ensure that AlignXP is violated in order to satisfy Wrap-FP.[4]

(24) NP-raising: $[_{AgrP}\ NP_i\ Agr\ \ [_{FP}\ AP\ F\ t_i\ ]]$

This hypothetical analysis however is inconsistent with the generalization requiring a pp-boundary after lexical maximal projections illustrated by the data in (22) above. For example, the left-peripheral topic in (22d) lies in the specifier of the CP-level functional projection TopicP (Rizzi 1997) yet it is not phrased together with the lower subject as mandated by Wrap-FP. Likewise the subject of (22a) lies in the specifier of TP (or possibly TopicP if they constitute left peripheral topics as argued by several analyses - cf. below), yet it cannot be phrased together with the following verb.

One could counter that in the TP case, TP is made lexical by the raising verb and hence it is no longer subject to Wrap-FP. Italian subjects however remain parsed in a pp of their own even when T is filled by an auxiliary and hence undoubtedly qualifies as functional, see the examples in (25) below. These failures are inevitable and follow from the similarities between the NP-raising structure and the structures for the sentences in (22) and (25). These similarities ensure that any constraint phrasing NP and AP together in the NP-raising structure will also incorrectly require a single pp in sentences like (22) and (25). Nor can this problem be solved via constraint-ranking, since ranking is fixed within a grammar and therefore the ranking for the NP-raising structure carries over to the data in (22) and (25). The only new constraint making the correct distinctions would be a version of Wrap-FP that only applies to DPs, but this would obviously be an ad-hoc uninformative solution.

---

[4] In the variant of Truckenbrodt's analysis where AlignXP is replaced by StressXP and Align-pp, the needed ranking would be Wrap-FP>>{StressXP, Align-pp}.

(25) a. (la gabbia) (é giá caduta) (Nespor and Vogel 1986:170)
the cage is already fallen
'The cage has already fallen'

b. (Carlo) (ha portato) (tre bassotti) (alla mostra) (Frascarelli 2000:19)
Carlo has brought three dachshunds to-the show
'Carl brought three dachshunds to the show'

The above discussion leaves little doubt about the underlying syntactic structure of the N-A sequences. All major models of the prosody-syntax mapping unambiguously associate the attested phrasing of N-A sequences with the N-raising representation repeated in (26), which furthermore perfectly fits the empirical generalization holding of Italian syntax-prosody mapping.

(26) N-raising: $[_{AgrP}\ N_i\ [_{FP}\ AP\ t_i\ [_{NP}\ t_i\ ]]]$

Given such a convergence between theoretical and empirical observations we conclude that the attested prosodic phrasing of N-A sequences provides unequivocal evidence for the occurrence of N-raising in Italian.

3.2 Syntactic representation of A-N sequences

Turning to A-N sequences, the joint '(A N)' phrasing attested in our experimental data may at first appear surprising since the initial AP might appear to have to trigger a pp-boundary at its right edge. Once again we should consider the possible underlying representations in detail. The structure in (27a) occurs if N moves to the head of the functional projection hosting AP, while structure (27b) occurs if N does not move.

(27) a. N-raising: $[_{FP}\ AP\ N_i\ [_{NP}\ t_i\ ]]]$

b. No raising: $[_{FP}\ AP\ F\ NP\ ]$

Structure (27b) is clearly inconsistent with the attested phrasing. This structure parallels the NP-raising structure discussed in the previous section, with a lexical XP in the specifier of a functional head that is not targeted by N-raising. In this case all models of syntax prosody mapping predict a pp-boundary after the initial AP for the reasons already explained in the previous section. Since the expected boundary is absent (27b) cannot be the correct representation.

The same models differ in their predictions with respect to (27a). Nespor and Vogel (1986), Selkirk (1986), and Ghini (1993) predict a post-AP boundary and therefore cannot account for the single pp found for A-N sequences. In

Truckenbrodt's model, instead, the phrasing of (27a) depends on the ranking of Wrap and AlignXP relative to each other as illustrated by Tableau 1 below. The raised N makes the entire FP lexical as far as prosodic constraints are concerned. Wrap therefore requires the entire FP to be contained in a single pp. Consequently languages where Wrap dominate AlignXP will phrase (27a) into a single pp even if this phrasing violates AlignXP.[5] Note that the same is not true of (27b), where Wrap only requires AP and NP to be each contained within a single pp but with no condition imposed on the entire FP. Even languages with the Wrap>>AlignXP ranking will therefore choose to satisfy AlignXP and place a boundary after AP as mentioned above.

Tableau 1	Wrap	AlignXP
☞ (AP N)	✔	*
(AP)(N)	*	✔

If AP can be phrased with the following noun in (27a), why are the left-peripheral topics and subjects of the sentences in (22) above parsed in a pp of their own? The crucial difference lies in their syntactic representation. In (27a) AP occurs as the specifier of a projection headed by the raised N. As mentioned above, this makes the projection prosodically lexical and hence subject to Wrap. The same is not true for the examples in (22). Consider for example sentence (22d), repeated as (28) below. The left-peripheral topic 'a Gegé' lies in a TopicP projection at the top of the clause as shown in (28b). The head of TopicP is not filled by a lexical head. Consequently Wrap places no condition on TopicP. The available trivial satisfaction of Wrap in turn enables the satisfaction of the lower ranked AlignXP, determining the pp-boundary immediately following the topic 'a Gegé'.

---

[5] The same analysis extends to Truckenbrodt's StressXP and AlignPP variant under the ranking {Wrap, AlignPP}>>StressXP. The condition Wrap>>StressXP ensures that (AP N) is parsed in a single pp as required by Wrap even if (AP)(N) would better satisfy StressXP by providing both items with their own pp-head. The condition AlignPP>>StressXP in turn ensures that within (AP N) the pp-head falls rightmost on N as attested despite the violation of StressXP on AP which is left without a pp-head. The opposite ranking would still predict a single pp (AP N) but by placing the pp-head on AP to satisfy StressXP (since AP is contained within the projection of N, the pp-head on AP also counts as satisfying StressXP relative to N; see Truckenbrodt 1995 for a detailed discussion). The relevant competition is provided in the tableaux below where the item carrying the pp-head is shown in bold.

	Wrap	AlignPP	StressXP
☞ (AP **N**)	✔	✔	*
(**AP** N)	✔	*	✔
(**AP**) (**N**)	*	✔	✔

(28) a. (a Gegé) (Páola) (gli parlerá) (dománi)
To Gegé Paola to-him will-speak.3sg tomorrow
'As for Gegé, Paula will talk to him tomorrow'

b. [$_{\text{TopicP}}$ [a Gegé]$_{\text{PP}}$ Topic$^{\text{o}}$ [….]]

The same analysis applies to the subject 'Paola' in (28a) above and any other Italian preverbal subject provided they too are analyzed as located in the specifier of a higher projection not reached by raised finite verbs. While there are differences about the location of this projection (namely whether it is part of CP or the inflectional field) and also about the preverbal subjects that it hosts (i.e. whether they are base-generated and controlling a lower resumptive *pro* or raised all the way up from specVP), its existence is advocated by many of the scholars investigating preverbal subjects in null subject languages; see for example Barbosa (1995), Alexiadou and Anagnostopoulou (1998), Rizzi (2004), Cardinaletti (2004), Frascarelli (to appear). For example, as noted by Alexiadou and Anagnostopoulou (1998) the higher position of overt subjects in languages like Italian follows immediately from the observation that they can be followed by a sentential adverb like 'probably' (cf. (29a)) or by an entire subordinate clause (cf. (29b)). This would be impossible if overt subjects were in a spec-head relation with the inflected verb.

(29) a. Gianni probabilmente ha incontrato Maria.
John probably has met Mary.
'Probably John has met Mary'

b. I bambini se Maria viene andranno via.
The children if Mary comes will go away.
'If Mary comes the children will go away'

The position of the above subjects on the other hand is accounted for if, as proposed for example in Frascarelli (to appear), preverbal subjects are sentential topics sitting in a dedicated topic-related projection ShiftP (for 'aboutness-shift topic') from where they control a lower *pro* subject as shown in (30a) (adapted from Frascarelli; see also Barbosa 1995 and Alexiadou and Anagnostopoulou 1998). The analysis also correctly predicts that the same sentence will be ungrammatical in languages that lack null subjects; see for example sentence (30b) from French (Alexiadou and Anagnostopoulou 1998).

(30) a. [$_{\text{ShiftP}}$ Gianni$_{\text{k}}$ [$_{\text{FP}}$ probabilmente [$_{\text{TP}}$ ha [$_{v\text{P}}$ *pro*$_{\text{k}}$ [$_{\text{VP}}$ incontrato Maria]]]]]

b. *Jean probablement a rencontré Marie
John probably has met Mary.
'Probably John has met Mary'

In conclusion, the single pp found for A-N sequences provides evidence for an N-raising analysis where N moves into the head of the functional projection hosting AP in its specifier. It also supports Truckenbrodt's model of syntax prosody mapping, as all other models incorrectly predict distinct pps for A and N. Finally, it provides further evidence for analyzing Italian referential preverbal subjects as located in a higher topic-related position dominating TP.

4 Head and phrasal movement in DPs

With respect to the debate on whether N-A order in Romance is due to N- or NP-raising our results clearly support the former hypothesis for the reasons provided in the previous sections. It is natural to wonder to which degree this result is consistent with the available syntactic arguments for and against N- and NP-raising. As we will show, once closely examined, these arguments do not support a simple conclusion for or against N- or NP-movement. Arguments provided for either of the two analyses have often been reanalyzed as arguments for the opposite analysis. See for example the stranding of prepositional complements by a raising noun in <N A PP> sequences, proposed as evidence for N-raising in Cinque (1994) and Pereltsvaig (2006) but reanalyzed in terms of NP-raising in Cinque (2006); or the presence of N-raising in English, argued for in Kishimoto (2000) but rejected in Larson and Marušič (2004) and Cinque (2006). There also appears to be robust evidence for phrasal movement, but it appears to apply to functional projections above NP and its application to NP itself is far less obvious.

In this section we examine some of the most relevant evidence concerning the presence or absence of N-movement in detail and eventually argue for an analysis of Italian DPs where phrasal movement of functional projections co-exists with strictly local N-raising. We start in 4.1 with a brief survey of the syntactic evidence supporting the availability of N-raising, then move to the syntactic evidence for phrasal movement and examine its compatibility with the N-raising analysis advocated in the previous sections.

4.1 Head movement

The availability of N-raising has been argued for in several independent studies and across many languages (e.g., Sproat and Shih 1990; Crisma 1991; Valois 1991; Bernstein 1991, 1993; Cinque 1994; Zamparelli 1995; Longobardi 1994, 2001,

2005; Kishimoto 2000; Rutkowski and Progovac 2006; Willis 2006; Pereltsvaig 2006; Airtiagoitia 2006). While evidence based on word order alone cannot distinguish between N- and NP-raising, direct interaction of the raising noun with higher syntactic heads provides the clearest possible syntactic evidence for the existence of N-raising. This interaction may occur in three different ways: (i) replacement of a higher head H with N, yielding a complementary distribution between H and N; (ii) incorporation of N into a higher head H, with H emerging as an affix of N; (iii) blocking of N-raising by a higher head intervening in the path of N. Below we describe an instance of each of the above cases but further instances are available in the literature cited above.

Head-replacement is well illustrated by Longobardi's studies of N-to-D movement, which also provide some of the strongest possible evidence for the availability of N-raising in Italian (Longobardi 1994, 1996, 2001, 2005). Longobardi notices that whenever a determiner is present, a proper name N occurring with possessives or with focusing adjectives like 'solo' (only) requires the D-A-N order. Yet the same noun allows for the N-A order when the determiner is absent, see the paradigm below from Longobardi (2001) (for additional cases of N-to-D across distinct languages see Longobardi 2001 and references listed there).

(31) a. La sola Napoli (è stata prescelta tra le città italiane).
the only Naples (is been chosen among the cities Italian)
'Only Naples was selected among Italian cities'

b. *La Napoli sola (è stata prescelta tra le città italiane).

c. Napoli sola (è stata prescelta tra le città italiane).

d. * Sola Napoli (è stata prescelta tra le città italiane).

As Longobardi remarks, N-to-D raising immediately explains the complementary distribution of N and overt D in DP-initial position. The same is not true for NP-raising since NP would move to specDP and thus allow for the occurrence of D to its right. Longobardi's analysis is more complex and informative than we can expose here, identifying for example the conditions that determine whether N-to-D raising is or is not available to specific classes of nouns. What is relevant in the context of this study is that it shows that N-raising can occur in Italian DPs, providing independent support for the N-raising analysis of N-A sequences advocated in this paper.

Further evidence for N-raising comes from cases where a raising noun incorporates into D, with D surfacing as a suffix of the noun. This is best illustrated by Scandinavian languages, including the Danish examples in (32a) below where the determiner 'en' emerges as a suffix of the raised noun 'hest' (horse) (Delsing 1993; Embick and Noyer 2001). The original post-determiner position of the noun

is visible in (32b), where N-raising is blocked by the intervening adjective 'rode' (red).

(32) a. hest-en
horse-def
'the horse'

b. den rode hest
def red horse
'the red horse'

The third type of interaction, N-raising blocked by an intervening head, is already illustrated by the above Danish data.[6] An interesting case is also examined in Pereltsvaig's (2006) analysis of Russian approximative inversion (but see also her discussion of Hebrew and Artiagoitia's 2006 discussion of Basque). As (33) shows, a noun raising to the left of its numeral specification determines an approximative interpretation of the numeral. Following Bailyn (2004), Pereltsvaig maintains that the numeral occurs in the specifier of a higher NumP projection when assigned genitive case but in the head of the same projection when assigned instrumental case. Correspondingly, N-raising to the even higher projection associated with the approximative interpretation is possible under genitive case but not instrumental case, where the intervening overt Num head blocks N-raising; see the examples in (33) and (34) from Pereltsvaig (2006:277, 283).

(33) a. desjat' kardinalov
ten cardinals
'ten cardinals'

b. kardinalov desjat'
cardinals ten
'approximately ten cardinals'

(34) a. Džejms Bond vypil rjumok desjat' vodki.
James Bond drank-up glasses.GEN ten vodka.GEN
'James Bond drank up approximately ten glasses of vodka.'

[6] The fact that in the above data N-raising is blocked by intervening adjectives might support a view of adjectives as heading projections that are a direct part of the main spine of the DP structure as in Artiagoitia (2006). The same analysis however does not apply to languages like Italian where adjectives show no corresponding blocking effects.

b. *Džejms Bond napilsja rjumkami desjat'ju vodki.
James Bond got-drunk glasses.INSTR ten.INSTR vodka.GEN
'James Bond got drunk from approximately ten glasses of vodka.'

In conclusion, there is substantive independent syntactic evidence for N-raising in both Italian and other languages. As discussed in the next section, this does not exclude the presence of phrasal movement within Italian DPs but neither is the presence of phrasal movement sufficient to exclude N-raising.

4.2 Phrasal movement

The evidence for DP-internal phrasal movement is compelling, see among others Laenzlinger (2000), Knittel (2005), and Cinque (2005, 2006). Here we examine Cinque's 2005 and 2006 studies because we find them particularly informative and also because they explicitly argue against N-movement. We show that while they provide clear evidence for the occurrence of phrasal movement in Italian DPs, on close inspection they do not exclude the N-raising analysis found necessary for the adjectives and nouns tested in our experiment.

As mentioned, Cinque (2005) convincingly argues for a universal merge order <Dem # A N>. His study also examines what movement restrictions are necessary to prevent the derivation of unattested orders reaching two important conclusions. The first one is that remnant movement must be disallowed. If this were not the case unattested orders would become derivable. For example, the impossible <Dem A # N> order becomes derivable as shown in (35) by first raising N or NP to the left of A and then moving the remnant FP containing the adjective to the left of #. Moved constituents are shown in square brackets.

(35) <Dem # A N> → <Dem # [N] A t_N> → <Dem [$_{FP}$ A t_N] # N t>

The second conclusion concerns the necessity of pied-piping in order to derive those attested orders that seem to defy the universal <Dem # A N> hierarchy (see also Shlonsky 2004). For example, as (36)a shows the order <N A # Dem> can be derived by pied-piping the complement of the projection hosting the raised noun (i.e. pied-piping of the '*picture of who*' type in Cinque's terminology), while (36)b shows how the order <# N A Dem> follows from pied-piping the projection immediately above the raising noun (pied-piping of the '*whose picture*' type). In contrast, no amount of pied-piping will ever derive orders that require a different initial merge-order. For example, the impossible <Dem A # N> or <# N Dem A> would respectively require merging A before # in one case and merging # before Dem in the other. Given these restrictions, the ten universally unattested orders can

be characterized as those that defy the <Dem # A N> hierarchy while remaining underivable via noun movement and pied-piping (Cinque 2005).

(36) a. <Dem # A N> → <Dem # [N] A> → <Dem [N A] #>
→ <[[N A] #] Dem>

b. <Dem # A N> → <Dem # [N] A> → <[# [N] A] Dem>

Cinque's analysis provides strong evidence for phrasal movement, since without it pied-piping could not occur and several attested orders would be incorrectly predicted impossible. Cinque, however, also claims N-movement to be impossible. This claim is made in the context of a more general theoretical goal aiming at keeping all syntactic movement phrasal. Yet we could not find any specific evidence directly linking N-movement to the derivation of one of the unattested orders (a similar conclusion is reached in Artiagoitia 2006:22).[7] As far as we can see all instances of NP-movements proposed by Cinque can be recast in terms of N-movement with no analytical loss but for the theoretical goal mentioned above. For example, nothing prevents the orders <Dem # [N] A>, <Dem [N] # A>, and <[N] Dem # A> derived via NP-movement by Cinque from being reanalyzed as derived by N-movement. Nor does N-movement enable the derivation of any of the unattested orders provided remnant movement remains banned. All in all, while keeping all movement phrasal might be theoretically desirable, the case against N-movement is not empirically supported. The evidence for N-raising presented in the previous sections and the evidence for phrasal movement in Cinque's research instead suggest that DP-internal phrasal and head-movement might co-exist, with phrasal movement affecting functional projections and head-movement restricted to N-raising.

Cinque (2006) further refines the argument for DP-internal phrasal movement by arguing for its obligatoriness in Italian. To understand his argument we have to first recapitulate some of Cinque's findings. In an impressive comparative study of the distribution of nouns and adjectives in English and Italian Cinque (2006) shows that adjectives divide into two main classes: adjectives expressing indirect modification, best analyzed as reduced relative clauses, and adjectives expressing

[7] Many of the claims against N-movement in Cinque (2005) reject it in combination with remnant movement. As such these claims are of course correct but the problem in these cases is remnant movement itself, which gives rise to incorrect predictions even in absence of N-movement (as also noticed by Cinque 2005: 324, fn30). In a recent personal communication Cinque agrees that the main argument favoring NP-movement over N-movement is conceptual, related to the goal of keeping all movement phrasal. He also notes that under N-movement the impossibility of remnant movement must be stipulated, whereas under NP-movement it can be derived from Kayne's (2005) closeness driven movement analysis (see Cinque 2005:326).

direct modification[8] (see also Alexiadou 2001 and Knittel 2005). Furthermore the set of possible pre-nominal and post-nominal adjectives in Italian and English are shown to differ according to the linear distributions shown in (37).

(37) a. Italian: Direct Mod. > **N** > Direct Mod. > Indirect Mod.

b. English: Indirect Mod. > Direct Mod. > **N** > Indirect Mod.

As Cinque points out, these distributions suggest that the two adjectival classes are merged in distinct positions within the DP. Yet, Cinque notes, it is impossible to posit a universal merge order while at the same time deriving the above distributions via N-movement alone. For example, the universal hierarchy in (38a) is incompatible with Italian because movement of N into the middle position yields the incorrect distribution of pre- and post-nominal adjectival classes, while movement into the leftmost position yields the incorrect order for post-nominal classes. The alternative hierarchy in (38b) in turn is incompatible with English because the order for pre-nominal classes becomes unreachable with or without N-movement.

(38) a. Indirect Mod. > Direct Mod. > **N**

b. Direct Mod. > Indirect Mod. > **N**

This impasse provides Cinque with a persuasive argument in support of phrasal movement in Italian. Assuming (38a) as the universal merge order, yielding the base order in (39), the Italian distribution is reached as shown in (40a) and (40b) below by obligatorily moving the lower section of the DP containing the direct modification adjectives (i.e. FP3) to the left of the phrase containing the indirect modification adjectives (i.e. into specFP1). The NP itself may or may not move to the left of the direct modification adjectives (i.e. in specFP3)[9]. If it remains

---

[8] A full description of the two classes is provided in Cinque (2006) and cannot be replicated here for reasons of space. In general indirect modification adjectives constitute reduced relative clauses and systematically associate with stage-level, restrictive, intersective, relative, and propositional interpretations. Direct modification adjectives instead associate with individual-level, non restrictive, non intersective, absolute, specificity-inducing, and evaluative interpretations. For example, Cinque points out how 'visible' in prenominal position in 'the visible stars include Aldebaran and Sirius' is ambiguous between a stage-level and an individual-level reading, whereas it may only take a stage-level reading when occurring postnominally in 'the (only) stars visible are Aldebaran and Sirius'. In Italian, a similar adjective necessarily has an individual-level reading when prenominal and only becomes ambiguous when postnominal.

[9] Whether movement past direct modification adjectives occurs or not also depends on the particular subclass of adjectives. Cinque (2006) notes that movement is obligatory with

unmoved the final order is <Dir.Mod N Ind.Mod>, see (40a). If it moves, the final order becomes <N Dir.Mod Ind.Mod.>, see (40b). Crucially, in both cases the final structure matches the attested distribution of Italian adjectives in (37a).

(39) $[_{FP1}\ F_1\ [_{FP2}\ AP_{Ind.Mod.}\ F_2\ [_{FP3}\ F_3\ [_{FP4}\ AP_{Dir.Mod.}\ F_4\ NP\ ]]]]$

(40) a. $[_{FP1}\ [_{FP3}\ F_3\ [_{FP4}\ AP_{Dir.Mod.}\ F_4\ NP\ ]]_i\ F_1\ [_{FP2}\ AP_{Ind.Mod.}\ F_2\ t_i\ ]]$

b. $[_{FP1}\ [_{FP3}\ NP_k\ F_3\ [_{FP4}\ AP_{Dir.Mod.}\ F_4\ t_k\ ]]_i\ F_1\ [_{FP2}\ AP_{Ind.Mod.}\ F_2\ t_i\ ]]$

Once again Cinque's analysis provides strong evidence for the presence of phrasal movement of functional projection within Italian DPs but does not provide direct evidence against N-movement with respect to adjectives of direct modification. Cinque's observation that unbounded N-movement determines the incorrect order <N Ind.Mod Dir.Mod> also applies to unbounded NP-movement (for example NP-raising to specFP3 followed by NP-movement to specFP1 while assuming no FP3 movement). What Cinque's study truly shows is that nouns cannot raise above indirect modification adjectives, whether via N- or NP-movement. Within the lower section of the DP, however, N-raising remains a viable hypothesis.[10]

4.3 A combined analysis of N- and phrasal raising for Italian DPs

Is it possible to simultaneously derive Cinque's results on phrasal movement and the head-raising representation shown necessary in the previous sections? One possibility is to combine the two analyses and maintain a representation of Italian DPs like (40) above where N raises as a head to F4 and optionally to F3, as required by the attested prosodic phrasing of N-A and A-N sequences, but never beyond F3, as required by Cinque's analysis of Italian adjectives. As Cinque notices, once movement is so restricted only direct modification adjectives can occur in both the N-A and A-N orders whereas indirect modification adjectives necessarily follow the noun. This is a welcome result as it explains why symmetric N-A and A-N sequences necessarily involve direct modification adjectives.

---

classificatory adjectives and adjectives of provenance/nationality but optional with higher direct modification adjectives of color, shape, size, value, etc.

[10] Cinque's results also raise the issue of the location of D in view of Longobardi's N-to-D raising. One possibility is that D in Italian is merged immediately above FP3; see for example the hypothesis allowing for the existence of an indefinite determiner projection in this position in Cinque (2006).

As for why N-raising co-exists with phrasal raising of higher functional projections, a possible answer is suggested by Cinque's (2005) proposal that DP-internal movement is subject to a version of Kayne's constraint on remnant movement (Kayne 2005, p. 54). Kayne's constraint restricts the set of syntactic items that can legitimately move to the specifier of a head H to the closest category distinct from the complement of H. Under a bare-structure representation à la Chomsky (1995) the noun of the simple N-A and A-N sequences examined in our experiment would simultaneously count as head and maximal projection. Raising the noun as a head would form a shorter chain and thus be preferable to phrasal movement on economy considerations, explaining the occurrence of N-raising. The same choice, however, is not available to higher functional projections because they necessarily involve a complement and therefore count as phrasal, explaining Cinque's findings on the phrasal movement of higher functional projections.

The proposed combined analysis also makes fine-grained predictions that could potentially highlight subtle syntactic differences not easily testable with syntactic means alone but potentially revealed by prosodic phrasing. The first prediction concerns N-A sequences involving unambiguous indirect modification adjectives. Following Cinque (2006), the combined analysis assigns to them a structure similar to (41) below, where AP is preceded by an entire NP encapsulated within the raising FP3 projection. Consequently N and A would be predicted to occur in two separate pps, since NP triggers a pp-boundary to its right. We thus expect the structural difference between N-A sequences involving direct modification adjectives, based on N-raising, and sequences involving indirect modification adjectives, formed via phrasal raising, to be reflected in their prosodic phrasing. If borne out, this prediction would at once provide significant support for Cinque's analysis as well as for the N-raising representation for N-A and A-N sequences advocated here.

(41) $[_{FP1} \quad [_{FP3}\ F3\ [_{FP4}\ F4\ NP\ ]]_i \qquad F_1 \quad [_{FP2} \quad AP \quad F_2 \quad t_i\ ]]$

The second prediction concerns the analysis of expressions like (42) below adapted from Cinque (1994) where an adjective intervenes between a noun and its complement.

(42) L'invasione brutale di Parigi
the invasion brutal of Paris
'the brutal invasion of Paris'

There are two possible structural analyses of (42). The more traditional one, proposed in Cinque (1994) and reconsidered in Pereltsvaig (2006), maintains that the noun raises above the adjective as a head, stranding its prepositional

complement behind as shown in (43). This analysis is inconsistent with Kayne's remnant movement constraint which would require the entire NP to move to specFP3, PP-complement included.

(43) $[_{FP3}\ \ N_i\ \ [_{FP4}\ \ AP\ \ F_4\ \ [_{NP}\ t_i\ PP\ ]]]$

Cinque (2006) proposes an alternative analysis based on Kayne (1999, 2000, 2002) where the preposition 'di' introducing the complement is supplied at the top of the DP and the complement of the noun moves to get case leaving the noun able to raise as a complement-free NP. The main derivational steps are shown in (44). Stage (i) provides the initial configuration. Stage (ii) follows from merge of K(ase) and attraction of the complement DP to its specifier. Stage (iii) follows from merge of the preposition 'di' (of) at the top of the structure and the attraction of the remnant 'la brutale invasione' (the brutal invasion). Stage (iv) follows from raising the entire NP above the adjective.

(44) i. [La [brutale [invasione [Parigi]]]

ii. $[[\text{Parigi}]_k$ K [la [brutale [invasione t_k]]]]

iii. $[[\text{La [brutale [invasione t]]}]_s$ di [[Parigi] K t_s]]

iv. $[[\text{La } [[\text{invasione t}]_i$ [brutale t_i]]$_s$ di [[Parigi] K t_s]]

Once again the two analyses predict a distinct prosodic phrasing of the resulting N-A sequence. The analysis of Cinque (1994) predicts a joint pp, whereas the Kaynian analysis in Cinque (2006) predicts two distinct pps, because the raised NP would trigger a pp-boundary at its right edge. If borne out, the latter prediction would thus provide at once evidence for the NP-raising analysis argued for in Cinque (2006) as well as new independent support for Kayne's model of syntactic derivations.

5 Conclusion

The robust generalizations available on Italian prosodic phrasing and current models of the syntax-prosody mapping converge in dictating an N-raising analysis for simple N-A and A-N sequences involving adjectives of direct modification. An alternative analysis based on NP-raising would directly contradict the joint phrasing that was experimentally attested, as well as current understanding of how prosodic phrasing is determined in Italian and other languages.

This result was shown to have important syntactic and prosodic implications. From a prosodic perspective, it lends support to Truckenbrodt's model of the

syntax-prosody mapping, which alone among the models examined here can predict the joint prosodic phrasing of both N-A and A-N sequences. From a syntactic perspective, we showed that the N-raising analysis is consistent with the equally necessary phrasal movement discussed in Cinque (2005, 2006). What emerges is a complex model of Italian DPs where limited, local, N-raising co-exists with phrasal movement of higher functional projections. Furthermore the occurrence of N- vs. NP-raising was shown to be potentially determined by the type of items contained in the DP itself, depending for example on the class of adjectives preceding or following the noun as well as the presence of a complement of the noun.

We also showed how prosodic phrasing can test and determine fine-grained properties of syntactic analysis that range from the syntactic representation of Italian overt subjects and complex DPs to the precise conditions constraining movement in UG and the distinct available models of syntactic derivations.

Appendix: Experimental Materials

- **[A N]$_{Subj}$ V Obj**

1. Un **po.TEN.TE pre.LA.TO** può imporre il suo punto di vista anche al papa
   A powerful prelate can impose the his point of view even to-the pope

2. Un **pro.VET.TO ten.NIS.TA** dovrebbe evitare un errore così plateale
   An experienced tennis-player should avoid a mistake so evident

3. Un **TI.pi.CO pre.TES.TO** comporta il dichiararsi malati anche se in ottima salute.
   A typical excuse involves the self-declaring sick even if in optimal health

4. Un **BRUT.TO POR.TO** diminuirebbe il valore touristo della nostra città
   An ugly harbor would-lower the value tourist of-the our town

5. Un **cor.RET.TO con.TAT.TO** determina l'accensione della spia verde.
   A correct contact determines the switching-on of the light green

- **[N A]$_{Subj}$ V Obj**

1. Un **pre.LA.TO po.TEN.TE** può imporre il suo punto di vista anche al papa
   A prelate powerful can impose the his point of view even to-the pope

2. Un **ten.NIS.TA pro.VET.TO** dovrebbe evitare un errore così plateale
   A tennis-player experienced should avoid a mistake so evident

3. Un **pre.TES.TO TI.pi.CO** comporta il dichiararsi malati anche se in ottima salute.
   An excuse typical involves the self-declaring sick even if in optimal health

4. Un **POR.TO BRUT.TO** diminuirebbe il valore touristo della nostra città
   An harbor ugly would-lower the value tourist of-the our town

5. Un **con.TAT.TO cor.RET.TO** determina l'accensione della spia verde.
   A contact correct determines the switching-on of the light green

- **Subj V [A N]$_{Obj}$ X**

1. Abbiamo contattato un **po.TEN.TE pre.LA.TO** con il permesso del papa
   (We) have contacted a powerful prelate with the permission of-the pope

2. Abbiamo bisogno di un **pro.VET.TO ten.NIS.TA** per la squadra olimpica
   (We) have need of an experienced tennis-player for the team Olympic

3. Gli studenti hanno usato un **TI.pi.CO pre.TES.TO** con pessimi risultati.
   The students have used a typical excuse with terrible results

4. Vorremmo evitare un **BRUT.TO POR.TO** proprio al centro della città
   (We) would-like to-avoid an ugly harbor at-the centre of-the town

5. Le componenti elettriche devono stabilire un **cor.RET.TO con.TAT.TO** prima di ogni accensione.
   The components electric must establish a correct contact before of every switching-on

- **Subj V [N A]$_{Obj}$ X**

1. Abbiamo contattato un **pre.LA.TO po.TEN.TE** con il permesso del papa
   (We) have contacted a powerful prelate with the permission of-the pope

2. Abbiamo bisogno di un **ten.NIS.TA pro.VET.TO** per la squadra olimpica
   (We) have need of a tennis-player experienced for the team Olympic

3. Gli studenti hanno usato un **pre.TES.TO TI.pi.CO** con pessimi risultati.
   The students have used an excuse typical with terrible results

4. Vorremmo evitare un **POR.TO BRUT.TO** proprio al centro della città
   (We) would-like to-avoid a harbor ugly at-the centre of-the town

5. Le componenti elettriche devono stabilire un **con.TAT.TO cor.RET.TO** prima di ogni accensione
   The components electric must establish a contact correct before of every switching-on

References

Abney, S. 1987. *The English Noun Phrase in its sentential aspect*. Doctoral dissertation, MIT.

Alexiadou, A. & E. Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP checking. *Natural Language and Linguistic Theory* 16, 491-539.

Alexiadou, A. 2001. Adjective syntax and noun raising: Word order asymmetries in the DP as the result of adjective distribution. *Studia Linguistica* 55(3), 217-248.

Artiagoitia, X. 2006. Basque adjectives and the functional structure of the noun phrase. Ms., University of the Basque Country.

Bailyn, J. F. 2004. The case of Q. *Proceedings of FASL* 12, 1–35.

Barbosa, P. 1995. *Null subjects*. Ph.D. dissertation, MIT.

Beckman, M. & J. Edwards. 1987. The phonological domains of final lengthening. *Journal of the Acoustical Society of America* 81.

Beckman, M. & J. Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. Beckman (eds.), *Between the grammar and physics of speech (Papers in laboratory phonology I)*, 152-178. Cambridge: Cambridge University Press.

Beckman, M. & J. Edwards. 1991. Prosodic categories and duration control. *Journal of the Acoustical Society of America* 89.

Bernstein, J. 1991. DPs in French and Walloon: Evidence for parametric variation in nominal head movement. *Probus* 3, 101-126.

Bernstein, J. 1993. *Topics in the syntax of nominal structure across Romance*. Ph.D. dissertation, CUNY.

Bhattacharya, T. 1998. Kinship inversion in Bangla. *Papers in Linguistics from the University of Manchester (PLUM)* 7, 143-156.

Boersma, P. 2001. PRAAT: a system for doing phonetics by computer. *Glot International* 5(9/10), 341-345.

Cardinaletti, A. 2004. Towards a cartography of subject positions. In L. Rizzi (ed.), *The cartography of syntactic structures, Vol. 2: The structure of CP and IP*, 115-165. Oxford: Oxford University Press.

Chomsky, N. 1995. Bare phrase structure. In G. Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*, 383-439. Oxford: Blackwell.

Cinque, G. 1994. On the evidence for partial N movement in the Romance DP. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (eds.), *Paths towards universal grammar*, 85-110. Georgetown: Georgetown University Press.

Cinque, G. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, G. 2002. *Functional structure in DP and IP: The cartography of syntactic structures, Vol. 1*. Oxford: Oxford University Press.

Cinque, G. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3), 315-332.

Cinque, G. 2006. The dual source of adjectives and phrasal movement in the Romance DP. Ms. University of Venice. (Revised ms. to appear in *Linguistic Inquiry*.)

Crisma, P. 1991. *Functional categories inside the noun phrase: A study on the distribution of nominal modifiers*. B.A. Thesis, University of Venice.

Cruttenden, A. 1997. *Intonation* (2nd edition). Cambridge: Cambridge University Press.

Delsing, L.-O. 1993. *The internal structure of noun phrases in the Scandinavian languages*. Doctoral dissertation, University of Lund.

Edwards, J. & M. Beckman. 1988 Articulatory timing and the prosodic interpretation of syllable duration. *Phonetica* 45, 156-174.

Embick, D. & R. Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32(4), 555-595.

Frascarelli, M. 2000. *The syntax-phonology interface in focus and topic constructions in Italian.* (Studies in Natural Language and Linguistic Theory 50). Dordrecht: Kluwer.

Frascarelli, M. To appear. Subjects, topics and the interpretation of referential pro. An interface approach to the linking of (null) pronouns. *Natural Language and Linguistics Theory*.

Ghini, M. 1993. M-formation in Italian: a new proposal. *Toronto Working Papers in Linguistics* 12, 41-78.

Halle, M. & J.-R. Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Hammond, M. 1984. *Constraining metrical theory: A modular theory of rhythm and destressing*. Ph.D. dissertation, University of California .

Hayes, B. 1988. Final lengthening and the prosodic hierarchy. *Journal of the Acoustical Society of America* 84.

Hayes, B. 1995. *Metrical stress theory*. Chicago: University of Chicago Press.

Hellmuth, S. To appear. Prosodic weight and phonological phrasing in Cairene Arabic. *Chicago Linguistic Society (CLS)* 40.

Kayne, R.S. 1999. Prepositional complementizers as attractors. *Probus* 11(1), 39–73.

Kayne, R.S. 2000. A note on prepositions, complementizers, and word order universals. In R.S. Kayne (ed.), *Parameters and universals*, 314-326. Oxford: Oxford University Press.

Kayne, R.S. 2002. On some prepositions that look DP-internal: English *of* and French *de*. *Catalan Journal of Linguistics* 1, 71-115.

Kayne, R.S. 2005. Some notes on comparative syntax, with special reference to English and French. In G. Cinque & R.S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 3–69. Oxford: Oxford University Press.

Kishimoto, H. 2000. Indefinite pronouns and overt N-raising. *Linguistic Inquiry* 31(3), 557-566.

Knittel, M.L. 2005. Some remarks on adjective placement in the French NP. *Probus* 17(2), 185–226.

Laenzlinger, C. 2000. French adjective ordering: Perspective on DP-internal movement types. *Generative Grammar in Geneva* 1, 55-104.

Larson, R.K. & F. Marušič. 2004. On indefinite pronoun structures with APs: Reply to Kishimoto. *Linguistic Inquiry* 35(2), 268-287.

Longobardi, G. 1994. Reference and proper names: A Theory of N-movement in Syntax and Logical Form. *Linguistic Inquiry* 25, 609-665.

Longobardi, G. 1996. The syntax of N-raising: A minimalist theory. *UiL OTS Working Papers in Theoretical Linguistics N. 96_005*. University of Utrecht: Utrecht Institute of Linguistics.

Longobardi, G. 2001. The structure of DPs: Some principles, parameters and problems. In C. Collins & M. Baltin (eds.), *The handbook of contemporary syntactic theory*, 562-601. Oxford: Blackwell.

Longobardi, G. 2005. Toward a unified grammar of reference. *Zeitschrift für Sprachwissenschaft* 24(1), 5-44.

McCarthy, J. & A. Prince. 1993. Generalized Alignment. In G. Booij & J. van Marle (eds.), *Yearbook of Morphology 1993*, 79-153. Dordrecht: Kluwer.

Nespor, M. & I. Vogel. 1986. *Prosodic phonology* (Studies in Generative Grammar 28). Dordrecht: Foris.

Pereltsvaig, A. 2006. Passing by cardinals: In support of head movement in nominals. *Proceedings of FASL 14: The Princeton Meeting*, 277-292.

Post, B. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.

Prince, A. & P. Smolensky. 1993. *Optimality theory. Constraint interaction in generative grammar*. RuCCS Technical Report #2, Rutgers University Center for Cognitive Science.

Prince, A. & P. Smolensky. 2004. *Optimality theory. Constraint interaction in generative grammar*. Oxford: Blackwell.

Rizzi, L. 1997. The fine structure of the left periphery. In L. Haegeman (ed.), *Handbook in generative syntax*, 281-337. Dordrecht: Kluwer.

Rizzi, L. 2004. Locality and left periphery. In A. Belletti (ed.), *Structures and beyond. The cartography of syntactic structures, Vol. 3*, 223-251. Oxford: Oxford University Press.

Rizzi, L. 2006. On the form of chains: Criterial positions and ECP effects. In L. Cheng & N. Corver (eds.), *Wh movement: Moving on*, 97-133. Cambridge, MA: MIT Press.

Rutkowski, P. & L. Progovac. 2006. Classifying adjectives and noun movement in Lithuanian. In C. Yim (ed.), *Minimalist views on language design: Proceedings of the 8th Seoul international conference on generative grammar*, 265-277. Seoul: Hankook/ Korean Generative Grammar Circle.

Samek-Lodovici, V. 2005. Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23(3), 687-755.

Selkirk, E. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3, 371-405.

Selkirk, E. 1995. Sentence prosody: Intonation, stress, and phrasing. In J.A. Goldsmith (ed.), *The handbook of phonological theory*, 550-569. Oxford: Blackwell.

Selkirk, E. 2000. The interaction of constraints on prosodic phrasing. In M. Horne (ed.), *Prosody: Theory and experiment*, 231-261. Dordrecht: Kluwer.
Selkirk, E. 2004. Bengali intonation revisited. In C. Lee, M. Gordon, & D. Büring (eds.), *Topic and focus: A crosslinguistic perspective*, 217-246. Dordrecht: Kluwer.
Shlonsky, U. 2004. The form of Semitic noun phrases. *Lingua* 114(12), 1465–1526.
Sproat, R. & C. Shih. 1990. The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (eds.), *Interdisciplinary approaches to language: Essays in honor of S. Y. Kuroda*, 565–593. Dordrecht: Kluwer.
Truckenbrodt, H. 1995. *Phonological phrases: Their relation to syntax, focus, and prominence*. Ph.D. dissertation, MIT.
Truckenbrodt, H. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2), 219-255.
Umeda, N. & A.M.S. Quinn. 1981. Word duration as an acoustic measure of boundary perception. *Journal of Phonetics* 9(1), 19-28.
Vaissière, J. 1983. Language-independent prosodic features. In A. Cutler & D.R. Ladd (eds.), *Prosody: Models and measurements*, 53-66. Berlin: Springer.
Valois, D. 1991. *The internal syntax of DP*. Ph.D. dissertation, University of California.
Wightman, C., S. Shattuck-Hufnagel, M. Ostendorf & P. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91, 1707-1717.
Willis, D. 2006. Against N-raising and NP-raising analyses of Welsh noun phrases. *Lingua* 116(11), 1807–1839.
Zamparelli, R. 1995. *Layers in the Determiner Phrase*. Ph.D. dissertation, University of Rochester.

Eliding the Noun in Close Apposition, or Greek Polydefinites Revisited

MARIKA LEKAKOU* & KRISZTA SZENDRŐI†

Abstract

In this paper we propose a treatment of Greek polydefinites as an instance of close apposition, as in *Burns the poet*. We argue that like close appositives, Greek polydefinites consist of two DPs, the only difference being that one of them contains noun ellipsis. We propose that both polydefinites and close appositives involve a process of Referential-role identification, in the spirit of the proposal by Higginbotham (1985) for theta-role identification in cases of adjectival modification. We show that our proposal can shed light on the ordering freedom of polydefinites, their discourse properties, the kind of adjectives that can appear in the construction, as well as the lack of polyindefinites.

1 Introduction

This paper discusses Greek polydefinites, i.e. combinations of an adjective and a noun where each features its own determiner, as in (1).

(1) a. to spiti to megalo
the house the big
b. to megalo to spiti
the big the house

Polydefinites co-exist in the language with monadics like (2), i.e. modification structures where only one determiner is present—although polydefinites have special semantic and pragmatic properties (see Kolliakou (2004); Campos and Stavrou (2004), and also section 4 of this paper).[1]

(2) to megalo spiti
the big house

*Meertens Instituut, Amsterdam. Email: marika.lekakou@meertens.knaw.nl

†University College London. Email: k.szendroi@ucl.ac.uk

[1]The terms ‘polydefinite’ and ‘monadic’, which we use throughout the paper, are both due to Kolliakou (2004).

Our proposal is to treat polydefinites as a case of close apposition, as in (3) from Greek and (4) from English:

(3) a. o aetos to puli (Stavrou, 1995)
the eagle the bird
b. to puli o aetos
the bird the eagle

(4) a. Burns the poet
b. the poet Burns

The affinity of polydefinites to close appositives has been noted in passing by a number of authors (Stavrou, 1995; Kolliakou, 2004; Panagiotidis, 2005), but it has not been exploited systematically. We propose that they are structurally very similar. What makes the polydefinite, which contains (overtly) one adjective and one noun, parallel to close appositives, which contain two nouns, is the fact that polydefinites contain noun ellipsis (see also Panagiotidis (2005)). We argue that these two key properties of polydefinites—the appositive relationship between their subparts, and the ellipsis site inside one of them—provide the answer to the following questions: (i) what is the structure of the polydefinite construction (ii) why are there no polyindefinites and (iii) what kind of adjectives can partake in the construction.

The paper is structured as follows. In section 2 we discuss the characteristics of close and loose appositives and suggest an analysis for the former in terms of R(eferential)-role identification. In section 3 we show that this analysis also applies to polydefinites in Greek. We discuss how the proposal treats the word order patterns that polydefinites give rise to, and the absence of poly*in*definites. In section 4 we consider what the restriction is on the adjectives that can partake in the construction, and what the pragmatic characteristics of polydefinites are. We show that the focus-like effects of polydefinites need not be taken to motivate a DP-internal focus position, as has been assumed by most authors: they follow straightforwardly from the properties ellipsis is known to have. Section 5 concludes the paper.

2 On apposition

2.1 The distinction between close and loose appositives

The literature on apposition distinguishes between close and loose apposition, exemplified in (5):[2]

[2]In the literature we find a variety of terms to refer to close vs. loose apposition , for instance restrictive vs. non-restrictive apposition (particularly in connection to the parallels between nominal appositives and relative clauses), integrated vs. supplementary appositives (Huddleston and Pullum, 2002; Potts, 2005), etc. We retain the terms close and loose apposition. Moreover, a number of different terms have been used to refer to the two sub-parts of appositives. In particular, for many authors 'host' or 'anchor' designates the (linearly) first subpart, and 'appositive', 'apposition' or 'supplement' designates

(5) a. Burns the poet (close apposition)
b. Burns, the poet (loose apposition)

Several differences have been noted in the literature between the two types of apposition in (5) (see among others Burton-Roberts (1975); Espinal (1991); Meyer (1992); McCawley (1998); Acuña–Fariña (1999); Huddleston and Pullum (2002); Keizer (2005); Potts (2005) for English, and Stavrou (1995) for Greek)—although we should point out that much more attention has been paid to loose than to close apposition. We will focus on the differences that seem most relevant.

The most widely-used diagnostic to tell close from loose apposition concerns the intonational properties of the two constructions. The two elements partaking in close apposition belong to a single intonational unit. Loose apposition, by contrast, involves an intonational pause between its two sub-parts. This property is reflected in orthography by means of a comma, as shown in (5), a dash or parentheses. Since loose apposition comprises two separate prosodic units, it is possible for each unit to feature its own stress. By contrast, in close apposition there can only be one stress assigned. (In English close appositives, main stress falls on the rightmost element, which is neutral stress assignment in this language.)

Given the presence of a prosodic boundary, it is not surprising that in loose apposition the two parts can be separated by expressions like *namely*, *that is (to say)*, *or rather*, *in other words* etc. As expected, this is impossible in the case of close apposition:[3]

(6) a. the head of department, namely Prof. Todorov
b. * Burns namely the poet

In fact, while nothing can intervene between the two parts in a close appositive, the two parts of a loose appositive need not even be adjacent:

(7) a. [The two dominical sacraments] stand out from all the rest – namely [baptism and Holy Communion]. (Huddleston and Pullum, 2002)
b. I met [the new head of department] the other day, [Prof. Todorov].

---

the second one. Unfortunately these terms have been coined with loose apposition in mind and do not reflect the ways in which the latter is different from close apposition, which is our primary focus here. We therefore refrain from adopting these terms and use the somewhat awkward 'first' and 'second' subpart/unit/nominal constituent/DP.

[3]To be more precise, Huddleston and Pullum (2002) discuss a distinction between specifying and ascriptive apposition (which incidentally corresponds to *epeksegesis* and *parathesis* of Greek traditional grammmars). *Namely* can actually occur in specifying, but not ascriptive loose apposition.

(i) a. The first constestant, (namely) Lulu, was ushered on stage. (specifiying)
b. Kim Jones, (??namely) a quite outstanding student, won a scholaship to MIT. (ascriptive)

We believe that these two kinds of loose apposition should receive the same treatment, however we will for the most part ignore this distinction.

Of particular interest to the syntax of these constructions is the fact that close apposition necessarily involves a relationship between two DPs, whereas any two categories can come together under loose apposition:

(8) a. He [$_{V}$ ate], or rather [$_{V}$ devoured], the whole pie. (adapted from Stavrou (1995))
b. It was [$_{PP}$ at about 7 o'clock], [$_{PP}$ just before sunset], that they left. (Burton-Roberts, 1975)
c. When the patient closed his eyes, he had absolutely no [$_{A}$ spatial] (that is, [$_{A}$ third-dimensional]) awareness whatsoever. (Huddleston and Pullum, 2002)
d. The goal is to produce individuals who not only [$_{TP}$ possess 'two skills in one skull'], that is, [$_{TP}$ are bicultural], but can also act as human links between their two cultures. (Huddleston and Pullum, 2002)
e. [$_{IP}$ John was speechless], I mean, [$_{IP}$ he was really surprised].

(9) a. * He [$_{V}$ ate] [$_{V}$ devoured] the whole pie.
b. * It was [$_{PP}$ at about 7 o'clock] [$_{PP}$ just before sunset] that they left.
c. * When the patient closed his eyes, he had absolutely no [$_{A}$ spatial][$_{A}$ third-dimensional] awareness whatsoever.
d. * The goal is to produce individuals who not only [$_{TP}$ possess 'two skills in one skull'] [$_{TP}$ are bicultural], but can also act as human links between their two cultures.
e. * [$_{IP}$ John was speechless] [$_{IP}$ he was really surprised].

Finally, and most crucially for our purposes, the two constructions differ with regards to their referential properties. In (nominal) loose apposition, the first nominal constituent picks out a unique entity and the second one provides supplementary information about that entity.[4] In close apposition, on the other hand, reference to a unique entity is determined by the two DPs together. This is the reason that belies the contrast in (10), taken from Potts (2005). We return to this particular difference between close and loose apposition presently.[5]

(10) a. Armstrong, the Texan, is a cyclist. #Armstrong, the Ohioan, is an astronaut.
b. Armstrong the Texan is a cyclist. Armstrong the Ohioan is an astronaut.

[4]This applies to ascriptive loose apposition. The reverse would hold in the case of specifying loose apposition.

[5]The contrast in (10) also relates to the restrictive nature of close apposition. The second subpart of loose appositives can never be restrictive, while the second subpart of close appositives (normally) must be restrictive. We will return to this issue when we discuss the set of admissible adjectives in polydefinites in section 4.

In view of the characteristics of loose apposition reviewed above, it seems that the two sub-parts do not stand in a tight syntactic relationship. There can certainly be no selectional requirement between them, since they can be of any category and since they can be separated from each other. In other words, the second element seems to behave like an adjunct, a parenthetical, or even a non-integrated constituent (see Dehé and Kavalova (2007) for references, and Potts (2005) for a recent analysis). It is thus not surprising that nominal loose appositives like (11a) have been treated on a par with supplements/interpolations/appendages, such as the rest of the examples in (11) (Huddleston and Pullum, 2002):

(11) a. Pat – the life and soul of the party – had invited all the neighbours.
b. The best solution, it seems to me, would be to readvertise the position.
c. Jill sold her internet shares in January – a very astute move.
d. Jill – and I don't blame her – left before the meeting had ended.

We follow Dehé and Kavalova (2007) and Ackema and Neeleman (2004) in considering loose apposition a parenthetical structure (contra de Vries (2006))—though ultimately the analysis of loose apposition is immaterial to what we say about polydefinites and close apposition.

If loose appositives are taken to be parentheticals, their intonational properties follow straightforwardly. Furthermore, this kind of approach to loose apposition is consonant with what seems to be its semantic contribution. Doron (1992, 1994) and Potts (2005) provide several arguments for the claim that the second part of a loose appositive is semantically a predicate nominal and not a referential nominal. This claim captures the fact that the second part of a loose appositive does not pick out a referent, but simply provides a supplementary description for the entity referred to. What does pick out a referent in a nominal loose appositive is its first sub-part.[6]

As noted by Kolliakou (2004) and many others, the first NP of a nominal loose appositive has a uniqueness presupposition. Consider, for instance, the sentences in (12). Example (12a), a loose appositive, is appropriate, because the name 'Guillem' picks out a unique individual in the world. By contrast, the close appositive in (12b) is not a felicitous continuation of the sentence in (12).

(12) Tonight I will speak of a great French artist.
a. Guillem, the dancer, ...
b. # Guillem the dancer ... .

Not only do close appositives lack a uniqueness presupposition for their first subpart; what picks out a unique referent is the construction as a whole, i.e. both subparts

[6]Cf. the generalization that Potts (2005: 132) offers:

(i) An expression *E* can appear as the predicate in a predicative copular construction if and only if *E* can appear in an NA's [non-integrated appositive's, ML & KS] appositive position.

jointly contribute to reference. This means that the close appositive *Guillem the dancer* can only be felicitous in a context where more than one individual in the world bears the name 'Guillem', and only one of them is a dancer. The following example from Kolliakou (2004: 274-275) gives such a context. In this case we observe the exact opposite pattern: the loose appositive is infelicitous because the uniqueness presupposition associated with its first part is not satisfied, whereas the close appositive is fine:

(13) Tonight I will speak of the Van Gogh brothers, the painter and the critic.
  a. Van Gogh the painter ... .
  b. # Van Gogh, the painter, ...

2.2 Close apposition in terms of R-role identification

Since in close apposition both parts contribute to the determination of reference, neither is a predicate; the two nominal parts are both referential. This means that the predicate-NP analysis that Doron (1992, 1994) pursues for the second subpart of loose appositives does not extend to close apposition. Indeed, we do not believe that close appositives involve a subject-predicate relation (contra for instance Panagiotidis (2005)). Rather, we suggest that both DPs involved in the construction are referential DPs. We take this to mean that both DPs have an R(eferential) role in the sense of Williams (1981, 1989); Higginbotham (1985); Zwarts (1992); Baker (2003). Let us briefly see what function R-roles fulfill in nominals.

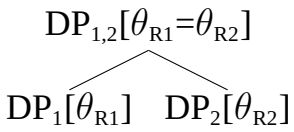
In line with the aforementioned authors, nominal elements have, in addition to other thematic roles they discharge, a referential role, the R-role, which is their external theta-role. The R-role is what enables a nominal element to act as a referential argument.[7] In Williams' system, which we adopt, when a nominal occupies an argument position, its R-role is bound by a thematic role of the selecting predicate, whereas when the nominal occurs as a predicate, it assigns the R-role to its subject.[8] We thus follow Williams (1989) and Baker (2003), contra Higginbotham (1985), in assuming that R-roles are not automatically saturated by (definite) determiners. Rather, the R-role survives until the topmost DP-layer of the nominal projection, where it gets bound by one of the theta-roles of the verb (in the case of argument nominals).

We propose that in close apposition an operation takes place which identifies the R-roles of two DPs. This operation can be thought of as complex argument formation. The proposal is schematically illustrated in (14).

[7] In Baker (2003), nouns come with a referential index, which is the syntactic correlate of the semantic fact that only nouns come with identity criteria.

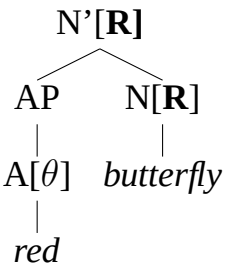
[8] Although the R-role is responsible for the (discourse-)referential properties of nominals, discourse reference is assigned outside the grammar, so that the syntax is independent of the system that ultimately assigns reference to nominal constituents.

(14)



Theta-identification is not new, and neither is identification that involves the R-role. Higginbotham (1985) proposes that theta-role identification is what happens in adjectival modification, where the R-role of the noun gets identified with a theta role of the adjective, as depicted in (15):

(15)

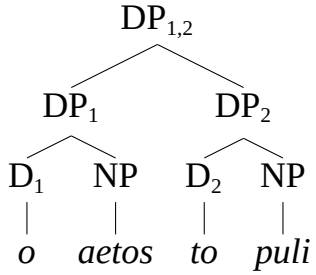


An important aspect of Higginbotham's proposal for theta-identification is that, semantically, it corresponds to the intersection of the set denoted by the noun and the set denoted by the adjective. So 'a big buttefly' is a thing that is big and a butterfly. See also Heim and Kratzer (1998) for predicate modification as set intersection.

Applied to close appositives, theta-identification amounts to identification of two R-roles. This creates a syntactically symmetric structure, as illustrated in (16b):

(16) a. o aetos to puli
the eagle the bird

b.



An interesting characteristic of the structure in (16b) is that it is multi-headed: the highest DP is a member of the projection lines of both DP_1 and DP_2. Multi-headed structures are argued for by Baker and Stewart (1999), who deal with serial verb constructions in terms of multi-headed verbal projections, namely multi-headed VoicePs, vPs and VPs. Moreover, multi-headed structures are employed by Neeleman and van de Koot (2002) for secondary predication. In particular in the framework of Neeleman and van de Koot (2002) (a) categorial features are copied up from daughter nodes to mother nodes in the projection line and (b) any identical features that are copied onto a node get identified (two identical features collapse into one). It

follows then that multi-headed structures like (16b) must involve phrases of the same category, as the categorial features would otherwise clash on the highest node.

Consider the case of secondary predication, as analyzed by Neeleman and van de Koot (2002). The following is an example from Dutch:

(17) dat Jan_i $Marie_j$ naakt $ontmoette_{i/j}$.
that Jan Marie naked met
'that Jan met Marie naked.'

One may reasonably ask why there is no violation of the Theta Criterion in secondary predication, or else, how can it be that two predicates can discharge their theta-roles in the presence of a single DP? If the argument of 'naked' gets identified with either the internal or the external argument of the verb 'meet', both predicates (the verbal and the adjectival one) can discharge their theta-roles in the presence of a single DP. A similar reasoning applies to (14): even though both DP_1 and DP_2 are referential and thus potential arguments, by identification of their R-roles it is the highest DP alone that acts as an argument.

Two welcome predictions are generated within this treatment of close apposition. Since close apposition involves R-role identification, it follows that only nominal phrases can be part of a close appositive, since only nominal elements have an R-role. Recall that indeed only nominal constituents can be brought together under close apposition, in contrast to loose apposition, which can involve any two (identical) categories.

Furthermore, since close apposition involves R-role identifcation, we predict that close apposition will fail when it involves two DPs which are independently identical in reference (i.e. when the two R-roles are already identical). This is indeed the case, as mentioned in Stavrou (1995).[9] (As expected, loose apposition is not subject to this restriction, cf. (18c) and (19c).)

(18) a. * i sikaminja i murja
the blueberry tree the mulberry tree
(Stavrou, 1995)
b. i sikaminja to dendro
the blueberry tree the tree
c. i sikaminja, (diladi) i murja
the blueberry tree namely the mulberry tree

(19) a. * Shakespeare the Bard
b. Shakespeare the poet
c. Shakespeare, the Bard

[9] *Sikaminja* is a dialectal synonym of *murja* (Stavrou 1995:225). The examples are glossed as in the original source, namely Stavrou (1995).

3 Greek polydefinites

We now turn to polydefinites in Greek. We first provide a description of the data and the questions they pose. We then propose an analysis of polydefinites in terms of close apposition, and show how the proposal derives the core properties of the construction.

3.1 The data

Polydefinites are constructions in which a definite DP has two subparts: one is a regular definite DP, the other is an adjective accompanied by an additional definite determiner. (The phenomenon is also referred to as 'determiner spreading', to reflect that there are as many 'extra' determiners as there are adjectives present.) One of the most well-known properties of polydefinites in Greek is the ordering freedom that they exhibit, see (1) repeated below as (20). The fact that in polydefinites the adjective can either precede or follow the nominal DP contrasts with the case of monadic definites, where the adjective necessarily precedes the noun it modifies, as shown in (21).

(20) a. to spiti to megalo
the house the big
b. to megalo to spiti
the big the house

(21) a. to megalo spiti
the big house
b. * to spiti megalo
the house big

The ordering freedom of constituents in a polydefinite persists even when more than one adjective is present, as illustrated in (22) (cf. Panagiotidis (2005)).

(22) a. to megalo to petrino to spiti
the big the stone the house
b. to megalo to spiti to petrino
the big the house the stone
c. to spiti to megalo to petrino
the house the big the stone
d. to spiti to petrino to megalo
the house the stone the big
e. to petrino to spiti to megalo
the stone the house the big
f. to petrino to megalo to spiti
the stone the big the house

As noted by Panagiotidis (2005), however, there is a restriction on this freedom: an adjective (still) has to precede the noun if it is not preceded by a determiner. The data are given in (23):

(23) a. * to spiti megalo to petrino
the house big the stone
b. * to spiti to megalo petrino
the house the big stone

A second property of polydefinites often noted in the literature is that not all adjectives can take part in it. According to Alexiadou and Wilder (1998), Kolliakou (2004) and many others, non-intersective adjectives such as 'alleged' are illicit, see (24).[10]

(24) i ipotithemeni (*i) tromokrates
the alleged the terrorists

Alexiadou and Wilder (1998) note the deviance of the polydefinite in (25) and propose a more general restriction stated in (26).

(25) i italiki (*i) isvoli
the italian the invasion

(26) An adjective permits determiner spreading only if it can be used predicatively.

Finally, Kolliakou points out that pragmatically non-restrictive adjectives are inadmissible in the polydefinite construction. In (27), *dilitiriodis* 'poisonous' is non-restrictive when applied to cobras, since as a matter of world knowledge there are no non-poisonous cobras. The polydefinite cannot be used under such circumstances.

(27) a. Idame tis dilitiriodis kobres.
saw.1PL the-PL.ACC poisonous-PL. ACC cobras-PL.ACC
'We saw the poisonous cobras.'
b. # Idame tis dilitiriodis tis kobres.
saw.1PL the-PL.ACC poisonous-PL.ACC the-PL.ACC cobras-PL.ACC

A final property of polydefinites is that, as the name suggests, they arise with definite determiners. As shown in (28), poly*in*definites do not occur in Greek—the *in*definite determiner cannot spread:

(28) a. ena megalo (*ena) spiti
a big a house
b. ena spiti (*ena) megalo
a house a big

In the next subsection we provide an analysis of polydefinites whose core ingredients are the appositive relationship between the two sub-parts, and the nominal ellipsis contained in the 'adjectival' part. In subsection 3.3 we show how the ordering freedom follows from the proposed analysis. In subsection 3.4 we discuss the

[10]The example is starred in Alexiadou (2006). However, as we show in section 4, (at least some) non-intersective adjectives can be felicitously used in polydefinite constructions, in appropriately manipulated discourse contexts. We will therefore contest the ungrammaticality of examples like (24), and also (25).

lack of polyindefinites. The set of admissible adjectives interacts with the pragmatics of the construction; both these issues are taken up in section 4.

3.2 Polydefinites are close appositives

Our proposal is that polydefinites are an instance of close apposition. Polydefinites are only special in that they involve an elided noun in one of their DP-subparts:

(29) a. [$_{DP}$ [$_{DP}$ to the megalo ∅] [$_{DP}$ to the spiti]]
big house

b. [$_{DP}$ [$_{DP}$ to the spiti] [$_{DP}$ to the megalo ∅]]
house big

One piece of syntactic evidence that the two DPs form a DP constituent is that the two DPs share Kase and a selecting Preposition (see (30) and (31)). Nevertheless, the two DPs are semantically and syntactically fully formed and there do not seem to be any arguments (contra Panagiotidis 2005) that what is involved is an asymmetric syntactic structure.[11]

(30) [$_{KP}$ K [$_{DP}$ [$_{DP}$ tu palju] [$_{DP}$ tu spitiu]]]
the-GEN old-GEN the-GEN house-GEN

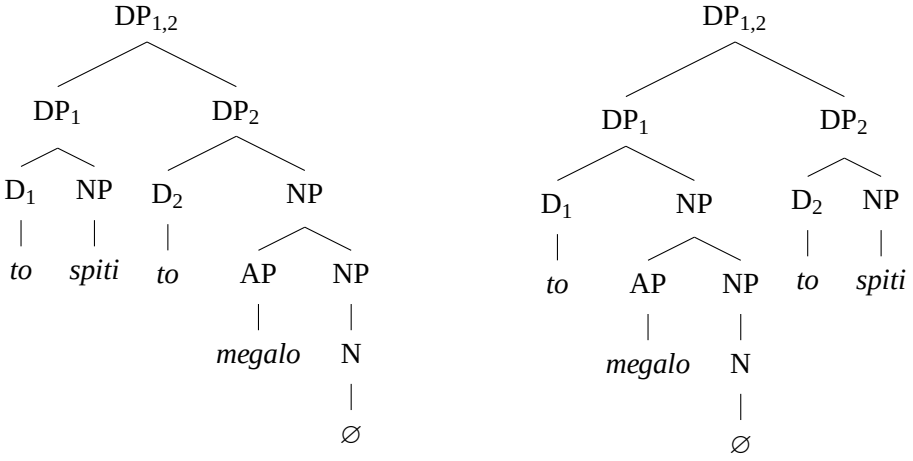
(31) a. [$_{PP}$ P me [$_{KP}$ K[$_{DP}$ [$_{DP}$ to kokino] [$_{DP}$ to podhilato]]]]
with the-ACC red-ACC the-ACC bicycle-ACC

b. * [$_{PP}$ P me [$_{KP}$ K[$_{DP}$ [$_{DP}$ to kokino] [$_{PP}$P me [$_{DP}$to podhilato]]]]]
with the-ACC red-ACC with the-ACC bicycle-ACC

The structure we assign to polydefinites is given in (32):

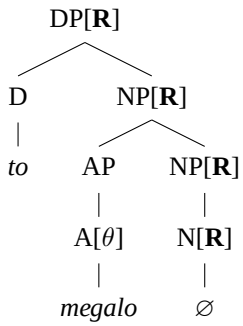
[11]Our proposal is similar to Panagiotidis (2005) in two ways: like him, we believe that there is an ellipsis site inside one of the DPs, and like him we believe that polydefinites instantiate a DP whose subparts are also DPs. This allows Panagiotidis too to draw a parallel with close appositives, which however he does not discuss in much detail. The crucial difference relates to the structure he proposes. According to him, close appositives and polydefinites involve a subject-predicate structure inside the DP (which thus resembles a small clause): the leftmost element is in the specifier of the larger DP and the subject of predication, and the rightmost element is the predicate. The 'adjectival' DP and the 'nominal' DP can occupy either position. However, Panagiotidis brings no syntactic arguments to support this structure for polydefinites/close appositives, and we doubt the validity of his semantic arguments.

(32)

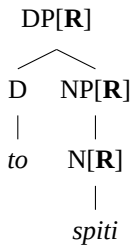


Let us see the workings of R-role identification in polydefinites in more detail. Higginbotham's theta-identification first takes place between the adjective and the null noun. This is illustrated in (33a). The resulting DP then undergoes R-role identification with the DP which contains the lexically realized noun, i.e. the DP in (33b). The result is (34).

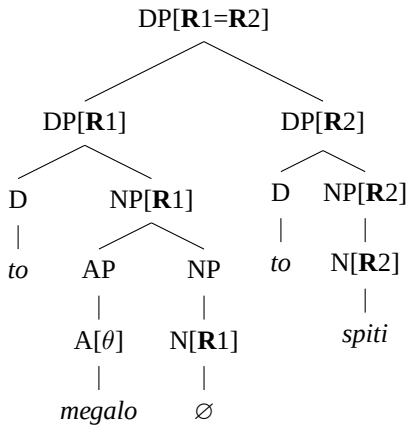
(33) a.



b.



(34)



3.3 Deriving the word order pattern

As already mentioned polydefinites display a freedom in word order which is not available in the case of monadic definites:

(35) a. to spiti to megalo
the house the big

b. to megalo to spiti
the big the house

(36) a. * to spiti megalo
the house big

b. to megalo spiti
the big house

Since we analyze polydefinites as an instance of close apposition, we expect the same ordering freedom in nominal appositives as well. This is indeed the case:

(37) a. o aetos to puli
the eagle the bird

b. to puli o aetos
the bird the eagle

(38) a. Burns the poet

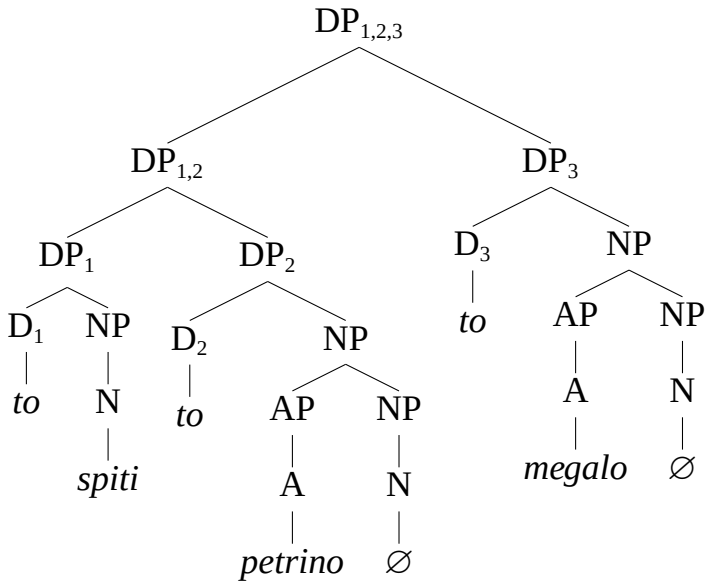
b. the poet Burns

The symmetric structure we propose for polydefinites/close appositives is perfectly consistent with their ordering freedom. Since the structure we propose is multi-headed, i.e. the two DPs are sisters, they can appear in either order.

Recall also that the ordering freedom persists when more than one adjective is present, so that the ordering possibilities multiply accordingly. Structurally, this means that we iterate R-role identification. The tree in (39b), for instance, represents the structure of (39a).

(39) a. to spiti to petrino to megalo
the house the stone the big

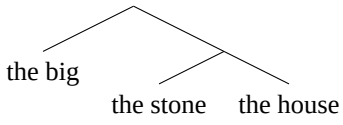
b.



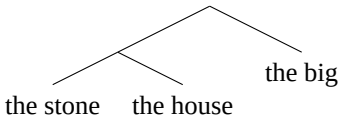
Since the ordering within the appositive/polydefinite is free, we can (a) permute the order of DP_3 with respect to $DP_{1,2}$, (b) permute the order within $DP_{1,2}$ prior to combination with DP_3, and (c) change the order with which each adjective combines with the noun. The full paradigm is repeated in (40). The corresponding tree structures are given in a more abstract form in (41).

(40) a. to megalo to petrino to spiti
the big the stone the house
b. to petrino to spiti to megalo
the stone the house the big
c. to megalo to spiti to petrino
the big the house the stone
d. to spiti to petrino to megalo
the house the stone the big
e. to petrino to megalo to spiti
the stone the big the house
f. to spiti to megalo to petrino
the house the big the stone

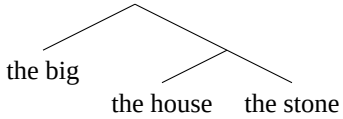
(41) a.



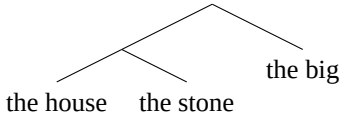
b.



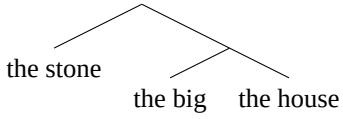
c.



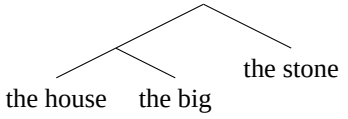
d.



e.



f.



As mentioned in the previous subsection the orderings with multiple adjectives obey the following generalization: if the adjectives follow the noun, they have to be preceded by a determiner (Panagiotidis, 2005):

(42) a. * to spiti megalo to petrino
the house big the stone
b. to megalo spiti to petrino
the big the house the stone

(43) a. * to spiti to megalo petrino
the house the big stone
b. to megalo petrino to spiti
the big stone the house

The explanation we have for these data is similar to the one provided in Panagiotidis (2005). In both the ungrammatical examples (42a) and (43a) we have two DPs, one of which contains an independently illicit structure. As far as (42a) is concerned, recall that in Greek monadic definites the adjective precedes and can never follow the noun:

(44) * to spiti megalo
the house big

However, this is not respected in (42a), which contains precisely this illicit DP. As for (43a), the DP containing the noun ellipsis features two adjectives, and that is independently disallowed (Panagiotidis, 2005).[12]

3.4 Why are there no poly*in*definites?

One question that has not received a satisfactory answer in previous treatments of polydefinites is the absence of poly*in*definites, cf. the examples in (45) repeated from earlier:[13]

(45) a. ena megalo (*ena) spiti
a big a house
b. ena spiti (*ena) megalo
a house a big

We believe that the present proposal can shed some new light on this issue. In particular, what is interesting from our perspective is that exactly the same (ill-understood)

[12]This cannot explain (43b), which Panagiotidis doesn't discuss. It seems to us to involve some sort of idiomatic, complex adjective. Note that changing the relative order of the two adjectives is not possible:

(i) * to petrino megalo to spiti
the stone big the house

[13]Ellipsis cannot be the problem with (45), because noun ellipsis is licensed in Greek with indefinite determiners, as well as with no determiner at all (Giannakidou and Stavrou, 1999; Panagiotidis, 2003).

restriction applies in the case of nominal close apposition: close apposition necessarily involves two definite DPs. As noted by Stavrou (1995), it is not possible for either one to be indefinite. (The examples in (46) are fine as loose appositives.)

(46) a. * o Nikos enas kathijitis
the Nikos a professor
b. * enas kathijitis o Nikos
a professor the Nikos

Since polydefinites are an instance of close apposition, the ban against indefinites is expected here too. On an approach that takes polydefinites to be an instance of close apposition, the lack of polyindefinites follows as a special case of this constraint. The exact nature of this constraint is a matter we leave for future research.[14]

4 Polydefinites in context

In this section we discuss the pragmatics of polydefinites and focus in particular on delimiting the set of admissible adjectives and on fleshing out the contribution of noun ellipsis.

4.1 The proper subset requirement

A well known fact about close apposition (see among others Kolliakou (2004: 274) for discussion and for references) is that there is a restrictive relation between the DPs that form its subparts. We refer to this constraint as the proper subset constraint:

(47) *The Proper Subset Constraint*
In a close appositive, the denotation of one of the DPs must be a proper subset of the denotation of the other DP.

Recall the following data, which contrast close to loose apposition:

(48) Tonight I will speak of a great French artist.
a. Guillem, the dancer, ...
b. # Guillem the dancer ... .

(49) Tonight I will speak of the Van Gogh brothers, the painter and the critic.

[14] One complicating factor is that it is not clear to us at this stage whether (46) illustrates a universal property of close apposition, or a peculiarity of Greek close apposition. In particular, although Greek disallows indefinites in close apposition, English has been claimed to be more liberal. Keizer (2005) gives the following example as a close appositive with an indefinite part:

(i) He has to put up with a soppy elder brother Robert who is forever moaning over some girl or other and a sister Ethel who has all the brisk no-nonsense superiority of a true Wodehouse gel.

a. Van Gogh the painter ... .
b. # Van Gogh, the painter, ...

In (48) our context involves only one person by the name 'Guillem', so 'dancer' cannot be restrictive on 'Guillem'. Therefore the close appositive fails. In (49) we introduce a set of two people, the Van Gogh brothers. One of them is a painter and the other is a critic. 'Painter' can thus act restrictively on 'van Gogh'. We may therefore felicitously use a close appositive to refer to one of them.

Given that we argue that polydefinites involve close apposition, the proper subset restriction should hold for polydefinites as well. That this is in fact the case can be illustrated, for instance, by considering again examples that contain pragmatically non-restrictive adjectives, like (50). (50) is infelicitous because the DP containing the adjective does not determine a proper subset of the denotation of the second DP; as we know, all cobras are poisonous. So, the reasoning behind the deviance of (50) is parallel to the reasoning behind the deviance of (48b).

(50) # i dilitiriodis i kobres
the poisonous the cobras

Similarly, Kolliakou (2004: 216-217) provides the following data to show that polydefinites are used in more restricted contexts than monadics. In particular, she shows that the example in (51), which involves a monadic, may be used in all the scenarios listed in (a)–(d), while (52), with a polydefinite, is only consistent with scenarios (a) and (b), i.e. only the situations in which the adjective is interpreted restrictively.

(51) O Yannis taise ta zoa. I mikres gates itan pinasmenes.
the Yannis fed the animals. the young cats were hungry
'Yannis fed the animans. The young cats were hungry.'
a. all the animals that Yannis fed were cats, but *there were young and non-young cats*
b. Yannis fed cats and non-cats, and *there were young and non-young cats*
c. all the animals that Yannis fed were cats, and there were only young cats
d. Yannis fed cats and non-cats, but all the cats were young ones

(52) O Yannis taise ta zoa. I mikres i gates itan pinasmenes.
the Yannis fed the animals. the young the cats were hungry
'Yiannis fed the animals. The young cats were hungry.'
a. all the animals that Yannis fed were cats, but *there were young and non-young cats*
b. Yannis fed cats and non-cats, and *there were young and non-young cats*

This is, of course, consistent with the idea that the adjectival part of the polydefinite must provide a restriction on the nominal part: the set of young cats is a proper subset of the set of cats in both (a) and (b), but not in (c) and (d).

4.2 The set of admissible adjectives

It follows from the proper subset constraint in (47) that only adjectives that can partition the noun-denotation of the non-elliptical DP will be able to appear in the polydefinite. In particular, it follows that the core cases will involve intersective adjectives. Intersective adjectives are those which allow the inference in (53a). An example is the adjective *good*, as seen in (53b):

(53) a. X is an Adjective Noun ⇒ X is Noun
b. X is a good pupil ⇒ X is a pupil

This property follows from our syntactic analysis of polydefinites involving complex argument formation/R-role identification: syntactically, by the identification of the R-roles, the two DPs form one argument; semantically, the denotation of the new DP will be the intersection of the sets denoted by the two sub-DPs.

But as we already saw in the previous section, the requirements on polydefinites are in fact somewhat stronger. It is not enough that the denotation of the polydefinite is obtained by set-intersection: adjectival modification in the polydefinite construction involves *restrictive* modification. This is stated in (47). We will argue presently that this relates to the ellipsis involved in polydefinites.

What is interesting to note is that manipulating the discourse context in the appropriate way can give us the desired effect even without an intersective adjective. The exceptional behaviour will arise whenever we can tamper with the potential of the noun denotation to be partitioned in disjoint subsets in a pragmatically plausible way. Leu (2007) points out that in some cases non-intersective non-predicative adjectives are acceptable:[15]

(54) O proigumenos o prothipurgos pethane.
the previous the president died
'The previous president died.

(54) is licit in a context where the speaker corrects another interlocutor who thought she overheard that the current president is dead. In this particular context, the noun denotation comprises two disjoint subsets, one containing the current president and the other the previous one. What makes the polydefinite available is that the mention of the current president in the previous discourse D-links the set of all Greek presidents in a salient way, and thus subsequent reference to the previous president satisfies the proper subset constraint in (47).

Similarly, (55) is possible in a context where the Dutch invasion is only one of the invasions endured.

[15]Leu seems to take *proigumenos* to mean 'former', though actually it corresponds to 'previous'. For reasons we do not at this point understand, *proin*, 'former', and *nin*, 'current', seem unable to partake in a polydefinite.

(55) i olandiki i isvoli mas kseklirise.
the dutch the invasion us wiped out
'The dutch invasion wiped us out.'

Note that examples such as (54) and (55) above undermine predicative analyses of polydefinites (Alexiadou and Wilder (1998); Campos and Stavrou (2004); Panagiotidis (2005) among others), since they involve non-predicative adjectives. What the examples above—and in particular the contexts that make them felicitous—suggest is that such dimensions as 'predicativity' are irrelevant. What matters is (the effect of the right context on) the potential of the noun denotation to be partitioned along the dimension contribution by the adjective, such that a restrictive interpretation of the adjective is possible.

4.3 Comparison with Kolliakou's Polydefiniteness Constraint

At this point it seems important to evaluate Kolliakou's own account for the unavailability of examples like (50) and the loss of readings in examples like (52). Kolliakou proposes that polydefinites are subject to a pragmatic constraint, her Polydefiniteness constraint, given in (56):

(56) *The Polydefiniteness Constraint* (Kolliakou, 2004, 273)
Greek polydefinites are unambiguously non-monotone anaphoric expressions: the discourse referent Y of a polydefinite is anaphoric to an antecedent discourse referent X, such that $Y \subset X$.

Given (56), a polydefinite is felicitous if: (i) it introduces in the discourse a proper subset of a given set and (ii) that set has been explicitly mentioned (and is highly salient/accessible).

Even though the constraint in (56) accounts for the data (though see the end of this subsection), we believe that Kolliakou's formulation has an important shortcoming. Although both polydefinites and close appositives more generally seem to be subject to the proper subset requirement (i.e. something like (i)), the second restriction (i.e (ii)) only seems to apply to polydefinites and not to close appositives in general. Examples like (57) repeated from above can occur in the absence of previous mention of a set of entities that properly includes eagles in (a) or a set of poets in (b).

(57) a. o aetos to puli
the eagle the bird
b. Burns the poet

We believe that the pragmatic restriction in (ii) can be derived from the only difference that we posit between polydefinites and close appositives, namely, that the former involves noun ellipsis. This is because, as Giannakidou and Stavrou (1999) have shown for Greek, noun ellipsis is subject to the recoverability condition of Hankamer and Sag (1976):

(58) An elided subconstituent α must recover its descriptive content by an antecedent γ previously asserted in the discourse. (Giannakidou and Stavrou, 1999, 307)

Given our claim that polydefinites involve nominal ellipsis, it is to be expected that this recoverability constraint also applies to them (which is in fact essentially what Kolliakou proposed).

So, treating the two pragmatic conditions separately allows for a transparent explanation for why they should hold: (i) holds because polydefinites involve close apposition and (ii) holds because they involve noun ellipsis. We will discuss noun ellipsis in particular in the following sections.

A different issue worth drawing attention to is that part (i) of Kolliakou's constraint differs from our own formulation of the proper subset requirement on close appositives. She states that a proper subset relation must hold between the set denoted by the polydefinite and *a previously mentioned set*, while we propose that a proper subset relation must hold between the set denoted by the polydefinite and *the set denoted by the 'nominal' DP*. The following subsections will make obvious in more detail how our account, in virtue of involving noun ellipsis, differs from hers.

4.3.1 The distribution of ellipsis and polydefinites

Positing nominal ellipsis in polydefinites does more than provide us with a transparent explanation of the pragmatic constraints of the construction. It also accounts for some interesting empirical facts.

As we would expect, there is considerable overlap between contexts that allow polydefinites and those that allow simple ellipsis. For instance, as Panagiotidis (2005) already noted, and as was mentioned earlier, ellipsis seems to be disallowed with more than one adjective. Crucially, this restriction is operative also in the case of polydefinites; compare (59a) with (59b).

(59) a. * to kenurjo kokino
the new red
b. * to kenurjo kokino to podilato
the new red the bicycle

But there are also discrepancies between the occurrence of polydefinites and that of simple ellipsis, which seem to us to be revealing. A context where one might expect a discrepancy between ellipsis and polydefinites is when the antecedent is not accessible enough to license simple ellipsis. This could be for instance because there is an intervener. In such cases, examplified in (60), the polydefinite is appropriate. This is not surprising as the nominal part of the polydefinite provides a way to recover the content of the elided noun.

(60) a. A. Boris na dialeksis mia asimena pena, mia xrisi pena ke
can-2SG SUBJ choose-2SG a silver pen a golden pen and

ena molivi. Ti dialejis ja ti Maria?
a pencil. what choose-2SG for the Maria
b. B. Dialego to molivi.
choose-1SG the pencil
c. A. Ke ja to Janni? Ti dialejis?
and for the Janni what choose-2SG
d. B. Dialego tin pena tin asimenia.
choose-1SG the pen the silver
e. B.# Dialego tin asimenia.
choose-1SG the silver

4.3.2 An asymmetry between polydefinites and close appositives

Recall that in the scenario with the feeding of the animals, repeated as (61) below, 'young' must be restrictive, while the set of cats need not be a proper subset of the set of animals mentioned in the previous discourse.

(61) O Yannis taise ta zoa. I mikres i gates itan pinasmenes.
the Yannis fed the animals. the young the cats were hungry
'Yiannis fed the animals. The young cats were hungry.'
a. all the animals that Yannis fed were cats, but there were young and non-young cats
b. Yannis fed cats and non-cats, and there were young and non-young cats

In other words, there seems to be an asymmetry in the polydefinite in that the proper subset requirement only applies to the adjectival part, while the set denoted by the nominal part is allowed so long as it is included in the denotation of the previously mentioned set (i.e. the set of animals, in this case), but it does not have to be a proper subset of it.

Note that no such asymmetry presents itself in the case of nominal appositives in general. Keizer (2005) gives examples of close appositives with either the definite DP or the proper name as the 'restrictive' DP:

(62) a. the actor Orson Welles
b. Orson Welles the actor

We believe that the asymmetry in Greek polydefinites follows from the fact that they involve noun ellipsis. We know at least since Williams (1997) (see also Giannakidou and Stavrou (1999) specifically for Greek noun ellipsis), that in the case of ellipsis the remaining non-elided elements must be informative (disanaphoric in Williams' terms). Thus noun ellipsis is licensed in (a), where the adjective is informative (or disanaphoric), but odd in (b), where it is not.

(63) a. I Maria forese to ble fustani ke i Eleni forese to prasino ∅.
the Maria wore the blue dress and the Eleni wore the green
[∅ = fustani]

b. # I Maria forese to ble fustani ke i Eleni forese to ble ∅.
the Maria wore the blue dress and the Eleni wore the blue
[∅ = fustani]

There are other ways an adjective can be uninformative, for instance, if it is pragmatically or inherently non-restrictive:

(64) a. # O Janis taise ta mikra zoa. Ta mikra itan pinasmena.
the Jannis fed the young animals. the young were hungry.
b. # Latrevo ta fronima pedia. Ta fronima ine evlojia theu.
adore-1SG the good children. the good are blessing God-GEN.
c. # Edo de tha vris kobres. I dilitiriodis zune alu.
here NEG FUT find-2SG cobras. the poisonous live elsewhere.
d. # I apopira dolofonias ixe stoxo ton proigumeno prothipurgo. O proigumenos den ine dimofilis.
the attempt murder-GEN had target the previous president. the previous NEG is popular

Since polydefinites involve noun ellipsis, polydefinites are also inappropriate if their adjectival part is uninformative in the above way. This is shown in (65).

(65) a. # O Janis taise ta mikra zoa. Ta mikra ta zoa itan pinasmena.
the Jannis fed the young animals. the young the animals were hungry.
b. # Latrevo ta fronima pedia. Ta fronima ta pedia ine evlojia theu.
adore-1SG the good children. the good the children are blessing God-GEN.
c. # Edo de tha vris kobres. I dilitiriodis i kobres zune alu.
here NEG FUT find-2SG cobras. the poisonous the cobras live elsewhere.
d. # I apopira dolofonias ixe stoxo ton proigumeno prothipurgo. O proigumenos o prothipurgos den ine dimofilis.
the attempt murder-GEN had target the previous president. the previous the president NEG is popular

To sum up what we have discussed so far: in a polydefinite construction, the denotation of one of the nominals must be a proper subset of the denotation of the other—this is because the polydefinite is a close appositive. That this part must be the adjectival part is due to the fact that the adjectival part has an independent requirement to be informative, since it involves noun ellipsis. This accounts for the asymmetry observed in (61).

4.4 Non-anaphoric or focused?

Most authors have argued that the adjectival part of the polydefinite construction is focused (see e.g. Kariaeva (2004), Ntelitheos (2004), Leu (2007) and many others). But we believe that the adjectival part is not obligatorily focused, just non-anaphoric, precisely as the noun ellipsis account we have been pursuing predicts. This is shown by the contrast in (66). In (66a) the polydefinite is licensed even in the absence of constrastive stress and an exhaustive interpretation, since the sentence 'the young cats were hungry' can be felicitously continued with 'as were the old ones'. So, one cannot maintain that the adjectival part is obligatorily focused in a polydefinite, since the effects that are usually associated with focus (e.g. exhaustive interpretation, contrastive stress) are not present. Now consider (66a), where the adjective in the polydefinite is contrastively stressed, and where it is impossible to continue with '... as were the old cats'. In other words, it is stress that provides the effects of focus, and not the polydefinite construction in and of itself.

(66) a. O Yannis taise ta zoa. I mikres i gates itan pinasmenes, opos
the Yannis fed the animals. the young the cats were hungry as
episis ke i megales (i gates).
also and the old the cats
'Yiannis fed the animals. The young cats were hungry, as were the old ones.'

b. O Yannis taise ta zoa. I MIKRES i gates itan pinasmenes, #
the Yannis fed the animals. the young the cats were hungry
opos episis ke i megales (i gates).
as also and the old ones

Another empirical problem with the claim that the adjective is necessarily focused in a polydefinite is noted by Kolliakou (2004:276). Her point is that the whole polydefinite itself can have other discourse functions. For instance in (66a), *i mikres i gates* 'the young the cats' is most likely a contrastive topic.

There are independent problems with analyses that try to derive the word order in the polydefinite construction with the help of a DP-internal designated Focus-position. As Lekakou (2000) argued for Greek and Szendrői (2001) for Hungarian and Italian, it is questionable whether a designated focus position exists even in the main clause projection line. In addition, Szendrői (2006) stresses that it is theoretically impossible to think of a focus-background partitioning DP-internally, as such notions are intrinsically propositional. Contrary to Schwarzschild (1999) and much subsequent work, we do not think that non-anaphoric is the same as focused. Rather, we follow Reinhart (2006) in assuming that givenness and focus are orthogonal notions. The adjectival part in a construction with noun ellipsis cannot be given or anaphoric, but at the same time it need not be focused.

5 Concluding remarks

In this paper we have argued for an approach to polydefinites in Greek on a par with close appositives. For both cases we have posited a mechanism of 'complex argument formation' via R-role identification, whereby the R(eferenctial) roles of two nominals get identified. In the case of polydefinites, one nominal is null, whereas in close appositives both are lexically filled. We have argued that our approach can successfully address the syntactic, semantic and pragmatic questions that polydefinites pose.

Since we capitalize on the affinity between polydefinites and close appositives, our account leads us to ask in a more pronounced way a bigger question: why do we seem to find polydefinites of the Greek type only in Greek, even though appositives are attested elsewhere (cf. Alexiadou (2006))? We do not have the definitive answer at this point; there seem to be several different factors which may prove relevant. For instance, Greek requires definite determiners with proper names, Greek requires spreading of ϕ-features throughout the DP, and so on. Although at this stage we cannot formulate a full answer that would explain the contrast between Greek and the other languages, we do have a direction to point towards as a way to resolve the contrast between English and Greek.

Keizer (2005) provides the following paradigm to illustrate the possible combinations in a close appositive:

(67) a. the actor Orson Welles
b. the word recession
c. my friend Orson Welles
d. my friend the actor
e. Orson Welles the actor
f. actor Orson Welles

The point is that, although in English too close appositives need to (or can in any event) be definite, it seems impossible to combine in a close appositive two definite DPs that are both headed by a definite determiner, as in (68).

(68) a. * the linguist the scholar
b. * the scholar the linguist

We believe that the answer to the ungrammaticality of (68) can inform us on why at least English lacks the Greek polydefinite. This direction becomes available only on an approach that treats polydefinites as an instance of close apposition.

References

Ackema, Peter, and Ad Neeleman. 2004. *Beyond Morphology*. Oxford: Oxford University Press.

Acuña–Fariña, Juan Carlos. 1999. On apposition. *English Language and Linguistics* 3:59–81.

Alexiadou, Artemis. 2006. On the cross-linguistic distribution of (in)definiteness spreading. Handout of talk given at OLT Syntax Workshop, Klagenfurt.

Alexiadou, Artemis, and Chris Wilder. 1998. Adjectival modification and multiple determiners. In *Possessors, Predicates and Movement in the DP*, ed. Artemis Alexiadou and Chris Wilder, 303–332. Amsterdam: John Benjamins.

Baker, Mark. 2003. *Lexical Categories. Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.

Baker, Mark, and Osamuyimen T. Stewart. 1999. On double-headedness and the anatomy of the clause. Ms. Rutgers University.

Burton-Roberts, N. 1975. Nominal apposition. *Foundations of Language* 13:391–419.

Campos, Hector, and Melita Stavrou. 2004. Polydefinites in Greek and Aromanian. In *Balkan syntax and semantics*, ed. Olga M. Tomic, 137–173. Amsterdam: John Benjamins.

Dehé, Nicole, and Yordanka Kavalova. 2007. Parentheticals: An Introduction. In *Parentheticals*, ed. Nicole Dehé and Yordanka Kavalova, 1–22. Amsterdam: John Benjamins.

Doron, Edit. 1992. Appositive predicates. *Belgian Journal of Linguistics* 7:23–33.

Doron, Edit. 1994. The discourse function of appositives. *IATL* 1:53–65.

Espinal, T. 1991. The representation of disjunct constituents. *Language* 64:726–762.

Giannakidou, Anastasia, and Melita Stavrou. 1999. Nominalization and ellipsis in the Greek DP. *The Linguistic Review* 16:295–331.

Hankamer, Jorge, and Ivan Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 391–426.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Higginbotham, James. 1985. On Semantics. *Linguistic Inquiry* 16:547–593.

Huddleston, Rodney, and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Kariaeva, Natalia. 2004. Determiner spreading in Modern Greek: split-DP hypothesis. Ms. Rutgers University. To appear in the Proceedings of the workshop on Greek Syntax and the Minimalist Seduction.

Keizer. 2005. The discourse function of close appositions. *Neophilologus* 89:447–467.

Kolliakou, Dimitra. 2004. Monadic definites and polydefinites: their form, meanning and use. *Journal of Linguistics* 40:263–333.

Lekakou, Marika. 2000. Focus in Modern Greek. Unpublished MA Dissertation, University College London.

Leu, Tom. 2007. From Greek to Germanic: poly(*in)definiteness and weak/strong adjectival inflection. Ms. New York University.
McCawley, James. 1998. *The Syntactic Phenomena of English*. Chicago: Chicago University Press.
Meyer, Christopher. 1992. *Apposition in contemporary English*. London: Longman.
Neeleman, Ad, and Hans van de Koot. 2002. The Configurational Matrix. *Linguistic Inquiry* 33:529–574.
Ntelitheos, Dimitris. 2004. Syntax of elliptical and discontinuous nominals. Unpublished MA dissertation, UCLA.
Panagiotidis, Phoevos. 2003. Empty Nouns. *Natural Language and Linguistic Theory* 21:381–432.
Panagiotidis, Phoevos. 2005. Determiner spreading as DP-predication. Ms. Cyprus College.
Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.
Reinhart, Tanya. 2006. *Interface Strategies*. Cambridge, Massachusetts: MIT Press.
Schwarzschild, Roger. 1999. GIVENness, AVOIDF and Other Constraints on the Placement of Accent. *Natural Language Semantics* 7:141–177.
Stavrou, Melita. 1995. Epexegesis vs. apposition. *Scientific Yearbook of the Classics Department*. Aristotle University of Thessaloniki.
Szendrői, Kriszta. 2001. Focus and the syntax-phonology interface. Doctoral Dissertation, University College London.
Szendrői, Kriszta. 2006. A flexible approach to discourse-related word order variations in the DP. Talk given at Workshop on DP-internal Information Structure, UiL OTS Utrecht.
de Vries, Mark. 2006. The syntax of appositive relativization. On specifiying coordination, false free relatives and promotion. *Linguistic Inquiry* 37:229–270.
Williams, Edwin. 1981. Argument structure and morphology. *The Linguistic Review* 1:81–114.
Williams, Edwin. 1989. The anaphoric nature of theta-roles. *Linguistic Inquiry* 20:425–256.
Williams, Edwin. 1997. Blocking and anaphora. *Linguistic Inquiry* 28:577–628.
Zwarts, Joost. 1992. X'-Syntax–X'-Semantics. Doctoral Dissertation, University of Utrecht.

Relatives and Pronouns in the English Cleft Construction[*]

MATTHEW REEVE

Abstract

In this paper I propose a new analysis of the English cleft in which the cleft clause is a restrictive relative clause adjoined to the clefted XP. I provide evidence showing that both 'matching' and 'raising' derivations are possible, that the initial *it* of clefts is not an expletive, and that the cleft clause undergoes obligatory extraposition to VP. Taken together, this evidence favours the present analysis over (a) analyses in which the cleft clause is extraposed from subject position, and (b) analyses in which the cleft clause is a complement of the copula or a focus head.

1 Introduction

The English cleft construction has been much discussed in the generative literature.[1] An example is given in (1a), which has the simple sentence paraphrase in (1b).[2]

(1) a. It was [$_{XP}$ THE SNAKE] [$_{YP}$ that the mongoose caught].
b. The mongoose caught the snake.

The cleft has the same propositional meaning as its simple sentence paraphrase, but typically makes the clefted XP prominent in some way. This prominence is often described as contrastive or identificational focus (see, e.g., Rochemont 1986 for the former term and É. Kiss 1998 for the latter), though a contrastive topic reading of the XP (in the sense of Büring 2003) is also possible. In addition, the cleft generates existential and exhaustiveness presuppositions; thus, (1a) presupposes

[*] I would like to thank my supervisor Ad Neeleman for much discussion and guidance. I am also grateful to David Adger, Theresa Biberauer, Paul Elbourne, Alison Hall, Vikki Janke, Hans van de Koot, Peter Svenonius, Rob Truswell and Reiko Vermeulen for discussion, judgements and help in general.

[1] The term 'cleft sentence' was coined by Jespersen (1927). As far as I know, the earliest generative work dealing specifically with clefts is Lees (1963).

[2] I will henceforth refer to XP in (1a) as the 'clefted XP', and to YP as the 'cleft clause'.

that there is something that the mongoose caught, and that the snake is all that the mongoose caught.

In this paper I propose a new analysis of clefts. In particular, I make the following claims: (i) the cleft clause is a restrictive relative (RRC), but is right-adjoined to the clefted XP rather than to an NP (in Abney's sense); (ii) the clefted XP is either base-generated in its surface position and associated in some way with a moved operator in the cleft clause (the 'matching' derivation), or raised out of the cleft clause to its surface position (the 'raising' derivation); (iii) the cleft clause is obligatorily extraposed and right-adjoined to VP; (iv) the cleft pronoun *it* is a standard pronoun rather than an expletive element. Since a central part of the analysis is that the cleft clause is a relative clause adjoined to the clefted XP (where X can be one of a number of lexical categories), I refer to it as the 'XP-relative analysis'. An illustration of the analysis is given in (2), which represents the structure of (1a):[3] [4]

(2) $[_{VP}\ [_{VP}$ it was $[_{DP}\ [_{DP}$ the snake$]_i$ $t_j]]$ $[_{CP}$ Op_i/t'_i that the mongoose caught $t_i]_j]$

In section 2, I present evidence that the cleft clause is an RRC adjoined to the clefted XP, proceeding from the observation that the cleft clause is internally essentially identical to an RRC, a fact which suggests that it is an adjunct. I then show that both matching and raising derivations must be available for clefts, which rules out analyses in which the cleft clause is extraposed from the subject. I also discuss a constraint on predicational interpretations of clefts which supports the idea that the cleft clause is attached to the clefted XP. In section 3, I show that cleft *it* cannot be an expletive. The evidence for this comes from the fact that cleft *it* is obligatory in Germanic, the fact that cleft *it* can control PRO, and the possibility of replacing cleft *it* with a demonstrative. In section 4, I claim that an obligatory process of extraposition takes place in clefts, basing this on evidence from topicalisation, VP-fronting, the insertion of parentheticals, and cleft extraposition in other Germanic languages. Section 5 contains a conclusion and some brief speculations on why adjunction of RRCs to non-NPs should be possible in clefts but impossible elsewhere.

---

[3] I assume that both subject and copular verb end up outside VP, presumably in SpecIP and I respectively. Alternatively, the copular verb might be generated in I and thus never appear inside VP.

[4] One important aspect of the XP-relative analysis is that the cleft clause adjoins to the clefted XP even under the raising derivation; in other words, the moved item rather than the target of movement projects. Although there has been debate about whether such operations are possible, I assume for the purposes of this paper that nothing rules them out in principle.

2 Where is the cleft clause generated?
2.1 Previous analyses

There are two dominant lines of analysis in the generative literature on clefts. The first, foreshadowed by Jespersen (1927), takes the cleft clause to originate in subject position, either as a free relative headed by a *wh*-operator or as part of a definite description (e.g., Akmajian 1970, Schachter 1973, Emonds 1976, Gundel 1977, Wirth 1978, Percus 1997).[5] These authors would thus take (1a) to have the same source as one of (3a-b):[6] [7]

(3) a. [$_{CP}$ What the mongoose caught] was the snake.
b. [The one [$_{CP}$ that the mongoose caught]] was the snake.

Under this view, (1a) would be derived from one of the underlying structures in (3) by extraposition of the CP, plus *it*-insertion or conversion of the definite description remnant to *it* (for 3a and 3b respectively). I will refer to analyses of this type as 'extraposition-from-subject' analyses.

The second type of analysis, which can be traced back to Jespersen (1937), takes the cleft clause to be base-generated as the sister of the copular verb or of a functional head encoding focus (e.g., Chomsky 1977, Halvorsen 1978, Delahunty 1982, Rochemont 1986, Heggie 1993, É. Kiss 1998, 1999). The clefted XP either occupies a left-peripheral position in the cleft clause (Chomsky, Heggie, É. Kiss) or is a sister of both the copular verb and the cleft clause (Halvorsen, Delahunty). I will refer to analyses of this type as 'expletive analyses', since they take the initial *it* to be an expletive rather than a semantically interpreted pronoun or the head of a definite description.

Finally, a small number of authors have claimed that the cleft clause is in some sense construed with the initial (non-expletive) *it*, but does not form an underlying syntactic constituent with it (e.g., Smits 1989, Hedberg 2000). As should be clear, this is the general approach I wish to defend here.

---

[5] Various problems with deriving clefts from *wh*-pseudoclefts were pointed out by Higgins (1973). As Percus (1997) suggests, deriving clefts from *th*-pseudoclefts instead may provide a way to avoid these. Nevertheless, the criticisms presented in this section apply to both types of analysis.

[6] I will refer to both types of sentence in (3) as 'pseudoclefts'. Where the two need to be distinguished, I will use '*wh*-pseudocleft' for sentences like (3a) and '*th*-pseudocleft' for sentences like (3b).

[7] In fact, Schachter (1973) proposes that the underlying structure of the cleft includes the initial *it*. Presumably extraposition would have to be obligatory under this analysis, given the ungrammaticality of the variant without extraposition (**It that the mongoose caught was the snake*).

2.2 Cleft clauses and relative clauses

As has been pointed out many times in the literature (e.g., Schachter 1973), cleft clauses bear a closer resemblance to relative clauses, and in particular to RRCs, than to other *wh*-constructions. Three types of evidence point to this conclusion. The first is that the set of *wh*-operators found in cleft clauses is a subset of that found in relative clauses:

(4)

Cleft	which, who(m), (?)where, (?)when, *how, *why
RRC	which, who(m), where, when, *how, *why
NRRC	which, who(m), where, when, *how, *why
Wh-*question*	what, which, who, where, when, how, why

Notably, this is also the case in other languages with English-like clefts (e.g., the other Germanic languages, French, Italian), and so cannot be considered to be an English-specific phenomenon.

Second, the alternation found in cleft clauses between a *wh*-operator and the complementiser *that* is also found in RRCs, but not in other '*wh*-constructions':

(5)
a. It was the snake which/that/0 the mongoose caught. (cleft)
b. the snake which/that/0 the mongoose caught (RRC)
c. the snake, which/*that/*0 the mongoose caught (NRRC)
d. What/*that/*0 did the mongoose catch? (*wh*-question)
e. The snake, *which**/***that/*0 the mongoose caught. (topicalisation)

Third, the '*that*-trace effect' found in complement CPs does not appear in clefts or RRCs:

(6)
a. It was the mongoose that *t* caught the snake.
b. the mongoose that *t* caught the snake
c. Which animal did you say (*that) *t* caught the snake?

It seems clear, then, that cleft clauses are a type of RRC. This being so, an analysis of clefts should ideally capture the various similarities between the two types of clause. Under expletive analyses, which treat cleft clauses as complements rather than adjuncts, these similarities are unexpected.[8] If, however, the cluster of

[8] Not everyone takes relatives to be adjuncts, of course – under Kayne's (1994) antisymmetry approach, relatives are CP complements of a determiner, and the head NP originates in the relative clause itself, raising to SpecCP. Kayne suggests that clefts should be analysed along similar lines, the difference being that the CP (the cleft clause) is selected by the copular verb rather than by a determiner. Again, though, this approach would not distinguish relatives and

properties distinguishing clefts and relatives from other *wh*-constructions is related to the fact that both cleft clauses and relatives are adjuncts to a non-verbal projection, in contrast to the other *wh*-constructions, then we have the beginnings of an account of the differences.[9]

If the cleft clause is right-adjoined to a non-verbal projection, then, we rule out expletive analyses of clefts, as well as analyses in which the cleft clause is base-generated as adjoined to a clausal projection (e.g., Smits 1989, who proposes that cleft clauses are base-generated adjuncts to the matrix VP).

2.3 The cleft clause is adjoined to the clefted XP

Having determined that the cleft clause is a type of RRC and thus an adjunct, we now have two alternatives: either the cleft clause is adjoined to the subject *it* and then extraposed (as in extraposition-from-subject analyses), or the cleft clause is adjoined to the clefted XP (as in the XP-relative analysis). There are various pieces of evidence suggesting that the latter possibility is correct. Most of the evidence for adjunction to the clefted XP is in a sense indirect, in that it supports the idea of a 'dual derivation' for clefts (i.e., that both matching and raising derivations are possible).[10] This in turn suggests that the cleft clause cannot have originated in the subject position, since under this view a raising derivation would be impossible, given the now widely-accepted view that movement to a non-c-commanding position is impossible.[11]

2.3.1 Subextraction and freezing. Pinkham and Hankamer (1975) observe that subextraction (i.e., extraction of a proper subconstituent) from the clefted constituent is possible only in certain types of cleft:[12]

---

clefts from other *wh*-constructions with respect to the occurrence of *wh*-operators and the *that*-trace effect. In any case, Borsley (1997) provides many reasons to think that Kayne's analysis of relatives is incorrect (though see the reply by Bianchi 2000).

[9] I have said 'non-verbal projection' because while all of DP, PP, AP and AdvP are apparently cleftable, VP and IP do not seem to be. The status of CP is unclear – apparent CP-clefts exist, but the CPs in question could be headed by a null DP.

[10] Indeed, such a duality of derivation would be expected if cleft clauses are a type of restrictive relative, as it has been shown by various authors, including Carlson (1977), Sauerland (1998) and Aoun and Li (2003), that both matching and raising derivations are required for standard restrictive relatives.

[11] When this view was not widely accepted, some authors, such as Schachter (1973) and Emonds (1976), analysed clefts as involving both raising of the clefted XP and extraposition of the cleft clause from subject position. I assume that such analyses are now not in contention.

[12] The slight deviance of (7a) is attributed by Pinkham and Hankamer to the general awkwardness of extracting from a clause-non-final constituent (Kuno 1973).

(7) a. ?Who was it a picture of that he decorated his door with?
b. *Who was it with a picture of that he decorated his door?

Making use of the generalisation later known as the Freezing Principle (Wexler and Culicover 1980), which rules out subextraction from a moved constituent, they conclude that there must be two types of derivation for clefts: one in which the cleft is essentially derived from an underlying pseudocleft (as in the analysis of Akmajian 1970), and a second in which the cleft is derived from its simple sentence paraphrase, the clefted XP being extracted to a clause-external position, and the *it* and copular structure being generated subsequently. The two types of derivation correspond to extraposition-from-subject and expletive analyses respectively. Crucially, in the first type of derivation (the 'deep cleft', which Pinkham and Hankamer claim is restricted to DP-clefts), the clefted XP does not move, and hence can be subextracted from, while in the second (the 'shallow cleft', which is possible in all cases), the clefted XP moves, preventing further subextraction from it.

The argument from subextraction has been criticised by Gundel (1977), who questions the validity of the data presented by Pinkham and Hankamer. I agree with her that the contrast between the examples in (7) does not seem to be sufficient to motivate different derivations for DP- and PP-clefts, but it seems that, at least for some speakers, the possibilities for subextraction are wider, not narrower, than Pinkham and Hankamer claim. For example, subextraction from PP-clefts does seem to be possible when the entire complement of the P is extracted, as in the examples in (8):[13]

(8) a. Which national park was it in that the mongoose caught the snake?
b. Which year was it in that Tony Blair became Prime Minister?
c. ?Whose picture was it with that he decorated his door?

The 'deep cleft' (matching) structure must therefore be available for at least some PP-clefts as well as DP-clefts. But crucially, when we turn to AP-clefts, we find strong motivation for the claim that some clefts allow only the shallow/raising structure. It has often been noted in the literature that AP-clefts are often marginal in acceptability, though they often improve if an appropriately contrastive context is set up. However, Heggie (1993) notes that AP-clefts in which the AP is a secondary predicate are often much better than, for example, AP-clefts in which the AP is the predicate of a copular clause.

[13] In general, it seems to be easier to subextract from locative clefted PPs than from other clefted PPs.

(9) a. It's drunk that John sounds intelligent.
b. It was undressed that the clown looked pitiful.
c. It's raw that Bill usually likes his meat.

AP-clefts with a secondary predicate AP thus provide a good testing-ground for the subextraction argument, at least for those adjectives, such as *drunk*, that take complements. It appears that subextraction from these clefted APs is unacceptable, as shown in (10):

(10) a. ?*What is it drunk on that John sounds intelligent?
b. ?*What was it dressed in that the clown looked pitiful?

This restriction cannot be a subcase of a general restriction on extraction from complements of APs, since such extractions are usually fully acceptable. Similar judgements apply to other types of clefts (e.g., CP- and IP-clefts). We thus seem to have at least some cases of clefts where the Freezing Principle is operative, and thus where only a raising derivation is available. In fact, the fundamental divide between derivationally ambiguous (matching/raising) and raising-only clefts seems to correspond roughly to the distinction between extended nominal (of category D, P) and extended non-nominal projections (of category A, Adv, C, I, etc.). A further indication that this is the case comes from the possibility of overt *wh*-operators, which only seems to be realised in DP- and (some) PP-clefts:

(11) a. It was a picture of John which/that he decorated his door with.
b. It was in the Kruger National Park ?where/that the mongoose caught the snake.
c. It's drunk *which/that John sounds intelligent.
d. It's drunk *which/that John seems to be.
e. It was that you left so early ??which/that bothered John.
f. It was far too slowly *which/that John walked home.

Furthermore, those PP-clefts which allow subextraction seem in the main to correspond to those which allow *wh*-operators (though there are exceptions, such as those with *with*):[14]

[14] Again, those PP-clefts that tolerate *wh*-pronouns most easily seem to involve locative PPs, which of course take *where*. As (b), shows, temporal PP-clefts with *when* are also possible. This suggests that only where a *wh*-operator is possible is a null operator also possible, since otherwise we would expect all PP-clefts and AP-clefts to allow a matching structure with a null operator, and thus to allow subextraction. The more limited distribution of the matching structure as compared with the raising structure is thus plausibly related to the limited inventory of *wh*-operators, though it is not clear what property of *wh*-operators prevents them from being linked to,

(12) a. It was in the Kruger National Park ?where/that the mongoose caught the snake.
b. It was in 1997 ?when/that Tony Blair became Prime Minister.
c. It was with a picture of John *which/that he decorated his door.
d. It was to China ?*where/that John travelled.
e. It was into the kitchen ?*where/that Bill walked.

In common with Pinkham and Hankamer's analysis, the XP-relative analysis allows for a derivational ambiguity of the type required, but with the advantage that it does not require radically different underlying structures for the two types of clefts. For example, it has been pointed out that Pinkham and Hankamer's raising derivation requires the derivation of *it* plus the copular structure by a transformation, which should rule out the insertion of extra elements in the matrix clause, such as modals and adverbials. This problem is avoided under the present analysis, since the *it* plus copular structure is not created by a clefting transformation, but is generated independently, in the same way as in copular sentences in general.

2.3.2 Scope ambiguities. É. Kiss (1999) points out that clefts, unlike RRCs and complement clauses, permit a universal quantifier in the embedded clause to take matrix scope. There is thus a contrast between (13a) and (13b-c):

(13) a. It was some paper by Chomsky that everybody wanted to read.
b. I have some paper by Chomsky that everybody wanted to read.
c. He said about some paper by Chomsky that everybody wanted to read it.

She takes this to indicate that a raising derivation, and hence syntactic reconstruction of the indefinite DP to a position c-commanded by the quantifier, is possible in clefts, but not in the other types of construction.[15] Of course, the possibility of raising is afforded by the XP-relative analysis as well as by her (expletive) analysis. Nevertheless, a potential problem for the view that such scope ambiguities are always due to syntactic reconstruction is posed by the fact that pseudoclefts of both types (*wh-* and *th-*) also display such ambiguities:

---

say, APs, especially given that in non-restrictive relatives APs (as well as VPs, IPs and CPs) can be linked to *which*. Furthermore, it is not clear why PP-clefts with *with* should allow subextraction but not a *wh*-operator.

[15] Or, at least, that the indefinite determiner *some* could not be part of the material reconstructed; in the restrictive relative in (14b), it is entirely possible that the NP *paper by Chomsky* could have raised (as in the analyses of Kayne 1994, Aoun and Li 2003 and Hulsey and Sauerland 2006, for example).

(14) a. What everybody wanted to read was some paper by Chomsky.
b. The one that everybody wanted to read was some paper by Chomsky.

Although some analyses of pseudoclefts have posited derivations in which the pseudoclefted constituent is extracted from the *wh*-clause or reconstructed into it at LF (e.g., Akmajian 1970, Bošković 1997), it has generally been accepted since Higgins (1973) that pseudoclefts do not have a raising derivation. How, then, can we account for the fact that both clefts and pseudoclefts display scope ambiguities?

A promising line of attack is suggested by data from Williams (1994), who notes that there is a gap in the scope connectivity paradigm for pseudoclefts:

(15) What bothered someone in the class was every paper by Chomsky.

Unlike (14a), (15) does not allow a reading under which the universal quantifier takes scope over the indefinite. Crucial for our purposes is the fact, not discussed by Williams, that the equivalent cleft does seem to allow this reading:[16]

(16) It was every paper by Chomsky that bothered someone in the class.

It is clear that under the XP-relative analysis, as well as under expletive analyses, this reading of (16) is predicted to exist, since the universal QP c-commands the indefinite at LF under both matching and raising derivations. This is important evidence against extraposition-from-subject analyses, since if (17) were derived from something like (15), the appropriate c-command relation would never arise.

There remains the problem of why inverse scope is possible in (14a). As Williams and Heycock and Kroch (1999) have pointed out, modifying the indefinite in (14a) with an overt distributive operator such as *different* blocks the reading whereby the universal takes scope over the indefinite:

(17) What everybody wanted to read was a different book by Chomsky.

(17) only allows the reading where there is a single book which is different from some other, contextually specified book, not the reading where each person is associated with a distinct book. Given that distributivity is generally possible with universal>indefinite scope readings, this is an important indicator that (14a) does not involve true scope reconstruction. Williams proposes that so-called 'inverse scope' arises in (16a) because the trace in the *wh*-clause generates a 'functional'

[16] It should be pointed out that universal quantifiers are somewhat deviant as clefted constituents (e.g., Schachter 1973), a fact which should be allowed for when considering the examples.

reading, the indefinite being unselectively bound. The fact that distributivity is incompatible with a functional reading thus explains why (17) does not allow a distributive reading.

Now compare the cleft equivalent of (17):

(18) It was a different book by Chomsky that everybody wanted to read.

It seems clear that (18) allows the distributive reading much more easily than (17), particularly if *different* receives main stress, thus being the element contrasted. This being the case, it seems that true inverse scope subordination is possible in clefts. To the extent that such scope subordination is a result of c-command at LF, therefore, this contrast between clefts and pseudoclefts supports the availability of a raising derivation for clefts, and argues against extraposition-from-subject analyses.[17]

2.3.3 Idioms. Brame (1968) and Schachter (1973) claim that the behaviour of idioms like *keep track of* provides evidence for a raising analysis of relatives:

(19) a. She's keeping careful track of her expenses.
b. *The careful track pleases me.
c. The careful track that she's keeping of her expenses pleases me.

The distinction in acceptability between (19a) and (19b) is evidence that *(careful) track* must be generated as the complement of *keep* for the idiomatic reading to be available. This in turn suggests that in (19c), *careful track* has raised from the complement position of *keep* inside the relative clause.

A similar argument can be constructed for clefts. As in relative clauses, the idiomatic sense of *keep track of* is possible if the object is clefted. In the equivalent pseudocleft, however, the idiomatic reading is not available. The inverse pseudocleft in (20c) shows that this cannot simply be a linearity effect:

(20) a. It was careful track that she kept of her expenses.
b. *What she kept of her expenses was careful track.
c. *Careful track was what she kept of her expenses.

[17] If the matching derivation accounts for the possibility of subextraction and the raising derivation accounts for inverse scope, it is expected that subextraction will block inverse scope readings. Thus, inverse scope should be impossible in (i):

(i) What was it a picture of that everybody bought?

My judgements are unclear on this, but I feel that the inverse scope reading is very awkward.

Idioms, then, provide further evidence that a raising derivation is available in clefts but not in pseudoclefts, ruling out extraposition-from-subject analyses of the former.

2.3.4 Wh-blocking. It has been observed by Aoun and Li (2003) that certain connectivity effects displayed by RRCs are blocked when a *wh*-operator occurs in the relative. For example, (21a) appears to allow inverse scope between the *every*-phrase and the *two*-phrase much more easily than does (21b), which features an overt *wh*-operator:

(21) a. I phoned the two patients (that) every doctor will examine.
b. I phoned the two patients who every doctor will examine.

Authier and Reed (2005) claim that this '*wh*-blocking' effect does not extend to clefts. In my judgement, though, *wh*-blocking effects are just as strong in clefts as in relatives:

(22) a. It was some paper by Chomsky which everybody wanted to read.
b. ?*It was careful track which she kept of her expenses.

The *wh*-blocking effect receives a natural account under the XP-relative analysis: when a *wh*-operator appears, only the matching derivation is possible, since under the raising derivation the *wh*-operator and clefted constituent would be competing for the same position (SpecCP of the cleft clause). Of course, under the matching derivation, syntactic reconstruction of the clefted constituent into the cleft clause will not take place, and hence connectivity effects which depend on reconstruction will not be visible. Under extraposition-from-subject analyses, on the other hand, it is not clear how *wh*-blocking could be accounted for, since in *wh*-pseudoclefts there must always be a *wh*-operator, and yet no *wh*-blocking effects are seen. And in *th*-pseudoclefts, it is not clear that the presence of a *wh*-operator makes any difference to the possibility of connectivity effects. In any event, if, as is argued above, there is no scope reconstruction in pseudoclefts, *wh*-blocking effects would not be expected anyway.

2.3.5 Specificational versus predicational readings. Since Higgins (1973), copular sentences have commonly been divided into various classes according to the semantic properties of the elements linked by the copula. Higgins himself distinguished four subtypes: specificational, predicational, identificational and identity sentences. For our purposes, it will be sufficient for now just to consider the first two subtypes, specificational and predicational sentences. Both clefts and

pseudoclefts proper are specificational sentences: that is, they provide a value (represented by the (pseudo)clefted constituent) for a variable in the cleft/*wh*-clause. In addition, a copular sentence with a *wh*-clause subject allows a predicational interpretation of the postcopular constituent (such sentences are commonly termed 'predicational pseudoclefts').

(23) a. What I am pointing at is the cat. (specificational)
b. What I am pointing at is feline. (predicational)

Though 'predicational clefts' are also possible under certain circumstances (e.g., Declerck 1983), the clefted constituent seems to be restricted to DP: an AP can never appear in such a position:

(24) a. It is a clever person that can solve puzzles like that.
b. *It is clever that can solve puzzles like that.

This is a very strong piece of evidence in favour of the XP-relative analysis, and a problem for extraposition-from-subject analyses. Take, for example, the analysis of Percus (1997), who posits a post-syntactic rule converting the remnant of the definite description subject to *it* after extraposition of the cleft clause.

(25) Spellout rule: [+def] 0 t_{CP} => *it*

Percus attempts to provide independent motivation for this conversion rule by claiming that it also applies in cases of clausal extraposition more generally – that is, cases where the underlying definite description subject would be argumental rather than an inverted predicate, which Percus assumes the definite description subject of clefts to be. Thus, sentence (26a) would result from extraposition of the clausal subject and application of the spellout rule in (25):

(26) a. It is widely believed on Wall Street that the shares are overvalued.
b. That the shares are overvalued is widely believed on Wall Street.

It is clear, though, that generalising the rule in this way means that 'predicational AP-clefts' such as (24) should also be generated, since the spell-out rule now applies both to argumental and predicational (superscriptional) definite description remnants, and should hence allow the creation of an argumental *it* subject, regardless of what follows the copula. If, on the other hand, the rule is to apply only to predicational definite descriptions – in effect restricting its application to specificational copular sentences – then it seems merely to be a stipulatory device.

Under the XP-relative analysis, however, sentences like (24) are excluded. The cleft clause is a relative clause on the clefted constituent, and so must contain a corresponding gap or operator. In (24), however, there is no gap or operator in the cleft clause corresponding to the. In the predicational DP-cleft examples, however, there is an appropriate gap or operator, and hence predicational DP-clefts are acceptable in principle.[18]

2.3.6 Extraposition of the cleft clause. One claim made in this paper is that the cleft clause is obligatorily extraposed, a view naturally shared by proponents of extraposition-from-subject analyses such as Percus (1997). Nevertheless, the two types of analysis make different predictions about the possible landing site(s) of cleft extraposition, since Percus's analysis claims that extraposition is from subject position (SpecIP), while the XP-relative analysis claims that it is from inside VP. Since Baltin (1981), it has generally been accepted that extraposition from subjects targets IP, while extraposition from objects targets VP.[19] Evidence for this comes from examples like (27) involving VP-ellipsis:

(27) a. Although not many people would ride with Fred who knew just him, some would ___ who knew his brother.
b. *Although he didn't call people up who are from Boston, he did ___ who are from New York.

In (27a), where the relative clause is extraposed from the subject *not many people*, VP-ellipsis does not obligatorily delete the relative clause, which suggests that it is not inside VP. In (27b), on the other hand, where the relative clause is extraposed from the object *people*, the relative clause must be included with the elided material, and therefore must be inside VP.

---

[18] It is marginally possible to have clefts where the gap or operator in the cleft clause plays a predicational role, given the correct context (see Declerck 1983 for discussion). In (i), the AP does have a corresponding gap in the cleft clause, and hence the sentence is expected to be acceptable. (ii) shows that this structure is also possible with predicate nominals:

(i) It's pretty that Mary is, more than anything else.
(ii) It is a teacher that he is, not a butcher!

It seems clear, however, that the clefted constituents in (i-ii) could not then be playing a predicational role with respect to the matrix clause. Again, this is as expected.

[19] In fact, Culicover and Rochemont (1990) show that extraposition from subject can target VP as well as IP ('target' being used here in a loose sense, since their analysis involves base-generation of the extraposed clause). This is irrelevant to the present argument, however, since it is the possibility of adjoining a clause to IP that is being tested.

Under the XP-relative analysis, extraposition of the cleft clause is assumed to be from inside VP, as for objects, and therefore the analysis predicts that the only possible landing site is VP, a prediction borne out by (28):[20]

(28) *Although it wasn't John who cooked the stew, it was ___ who baked the cake.

Since VP-ellipsis must delete the cleft clause, it can be concluded that the clause can be extraposed no higher than VP. This is further evidence against extraposition-from-subject analyses, which also predict IP to be a possible landing site.

2.3.7 Summary. In this section I have presented evidence that the cleft clause is a type of restrictive relative adjoined to the clefted XP, rather than a complement as in expletive analyses. Second, I have shown that both matching and raising derivations are possible, which causes trouble for extraposition-from-subject analyses.

3 The status of the cleft pronoun

3.1 Evidence that cleft *it* is not an expletive

This section presents evidence that cleft *it* is not an expletive, but should instead be analysed as a full pronoun. Naturally, this presents a problem for expletive analyses, but can be handled by the XP-relative analysis, assuming that *it* and the cleft clause can be associated in some way. I present some speculations on how this might be achieved in section 5.

3.1.1 'Expletives' in English. There are two subject elements which have in the past been considered to have expletive uses in English: *it* and *there*, the former found in weather sentences, extraposition contexts, and clefts, and the latter typically found in existential sentences. These various contexts are illustrated in (29):

(29) a. It rained. (weather *it*)
b. It is good that you have come. (extraposition *it*)
c. It was this car that she wanted to buy. (cleft *it*)
d. There are many solutions to that problem. (existential *there*)

[20] I am of course assuming here that both subject and copular verb are not inside VP by the time VP-ellipsis applies.

While *there* in (29d) is generally accepted to be a true expletive, weather and extraposition *it* have often been claimed to be 'quasi-arguments' (a term originating from Chomsky 1981). The aim of this section is to show that cleft *it* patterns with these 'quasi-arguments' (and, in fact, with real arguments as well) rather than with expletives.

3.1.2 Control. It has been pointed out that weather *it* and certain instances of extraposition *it* can control PRO – this is illustrated by (30a-b) (Chomsky 1981, Bennis 1986).[21] The acceptability of (30c) shows that cleft *it* also has this property:

(30) a. It sometimes rains after [PRO snowing]. (Chomsky 1981:325)
b. It was clear [PRO after having been explained to us] that we were wrong.
c. It was John who Bill spent all his time with [despite PRO being Mary the day before].

Chomsky (1981:323) notes that obligatory PRO generally takes an argument as its binder, pointing out that *it* in *it seems that…* (plausibly another type of extraposition *it*) and existential *there*, as in *there are three men…*, cannot bind PRO. Furthermore, PRO normally takes on the referential properties of its binder.[22] In this respect, (30a) seems odd, since weather *it* is apparently non-referential. Chomsky thus refers to weather *it* as a 'quasi-argument', meaning that it receives a theta-role but is non-referential. Thus the fact that cleft *it* can control PRO does not entail that this *it* is referential, but suggests that it at least has the status of a quasi-argument rather than an expletive (though it is not clear whether it should receive a theta-role, a question which I touch on in section 5). The control diagnostic thus distinguishes the three types of *it* in (30) from *seem*-type *it* and existential *there*.

3.1.3 The obligatoriness of cleft pronouns. The three types of *it* discussed above – weather, extraposition and cleft *it* – can be distinguished in some Germanic languages with respect to their obligatoriness in certain contexts.[23] Previous research on *IT* in Germanic has tended to focus on the first two types, to the exclusion of cleft *IT*. It has been noted that some of these languages allow or require certain instances of *IT* to be dropped in inversion contexts. Notably, though,

[21] This also appears to be true in Dutch (Bennis 1986), German (Safir 1985) and Danish (Vikner 1995:228).

[22] Chomsky points out that it is important to distinguish this sense of 'referential', meaning 'denoting a member of the domain D of individuals that serve as values of variables and as denotata', and the sense pertaining to relations between language and the real world. The term is used here in the former sense only.

[23] For ease of exposition, when I wish to refer to these three types of *it* and their other Germanic equivalents together, I will use *IT* as a cover term.

cleft *IT* can never be dropped in any of these languages, even under inversion, suggesting that it does not occur purely for EPP-related reasons, but has an interpretation essential to the interpretation of the cleft as a whole.

The Germanic languages apart from English are verb-second (V2) languages, and thus the presence of an initial adverbial in a matrix clause triggers 'subject-verb inversion'. In this situation, *IT* subjects vary in their behaviour between languages and between uses of *IT*: in some cases the pronoun must be dropped, in other cases it may optionally be dropped, and in still other cases it must be retained, as shown in the examples in (31-33), taken from Vikner (1995). In Danish, *det* must be retained under inversion in weather and extraposition sentences. In German, *es* must be retained under inversion in weather sentences, but can be dropped in some extraposition sentences. Finally, in Icelandic, weather *það* must be dropped under inversion, while extraposition *það* is optionally dropped:

(31) a. I går regnede *(det).
yesterday rained it
b. Naturligvis er *(det) godt at du er kommet.
of-course is it good that you are come

(32) a. Gestern regnete *(es).
yesterday rained it
b. Natürlich ist (es) gut, dass du gekommen bist.
of-course is it good that you come are

(33) a. Í gær rigndi (*það).
yesterday rained it
b. Að sjálfsögðu er (það) gott að þú ert kominn.
of-course is it good that you are come

A fact which has to my knowledge not been pointed out before is that in all three languages (in fact in all modern Germanic languages), cleft *IT* must be retained under inversion.[24]

(34) a. I går var *(det) DEN JAKKE som Björn købte.
yesterday was it the jacket that Björn bought
b. Gestern war *(es) DIESER WAGEN, den sie kaufen wollte.
yesterday was it this car which she to-buy wanted

[24] Peter Svenonius (p.c.) confirms this for Icelandic. What this suggests is that cleft *það* might be an instance of demonstrative *það* rather than an *it*-like pronoun, since demonstrative *það* is also obligatory under inversion. Given that demonstratives are optionally possible as cleft pronouns in (at least) English, Dutch and German, this would not be surprising.

c. Í gær var *(það) JÓN sem ég hitti í bænum.
yesterday was it Jón that I saw in the-town

What can we conclude from this? If cleft *IT* is the only type of *IT* which can never be dropped under inversion in Germanic, its occurrence is clearly independent of EPP-related considerations.[25] Therefore it is likely to be playing some role other than a purely syntactic one.[26] There are various explanations one could offer for the fact that weather and extraposition *IT* can sometimes be dropped, but unfortunately I cannot discuss the issue further here. The important thing is that the presence of cleft *IT* seems to be essential to the interpretation of clefts, rather than just a placeholder as expletive analyses would suggest.

3.1.4 Demonstratives in clefts. Yet another fact about clefts which suggests that the initial *it* is not an expletive is that it can alternate with demonstratives, as shown in (35a) (Smits 1989, Hedberg 2000). This option is not limited to English; Smits (1989) reports that it is also available in Dutch and German, as shown in (35b-c).[27]

(35) a. It/that was THE SNAKE that the mongoose caught.
b. Het/dat was DE REGISSEUR die zo hardgrondig vloekte.
c. Das bin ICH, den du dort am Rande siehst.

Some instances of weather *IT* and extraposition *IT* also have this property. For instance, German and Dutch weather *es*/*het* can sometimes be replaced by a demonstrative, as shown in (36a-b) (Vikner 1995, quoting Gisbert Fanselow, p.c.),

---

[25] It has often been assumed that the Germanic languages which drop *IT* under inversion have a non-referential *pro* in these cases (e.g., Vikner 1995). Some recent work (e.g., Biberauer 2003) has challenged the desirability of postulating such a *pro* in Germanic, however. In any case, the non-occurrence of a non-referential *pro* in clefts would still be consistent with the idea that cleft *IT* does more than just satisfy the EPP.

[26] A distinction between cleft pronouns and weather/extraposition pronouns is also evident in Russian (as pointed out by Gundel 1977). The '*èto*-cleft', which is interpretatively similar to the English cleft (though it does not tolerate relative operators or complementisers), obligatorily features the pronoun *èto* 'this, it', yet *èto* does not appear in weather and *seem*-type extraposition sentences, and is optional in other types of extraposition sentences (Mezhevich 2003). Yet another language in which cleft pronouns can be distinguished from weather and extraposition pronouns is French – while the latter two contexts only permit *il*, clefts only permit *ce*.

[27] This option seems to be available only if no element of the cleft clause is in focus, as shown by the unacceptability of (i-ii) with *that*, (i) being a contrastive topic cleft, in which *the mongoose* is contrastively focused, and (ii) being an example of Prince's (1978) 'informative-presupposition' clefts, in which the cleft clause contains new information focus:

(i) It/*that was the snake that THE MONGOOSE caught.
(ii) It/*that was just about 50 years ago that Henry Ford gave us the weekend.

and English extraposition *it* can sometimes be replaced by *that*, as in (36c). (Again, in the latter case no part of the extraposed clause can be in focus.)

(36) a. Es/das regnet ja doch wieder.
b. Dit regent toch weer.
c. It/that was good that you baked a cake for me.

As expected, 'seem'-type *it* and existential *there* cannot be replaced by demonstratives:

(37) a. It/*that seems that John is standing in the corner.
b. There/*that are many solutions to that problem.

Alternation with a demonstrative is, of course, a characteristic of argumental pronouns, and, as (37) indicate, is not a characteristic of expletive elements. Once again, therefore, cleft *it* appears to pattern with argumental pronouns rather than with expletives.[28]

3.1.4 Summary. In the preceding subsections, it has been shown that cleft *IT* patterns with argumental pronouns, as well as with 'quasi-arguments' such as weather *IT* and (some instances of) extraposition *IT* in at least three ways: it can control PRO, it cannot be dropped in inversion contexts, and it can alternate with demonstratives. In all these respects, it differs from existential *there* and 'seem' *it*.

4 Extraposition of the cleft clause

4.1 Evidence for obligatory extraposition

[28] A demonstrative introducing a cleft may be used deictically:

(i) That's JOHN that's standing in the corner.

Since deictic elements cannot generally have linguistically represented antecedents, this could be considered to be further evidence against the view that the cleft pronoun must be construed syntactically with the cleft clause in some way, either as part of an underlying definite description (e.g., Akmajian 1970, Percus 1997), or as a pronoun taking the cleft clause as an extraposed complement (Hedberg 2000). In fact, the possibility of deixis suggests that the cleft pronoun should be interpreted in the same way as the pronoun in right-dislocation sentences (as claimed by Gundel 1977), since RD-pronouns can also be deictic demonstratives:

(ii) That's JOHN, the one that's standing in the corner.

In this section, I attempt to show that clefting involves obligatory extraposition of the cleft clause to VP.

4.1.1 Topicalisation and VP-fronting. Data from McCawley (1981) suggest that the clefted constituent and cleft clause do not make up a surface constituent. First, he notes that both restrictive and appositive relatives can be topicalised along with their antecedents, but the antecedent cannot be topicalised alone. With clefts, however, the situation is reversed: the clefted constituent and cleft clause cannot be topicalised together, but the clefted constituent can be topicalised alone.

(38) a. The fish that I caught, Bill ate.
a´.*The fish, Bill ate that I caught.
b. Stella, who I love, many people can't stand.
b´.*Stella, many people can't stand, who I love.
c. *Bill who I talked to, it may have been.
c´.?Bill, it may have been who I talked to.

As McCawley observes, this suggests that a restrictive or appositive relative forms a constituent with its antecedent, but the cleft clause does not form a constituent with the clefted constituent, or at least not a constituent of the right type for topicalisation. Obligatory extraposition in clefts accounts for this contrast, since after this operation the clefted XP and cleft clause do not form a surface constituent; thus, topicalising the two together would either be an instance of double topicalisation (of DP and CP separately) or of VP-topicalisation (VP being the smallest constituent containing both clefted XP and cleft clause), both of which are generally impossible in English. There is, however, a construction in English which does allow dislocation of VP, namely VP-fronting. As (39) show, it seems that in a VP-fronting-type environment the clefted constituent and cleft clause can be moved as a unit:

(39) a. John said that he would arrive early, and arrive early he did.
b. ?John said that it was Bill who I talked to, and Bill who I talked to it was.

If the moved material in (38c) and (39b) made up a surface DP, (38c) would be expected to be acceptable, and (39b) would be expected to be unacceptable, contrary to what is actually found. We thus have evidence that VP in clefts is the smallest constituent containing both the clefted constituent and cleft clause. This in turn implies that the cleft clause has undergone obligatory extraposition to VP.

4.1.2 Parentheticals. The second contrast McCawley discusses relates to the insertion of parentheticals between the relative and the antecedent. As the data in (40) show, this is less acceptable for restrictive and appositive relatives than for clefts (and pseudorelatives):

(40) a. *Tom cooked a dish, as you know, that I always enjoy.
b. ?Tom cooked twice-cooked pork, as you know, which I always enjoy.
c. It was Sam, as you know, that Lucy was talking to.

Given that parentheticals are presumably adjoined to the extended verbal projection (C, I, V), the fact that (40c) is acceptable suggests that there is some such adjunction site between the clefted constituent and cleft clause. If extraposition of the cleft clause is to VP, then this projection would also be VP.[29]

4.1.3 Extraposition in SOV languages. It is obviously not possible to discern from the surface form of English clefts whether extraposition of the cleft clause is obligatory, since it would in many cases be string-vacuous. In SOV languages such as German and Dutch, on the other hand, there are certain types of clauses in which it is obvious whether extraposition has taken place – namely, those in which there is an auxiliary and hence a clause-final infinitive, and embedded finite clauses, in which the finite verb must appear clause-finally. Thus, if extraposition has taken place in such clauses, the extraposed phrase will always follow the verb.

In German and Dutch, as in English, extraposition of relative clauses is normally optional where it is possible; for example, the German examples in (41), from Kiss (2005), are equally acceptable:

(41) a. …weil jeder Mann, der schläft, schnarcht.
because every man who sleeps snores
b. …weil jeder Mann schnarcht, der schläft.
because every man snores who sleeps

In clefts, however, extraposition of the cleft clause is obligatory, as shown by the following German examples (Smits 1989:282/320):

[29] It must be admitted that VP seems a less likely adjunction site for parentheticals than IP. One possibility is that parentheticals are not directly integrated into the syntactic representation at all, but that the relative acceptability of insertion sites is related to the syntax-phonology mapping. Under this view, one might propose that the presence of a right VP boundary, resulting from extraposition, between the clefted constituent and cleft clause, is what allows insertion of a parenthetical, because it is mapped to the right boundary of an intonational phrase. I leave this question for future research.

(42) a. Jutta sagt, dass es DIESER WAGEN war, den sie kaufen wollte.
Jutta says that it this car was which she buy wanted
b. *Jutta sagt, dass es DIESER WAGEN, den sie kaufen wollte, war.
Jutta says that it this car which she buy wanted was

Of course, the fact that extraposition is obligatory in German and Dutch clefts does not mean that the same holds of English clefts, but given the strong syntactic and interpretative similarities between clefts in these three languages, the facts in (41-2) strongly suggest this possibility.

5 Conclusion and a loose end

5.1 Conclusion

In this paper, I have attempted to defend the XP-relative analysis over alternatives such as the extraposition-from-subject analysis of Percus (1997) and the expletive analysis of É. Kiss (1999). It was shown that the cleft clause must be adjoined to the maximal projection of the clefted XP, given the evidence that it is a relative modifier rather than a complement clause (contra expletive analyses), and that a raising as well as a matching derivation must be available (contra extraposition-from-subject analyses). Furthermore, it was shown that cleft *it* is unlikely to be an expletive. Finally, an obligatory process of extraposition was shown to apply to the cleft clause, adjoining it to VP.

The conclusions reached so far clearly raise some difficult questions, not the least of which is how the cleft construction as a whole is semantically interpreted. Related to this is the question of how the possible information-structural interpretations of clefts are derived. Then there are further questions relating to the syntax of the construction: why should 'relative clause' extraposition be obligatory in clefts but not elsewhere, and why can an apparent RRC be adjoined to a non-NP (DP, PP, etc.) in clefts, but not more generally? For reasons of space I will only discuss the last of these questions here: adjunction to non-NPs.

5.2 A loose end: adjunction of RRCs to non-NPs

It was proposed in section 2 that cleft clauses are a type of RRC. Given that RRCs are normally assumed to be adjoined to NP (where NP is the complement of D), an obvious question raised by the XP-relative analysis is why clefts allow adjunction of an RRC to DP, PP, AP and CP.

Although I cannot develop a full answer to this question here, I believe that part of the answer lies in the unusual thematic properties of copular constructions, and of equatives in particular. Heycock and Kroch (1999) argue convincingly that pseudoclefts should be analysed as equatives – sentences asserting identity between the two elements related by the copula – rather than as inverted predications, as proposed for example by Moro (1997). If so, then it seems reasonable to assume that clefts are also equatives, given the strong interpretative parallels between pseudoclefts and clefts. It has been observed in the literature that there is no obvious sense in which the elements linked by the copula in an equative can be said to receive or assign theta-roles (see, e.g., Pereltsvaig 2001:182ff). Therefore, adjunction of cleft clauses to DP, PP, AP and CP might be permitted because these are never arguments when clefted, and so such adjunction does not fall foul of the ban on adjunction to arguments proposed by Chomsky (1986:6) and McCloskey (1992) (who calls it the Adjunction Prohibition).[30] In general, DP, PP, AP and CP are either arguments or predicates, and adjunction of RRCs to them will be ruled out by the Adjunction Prohibition in the former case.

One way of implementing the effects of the Adjunction Prohibition on clefts might be to adopt the proposal of Neeleman and van de Koot (2002) that theta-role assignment involves satisfaction of a 'thematic function' (which percolates upwards from V to VP) by direct domination of a node of the appropriate category. Consider the following structures, which represent adjunction of an RRC to NP and DP respectively, where the DP in both cases is complement of V:

(43) a. $[_{VP}$ V $[_{DP}$ D $[_{NP}$ NP CP]]]
b. $[_{VP}$ V $[_{DP}$ $[_{DP}$ D NP] CP]]

Suppose that RRCs (and cleft clauses) bear a thematic function which must also be satisfied (Higginbotham 1985 proposes that this is via 'theta-binding' by a determiner), but that it must be satisfied by a zero-level category (a head). Thus, when an RRC is adjoined to NP, as in (43a), its thematic function will percolate first to NP (where it will collapse with the thematic function of NP to form a single function, according to the system of Neeleman and van de Koot), and then to DP. Having reached this position, the thematic function directly dominates the D node, which can thus satisfy the function. On the other hand, when an RRC adjoins to a DP, as in (43b), the thematic function of the RRC percolates up to the higher of the two DP nodes, which does not directly dominate the D node. The thematic function is thus not yet satisfied by this point and must percolate further, this time to VP.

[30] Furthermore, this might account for why, under the raising derivation, raising of the XP from its base position via SpecCP (an A′-position) to its surface position does not violate the ban on improper movement, since this latter position is arguably neither a theta-position nor a Case position, the two types of A-positions generally assumed.

Where DP is an argument, V bears its own thematic function(s), in which case the thematic function of the RRC must collapse with one of V's thematic functions into a single function. This will presumably cause problems. Either the RRC's function will collapse with a function of V which is satisfied by an XP distinct from that to which the RRC is adjoined (presumably causing a violation of the theta-criterion, since the RRC will 'modify' two distinct arguments), or the RRC's function will collapse with the function of V which is satisfied by the XP to which the RRC is adjoined (which will be ruled out by Neeleman and van de Koot's condition banning satisfaction of a function by a node which contains that function). Thus, adjunction of an RRC to an argumental DP should be ruled out. In the case of clefts, however, V is assumed to bear no thematic functions.[31] Thus, in a DP-cleft with the structure in (43b), the thematic function of the cleft clause can percolate up to VP and then to IP without collapsing with any other thematic functions. Once it has reached the IP node, it may be satisfied by a D element in subject position. Under the bare phrase structure assumption that pronouns are zero-level D elements, the function may be satisfied or theta-bound by *it*, demonstratives and expletive *there*, accounting for the existence of *it*-clefts, demonstrative clefts and *there*-clefts (Davidse 2000). On the other hand, the function will not be satisfied by phrasal subjects. Thus, sentences such as (44) will be ruled acceptable, while sentences such as (45) will be ruled out:

(44) a. [$_D$ It] was THE SNAKE that the mongoose caught.
b. [$_D$ That] was THE SNAKE that the mongoose caught.
c. [$_D$ There] was only THE SNAKE that the mongoose caught.

(45) a. *[$_{DP}$ That/The animal] was THE SNAKE that the mongoose caught.
b. *[$_{DP}$ What John is] is A DOCTOR that John is.
c. *[$_{DP}$ THE SNAKE that the mongoose caught] is THE SNAKE that the mongoose caught.

So far, however, we might be led to expect, contrary to fact, that RRCs can in general adjoin to predicative XPs (e.g., AP), since these are also non-arguments. One explanation that suggests itself is based on the observation made earlier that AP-clefts only allow a raising derivation. A predicative AP has a thematic function to discharge, which it must do in its base position. This means that after raising, the AP will have no thematic function left to discharge. In a cleft, this is not a problem on the assumption that clefts are equatives, since the equated elements in an equative neither satisfy nor discharge a thematic function. Apart from this

[31] If equative *be* is indeed of category V, that is. The argument would not be affected if *be* were an I element, though previous arguments in the paper might be.

exceptional case, however, APs which have raised from inside a clause to head the structure will not be usable as predicates, and hence plausibly not usable at all outside clefts. Note that this explanation also extends to predicative-DP-clefts, which are also plausibly raising-only, given that they do not tolerate *wh*-operators in the cleft clause:

(46) *It is a doctor which he is.

Adjunction of an RRC to predicative DPs in general is thus ruled out in the same way as adjunction to predicative APs. The XP-relative analysis thus partially divorces the acceptability of adjunction to an XP from its syntactic category, instead relating it to the thematic status of the XP.[32]

Clearly much more work needs to be done to establish this as a viable analysis, however. In particular, the apparently paradoxical fact that the cleft clause seems to take both the initial *it* and the clefted XP as 'antecedents' must be captured somehow. Nevertheless, if this and other theoretical problems can be resolved, the XP-relative analysis seems to be well suited to capturing the various puzzling properties of English clefts.

References

Akmajian, Adrian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1. 149-168.

Aoun, Joseph & Yen-hui Audrey Li. 2003. *Essays on the representational and derivational nature of grammar*. Cambridge, MA: MIT Press.

Authier, J.-Marc and Lisa Reed. 2005. The diverse nature of non-interrogative *wh*. *Linguistic Inquiry* 36. 635-647.

Baltin, Mark. 1981. Strict bounding. In C. L. Baker and John J. McCarthy (eds.), *The logical problem of language acquisition*, 257-295. Cambridge, MA: MIT Press.

Bennis, Hans. 1986. *Gaps and dummies*. Dordrecht: Foris.

Bianchi, Valentina. 2000. The raising analysis of relative clauses: a reply to Borsley. *Linguistic Inquiry* 31. 123-140.

Biberauer, Theresa. 2003. *Verb Second (V2) in Afrikaans: a Minimalist investigation of word-order variation*. Ph. D. dissertation, University of Cambridge.

Borsley, Robert D. 1997. Relative clauses and the theory of phrase structure. *Linguistic Inquiry* 28. 629-647.

Bošković, Željko. 1997. Pseudoclefts. *Studia Linguistica* 51. 235-277.

Brame, Michael. 1968. A new analysis of the relative clause: evidence for an interpretive theory. Ms, MIT.

Büring, Daniel. 2003. D-trees, beans, and B-accents. *Linguistics and Philosophy* 26. 511-545.

Carlson, Greg N. 1977. Amount relatives. *Language* 53. 520-542.

[32] This view has precedents in the literature, notably Heggie (1993).

Chomsky, Noam. 1977. On *wh*-movement. In: Peter W. Culicover, Thomas Wasow and Adrian Akmajian (eds.), *Formal syntax*, 71-132. New York: Academic Press.
Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
Culicover, Peter W. and Michael S. Rochemont. 1990. Extraposition and the Complement Principle. *Linguistic Inquiry* 21. 23-47.
Davidse, Kristin. 2000. A constructional approach to clefts. *Linguistics* 38. 1101-1131.
Declerck, Renaat. 1983. Predicational clefts. *Lingua* 61. 9-45.
Delahunty, Gerald. 1982. *Topics in the syntax and semantics of English cleft sentences*. Ph.D. dissertation, University of California, Irvine.
É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74. 245-273.
É. Kiss, Katalin. 1999. The English cleft construction as a focus phrase. In Lunella Mereu (ed.), *Boundaries of morphology and syntax*, 217-229. Amsterdam: Benjamins.
Emonds, Joseph. 1976. *A transformational approach to English syntax: root, structure-preserving, and local transformations*. New York: Academic Press.
Gundel, Jeanette. 1977. Where do cleft sentences come from? *Language* 53. 34-59.
Halvorsen, Per-Kristian. 1978. *The syntax and semantics of cleft constructions*. Ph.D. dissertation, University of Texas, Austin.
Hedberg, Nancy. 2000. On the referential status of clefts. *Language* 76. 891-920.
Heggie, Lorie. 1993. The range of null operators: evidence from clefting. *Natural Language and Linguistic Theory* 11. 45-84.
Heycock, Caroline and Anthony Kroch. 1999. Pseudocleft connectedness: implications for the LF interface level. *Linguistic Inquiry* 30. 365-397.
Higginbotham, James. 1985. On semantics. *Linguistic Inquiry* 16. 547-593.
Higgins, Roger. 1973. *The pseudo-cleft construction in English*. Ph.D. dissertation, MIT.
Hulsey, Sarah and Uli Sauerland. 2006. Sorting out relative clauses. *Natural Language Semantics* 14. 111-137.
Jespersen, Otto. 1927. *A modern English grammar on historical principles. Part III: Syntax. Second Volume*. Heidelberg: Carl Winters.
Jespersen, Otto. 1937. *Analytic syntax*. New York: Holt, Rinehart and Winston.
Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
Kiss, Tibor. 2005. Semantic constraints on relative clause extraposition. *Natural Language and Linguistic Theory* 23. 281-334.
Kuno, Susumo. 1973. Constraints on internal clauses and sentential subjects. *Linguistic Inquiry* 4. 363-386.
Lees, Robert B. 1963. Analysis of the 'cleft sentence' in English. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 16. 371-388.
McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53. 99-149.
McCloskey, James. 1992. Adjunction, selection and embedded verb second. Ms., University of California, Santa Cruz.
Mezhevich, Ilana. 2003. On Russian 'expletive' *èto* and post-verbal clauses. FASL presentation.
Moro, Andrea. 1997. *The raising of predicates*. Cambridge: Cambridge University Press.
Neeleman, Ad and Hans van de Koot. 2002. The configurational matrix. *Linguistic Inquiry* 33. 529-574.
Percus, Orin. 1997. Prying open the cleft. *Proceedings of NELS* 27. 337-351.
Pereltsvaig, Asya. 2001. On the nature of intra-clausal relations: a study of copular sentences in Russian and Italian. Ph. D. dissertation, McGill University.

Pinkham, Jessie and Jorge Hankamer. 1975. Deep and shallow clefts. *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society* 11. 429-450.
Rochemont, Michael. 1986. *Focus in generative grammar*. Amsterdam: John Benjamins.
Safir, Kenneth. 1985. *Syntactic chains*. Cambridge: Cambridge University Press.
Sauerland, Uli. 1998. *The meaning of chains*. Ph. D. dissertation, MIT.
Schachter, Paul. 1973. Focus and relativization. *Language* 49. 19-46.
Smits, R. J. C. 1989. *Eurogrammar: the relative and cleft constructions of the Germanic and Romance languages*. Dordrecht: Foris.
Vikner, Sten. 1995. *Verb movement and expletive subjects in the Germanic languages*. Oxford: Oxford University Press.
Wexler, Kenneth and Peter W. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
Williams, Edwin. 1994. *Thematic structure in syntax*. Cambridge, MA: MIT Press.
Wirth, Jessica. 1978. The derivation of cleft sentences in English. *Glossa* 12. 58-81.

*Japanese wa-phrases that aren't topics**

REIKO VERMEULEN

Abstract

The Japanese particle *wa* is widely considered a marker for topic. This paper presents new evidence that contrary to this general view, *wa* does not only mark topic, but it also marks items that are simply discourse anaphoric. The evidence comes from the observation that a *wa*-phrase displays distinct syntactic behaviour depending on whether it is interpreted as a topic or discourse anaphoric. It is furthermore argued that topics in Japanese, contrastive or not, must appear in clause-initial position, which, as will be demonstrated, has the desired consequence that a clause may contain no more than one topic.

1 Introduction

The Japanese particle *wa* has been considered at least since Kuroda (1965) to be 'the' marker for topic in this language. In this paper, I argue that it does not in fact only mark topic: it also marks items that are simply discourse anaphoric independently of whether they are also topics. I will demonstrate that the pragmatic consideration of whether a given *wa*-phrase is interpreted as a topic or discourse anaphoric dictates the syntactic distribution of the phrase.

It is well-known that at the level of information structure, a topic-comment structure cannot be part of a background, but a focus-background structure can be inside a comment, an observation that was initially noted by the Prague School (Lambrecht 1994, Hajičová, et al 1998). In relation to how such constraints may be represented in the syntax, it has been argued by Rizzi (1997), and more recently by Neeleman & van de Koot (to appear), that the sister constituent of a fronted topic is interpreted as the comment, and that of a fronted focus is interpreted as the background. These two considerations together make predictions regarding the syntactic distribution of topic and focus, which are schematised in (1): a focus can

* Earlier versions of this paper were presented at Information Structure/Syntax Workshop at Queen Mary, University of London and On Linguistic Interfaces Conference at Belfast. I thank the participants for useful comments. Many thanks to Hans van de Koot, Ad Neeleman, Hitoshi Shiraki and Kriszta Szendrői for numerous helpful discussions. I also wish to thank my Japanese informants for their patient help. This paper is part of the output of an AHRC-funded project, 'A Flexible Theory of Topic and Focus Movement' (Grant nr. 119403).

follow a fronted topic, but a topic cannot follow a fronted focus. Neeleman & van de Koot show in detail that the predictions are borne out for Dutch. The cross-linguistic observation that topics generally precede foci also partially confirm these predictions (Hajičová, et al 1998).

(1) a. topic_i [$_{\text{comment}}$ FOCUS [$_{\text{background}}$ t_i]]
b. *FOCUS_i [$_{\text{background}}$ topic [$_{\text{comment}}$ t_i]]

In Japanese, however, a phrase marked with the putative topic marker *wa* can follow a fronted focus. In both of the following examples, taking a constituent that answers the *wh*-part of an immediately preceding question to be focus, the object *John-o* 'John-acc' is a focus. As shown in (2b), it is possible for a *wa*-phrase to follow it. Small capitals indicate stress.

(2) sono inu-ga dare-o kande-simatta no?
that dog-nom who-acc bite-closed Q
'Who did the dog bite?'
a. sono inu-wa kinoo kooen-de JOHN-O kande-simatta
that dog-wa yesterday park-at John-acc bite-closed
b. JOHN_i-O sono inu-wa kinoo kooen-de t_i kande-simatta
John-acc that dog-wa yesterday park-at bite-closed
'The dog bit John in the park yesterday.'

Thus, either the predictions in (1) are not correct for Japanese or the post-focal *wa*-phrase in (2b) is something other than a topic. I will argue for the latter position. More specifically, I claim that in sentences such as above, only the pre-focal *wa*-phrase is a topic, in the sense that it is what the rest of the sentence is about, while the post-focal *wa*-phrase is simply a discourse anaphoric item, in the sense that it has been previously mentioned (Reinhart 1981, Lambrecht 1994). The claim is based on the evidence that a *wa*-phrase that is interpreted as discourse anaphoric has a different syntactic distribution from one that is interpreted as a topic, and that a post-focal *wa*-phrase behaves like a discourse anaphoric *wa*-phrase.

On the basis of this syntactic evidence, I will argue furthermore that a topic *wa*-phrase must always appear in clause-initial position, while non-topical *wa*-phrases need not. This implies that there can be only one topic per clause, as there is only one clause-initial position. Thus, in a sentence containing multiple *wa*-phrases, it is predicted that only the left-most *wa*-phrase behaves like a topic. I will demonstrate that this is indeed true.

The rest of the paper is organised as follows. Section 2 investigates the general distribution of topics in Japanese, where the predictions in (1) are also shown to be

generally borne out in this language. I will argue that topics must in fact always appear in clause-initial position, a more restricted distribution than is suggested by (1). Section 3 is concerned with the well-known observation that topics can be associated with a position inside an island (Kuno 1973, Saito 1985). It is shown there that only topic *wa*-phrases display such a characteristic and not discourse anaphoric *wa*-phrases. Post-focal *wa*-phrases behave on a par with discourse anaphoric *wa*-phrases in this respect. Using the characteristics of topic *wa*-phrases identified in Sections 2 and 3, Section 4 shows that only the left-most *wa*-phrase in a sequence of multiple *wa*-phrases is the topic. Section 5 discusses implications of the findings reported here for a theory of the syntax-information structure interface. In particular, I will argue that the observations are difficult to capture under what is commonly known as the cartographic approach (e.g., Rizzi 1997, 2004, Belletti 2004), in which discourse-related information is explicitly represented in the syntax and discourse anaphoric items are often treated as topics. In Section 6, I note some puzzling contrasts between subject and object *wa*-phrases with respect to their interpretation and offer some suggestions. Section 7 concludes the paper.

2 The distribution of topics

There are certain discourse contexts in which an item must be interpreted as a topic. For instance, a request such as *tell me about X* forces *X* to be interpreted as a topic in the following utterance (Reinhart 1981). *X* here is often referred to as an 'aboutness' topic. In Japanese, in responding to such a request, the item *X* must be marked with *wa* and appear in a left-peripheral position ('thematic' *wa*-phrase in Kuno's (1973) terminology). The point is demonstrated below. Here, a request about a particular dog, *sono inu* 'that dog', is being made. (4a), in which *sono inu-wa* 'that dog-wa' appears in clause-initial position, is felicitous, while (4b), in which the *wa*-phrase occupies a non-clause-initial position, is infelicitous.

(3) sono inu-nituiteosiete-kudasai
that dog-about tell-please
'Tell me about that dog.'

(4) a. sono inu-wa kinoo kooen-de John-o kande-simatta
that dog-wa yesterday park-at John-acc bite-closed
b. # John$_i$-o sono inu-wa kinoo kooen-de t_i kande-simatta
John-acc that dog-wa yesterday park-at bite-closed
'The dog bit John in the park yesterday.'

(4b) is not ungrammatical, as shown by the acceptability of the same example in (2b). Although *John-o* is stressed in the latter and not in the former, the lack of

stress on the object is unlikely to be the source of infelicity, as it is possible in Japanese to A-scramble object to a position in front of the subject without placing a stress on it (Tada 1993, Saito 1992, Ishihara 2001). This characteristic is often noted for sentences where the subject bears the nominative case marker *ga*. However, it is also possible when the subject is a *wa*-phrase. The utterance in (6) is a well-formed response to (5). Here, the object is unstressed and the fact that it can bind into the subject *wa*-phrase shows that it has undergone A-scrambling.

(5) [Mary-to Bill]$_i$-o [otagai$_i$ -no sensee]-ga hometa no?
Mary-and Bill-acc each other-gen teacher-nom praised Q
Lit.: 'Did each other's teachers praise Mary and Bill?'

(6) Iya, [Mary-to Bill]$_i$-o [otagai$_i$ -no sensee]-wa sikatta rasii.
No, Mary-and Bill-acc each other-gen teacher-wa told.off seem
Lit.: 'No, it seems that each other's teachers told off Mary and Bill.'

Exactly the same observation obtains if the object is to be interpreted as an aboutness topic. As (8) shows, the object, about which a request is made in (7), must appear with *wa* and in a left-peripheral position. The nature of the empty category in (8a) will be discussed in the next section.

(7) sono boosi-nituite osiete-kudasai
that hat hat-about tell-please
'Tell me about that hat.'

(8) a. sono boosi$_i$-wa John-ga kinoo e_i kaimasita
that hat-wa John-nom yesterday bought
b. #John-ga sono boosi-wa kinoo kaimasita[1]
John-nom that hat-wa yesterday bought
'John bought that hat.'

Items usually referred to as contrastive topics, namely those that generally bear B-accent in languages such as English and German (Jackendoff 1972, Büring 1997, 2003 and references in the latter) display comparable behaviour. Typical functions of contrastive topics include introducing a new topic of discourse, narrowing down the referent of a topic or shifting the topic from one item to another. Contrastive topics in Japanese are marked with *wa* and carry a heavy stress (Kuno 1973, among

[1] For reasons unknown to me, it appears that an object *wa*-phrase does not easily sit adjacent to a verb. Throughout the paper, adverbials are inserted between object and verb to avoid this effect. I assume following Neeleman & Reinhart (1998), that a structure in which an argument has scrambled across an adverbial can be base-generated, hence the absence of an empty position below the adverbial in (8b). This does not affect the discussion in the main text.

others, cf. Kuroda 2005). It turns out that they too must appear in a left peripheral position. In the following discourse, information in relation to John is requested in (9). Not knowing the relevant information regarding John, a speaker may provide information with respect to Bill, as in (10). In doing so, he has shifted the topic of discourse from John to Bill, making *Bill-wa* a contrastive topic.[2] As demonstrated by the contrast between (10a) and (10b), *Bill-wa* cannot follow the fronted object *mame-o* 'beans-acc'.

(9) John-wa/ga kinoo-no party-de nani-o tabeta no?
John-wa/nom yesterday-gen party-atwhat-acc ate Q
'what did John eat at the party yesterday?'

(10) Hmm, John-wa doo-ka sira-nai-kedo,
well, John-top how-whether know-not-but,
'Well, I don't know about John, but...'
a. BILL-WA 8-zi-goro MAME-O tabeteita (yo)
Bill-nom 8 o'clock-around beans-acc eating particle
b. #MAME$_i$-O BILL-WA 8-zi-goro t_i tabeteita (yo)
beans-acc Bill-wa 8 o'clock-around eating particle
'As for Bill, he was eating beans around 8 o'clock.'

Similarly, where the object introduces a new topic as in (12), it must precede the subject.

(11) kinoo-no party-de dare-ga pasta-o tabeta no?
yesterday-gen party-at who-nom pasta-acc ate Q
'Who ate the pasta at the party yesterday?'

(12) Hmm, pasta-wa doo-ka sira-nai-kedo,
well, pasta-top how-whether know-not-but,
'Well, I don't know about the pasta, but...'
a. #BILL-GA MAME-WA 8-zi-goro tabeteita (yo)
Bill-nom beans-wa 8 o'clock-around eating particle
b. MAME$_i$-WA BILL-GA 8-zi-goro t_i tabeteita (yo)
beans-wa Bill-nom 8 o'clock-around eating particle
'As for the beans, Bill was eating them around 8 o'clock.'

The examples in (9)-(12) demonstrate that the predictions in (1) are generally borne out in Japanese. In particular, (10b) shows that a topic cannot follow a fronted focus, but a fronted topic can precede a focus, as in (12b). However, it is obvious that the distribution of topics is much more restricted than the schema in

[2] The set-up of the context is due to Neeleman & van de Koot (to appear).

(1) suggest. (12a) shows that a topic cannot follow a focus that is in-situ, and (4b) and (8b) illustrate that a topic cannot even follow a non-focus argument.

The general view in the literature is that aboutness topics must appear sentence-initially, but contrastive topics need not (Heycock to appear and references therein, but cf. Watanabe 2003). As the examples in (10) and (12) clearly demonstrate however, contrastive *wa*-phrases must also appear clause-initially in cases they are interpreted as topics. I propose therefore that a topic *wa*-phrase, contrastive or non-contrastive, must appear in clause-initial position, but a non-topical *wa*-phrase need not.[3,4] I formulate the claim as a constraint in (13), where *YP-wa* is a topic.

(13) Topic *wa*-phrases
* XP YP-wa

One consequence of the above constraint is that a post-focal *wa*-phrase, such as the one in (2b), cannot be a topic. So, what is it? A difference between (2), where a *wa*-phrase can appear post-focally, and (4), in which the same sentence is infelicitous, is that in the latter, the context forces the statement to be about the item marked by *wa*. In (2), on the other hand, it is merely mentioned in the preceding question and there is no sense in which the responses are forced to be about the *wa*-phrase. I conclude therefore that an unstressed *wa*-phrase that can appear in a post-focal position is simply discourse anaphoric and not a topic.

The constraint also predicts that there can only be one topic per clause, as there is only one clause-initial position. I will return to this prediction in Section 4, where it is shown to be correct.

Note that with a constraint such as the one in (13), the considerations in (1) seem to lose their predicative force. For instance, a sentence in which a topic follows a fronted focus, predicted to be infelicitous by (1b), can equally be ruled out by the fact that the topic is not in clause-initial position. One may therefore wonder whether the claim that there is a one-to-one mapping between syntax and information structure when there is topic / focus displacement, which led to the predictions, is relevant at all in Japanese. It indeed seems irrelevant, if the

---

[3] It is unclear whether the constraint in (13) applies to cases in which XP is an adverbial. Thus, in (8a), for instance, an adverbial such as *yesterday* cannot precede *sono boosi-wa* 'that hat-wa'. On the other hand, in (12b), the adverbial *8-zi-goro* '8 o'clock-around' can precede *mame-wa* 'beans-wa'. I will leave this issue with adverbials for further research, maintaining for now that (13) applies generally to arguments. Moreover, I will also ignore what Kuroda (1992) calls 'mini-topics', which are *wa*-phrases that modify objects and appear in pre-object position, as they do not seem to behave on a par with the tests discussed.

[4] Heycock (to appear) also notes with different examples involving subject *wa*-phrases that a contrastive *wa*-phrase, if sentence-initial, can function as a topic.

constraint is taken simply to be a restriction on the positioning of a topic and also a primitive in the grammar. However, if it is to be derived from other general properties of language, appealing to the idea that the sister constituent of a displaced topic is marked as the comment seems attractive. I speculate here that this particular idea may allow us to obtain a more insightful explanation for the existence of a constraint like (13) in Japanese. This language displays some properties associated with topic-prominent languages (Li & Thompson 1976). One such property is that the syntax typically reflects topic-comment structure. It follows then that a topic must appear in clause-initial position, marking its sister, namely all of the rest of the sentence, as the comment. Note that if the constraint is taken to be a primitive, the fact that the topic targets clause-initial position, as opposed to, say, a post-focal position, would have to be seen as a result of an arbitrary choice.

3 Topicalisation and island

In addition to the positions in which they can appear, there is a further syntactic difference between topic and discourse anaphoric *wa*-phrases. It concerns the structure in which they are licensed. There is consensus in the literature that aboutness topics are base-generated in a left-peripheral position, binding a *pro* internally to the clause in their thematic position, as illustrated below. Contrastive topics, on the other hand, are generally assumed to be derived by movement, based on evidence from facts involving Weak Crossover, resumptive pronouns and parasitic gaps (Hoji 1985, Saito 1985).

(14) $\text{Topic}_i \quad [_{IP} \quad \text{pro}_i \quad]$

This analysis explains the well-known observation that a topic can be associated with a position inside an island, such as a relative clause.[5] Thus, in the following example, *sono sinsi* ‘that gentleman’ is interpreted as the subject inside the relative clause. The existence of the empty pronominal *pro* can be seen from the fact that it is possible to overtly realise it (Perlmutter 1972, Kuno 1973, Saito 1985).

[5] Kuroda (1988), Sakai (1994) and Watanabe (2003) argue that topicalisation always involves movement. However, the possibility of linking to a position inside a relative clause is still considered to be a characteristic of (a construction that can feed into) topicalisation.

(15) sono sinsi$_i$-wa [$_{TP}$[$_{NP}$ Ø$_j$ [$_{TP}$ pro / kare$_i$-ga e_j kitei-ta] yoohuku]-ga
that gentleman-Top he-nom wearing-Past suit-GA
yogoretei-ta.
dirty-Past
'Speaking of that gentleman, the suit (he) was wearing was dirty.'
(modified from Kuno (1973: 249))

If the structure in (14) is associated with 'topics' rather than *wa*-phrases in general, which is in line with the claim in (13), then we predict a contrast between topic *wa*-phrases and discourse anaphoric *wa*-phrases. It should be possible for a topic *wa*-phrase, but not a discourse anaphoric *wa*-phrase, to be associated with a position inside a relative clause. The prediction is borne out. In responding to the request regarding *sono kodomo* 'that child' in (16), *sono kodomo-wa* 'that child-wa' can indeed be associated with a position inside a relative clause, as in (17).

(16) sono kodomo-nituite osiete-kudasai.
that child-about tell-please
'Tell me about that child.'

(17) sono kodomo$_i$-wa kyoo [$_{NP}$ [$_{TP}$ pro/kare$_i$-ga e_j kinoo katta] inu$_j$]-ga
that child-wa today he-nom yesterday bought dog-nom
John-o kande-simatta.
John-acc bite-closed
'As for that child, the dog that (he) bought yesterday bit John today.'

By contrast, as an answer to the question in (18), which merely mentions *sono kodomo* 'that child', the same sentence is infelicitous, as illustrated in (19).

(18) [$_{NP}$ [$_{TP}$ sono kodomo-ga e_j kinoo katta] inu$_j$]-ga dare-o kanda no?
that child-nom yesterday bought dog-nom who-acc bit Q
'Who did the dog that the child bought yesterday bite?

(19) #sono kodomo$_i$-wa kyoo[$_{NP}$ [$_{TP}$ pro$_i$/kare-ga e_j kinoo katta] inu$_j$]-ga/wa
that child-wa today he-nom yesterday bought dog-nom/wa
JOHN-O kande-simatta.
John-acc bite-closed
'The dog that the child bought yesterday bit John today.'

If all *wa*-phrases were topics and licensed uniformly in the syntax as in (14), their distribution in the syntax should not differ.

The above syntactic difference predicts furthermore that if a *wa*-phrase following a fronted focus is indeed a discourse anaphoric item, as I have argued it is in the previous section, then it should be impossible for it to be construed as an argument

inside a relative clause. The prediction is borne out. The sentence in (20) is plainly infelicitous, regardless of whether it follows the request in (16) or (18).

(20) #JOHN$_k$-O sono kodomo$_i$-wa kooen-de [$_{NP}$[$_{TP}$ pro$_i$ e_j kinoo katta]
John-acc that child-wa park-at yesterday bought
inu$_j$]-ga t_k kanda.
dog-nom bit
'The dog that this child bought yesterday bit John in the park.'

Crucially, the example becomes acceptable if the *wa*-phrase preceded the fronted focus *John-o*, as demonstrated by (21), allowing the *wa*-phrase to be interpreted as a topic. An appropriate preceding request would be (16).

(21) sono kodomo$_i$-wa JOHN$_k$-O kooen-de [$_{NP}$[$_{TP}$ pro$_i$ e_j kinoo katta]
that child-wa John-acc park-at yesterday bought
inu$_j$]-ga t_k kanda.
dog-nom bit

In sum, there are clear syntactic differences between topic *wa*-phrases and discourse anaphoric *wa*-phrases. The former, but not the latter, must occupy clause-initial position and can be associated with a position inside an island.

4 One topic per clause

I now turn to the prediction that follows from the constraint in (13) that there can be only one topic in a clause. It has often been noted that a clause in Japanese may contain multiple *wa*-phrases. Such clauses may contain multiple contrastive *wa*-phrases, but sound awkward with more than one non-contrastive *wa*-phrase (Kuno 1973, Tomioka 2007 and Heycock to appear, cf. Kuroda 1988). In the following examples, the object *Bill-wa* carries a heavy stress and is interpreted contrastively, while the subject *sono inu-wa* 'that dog-wa' is not stressed and does not imply any contrast.[6] The order between the two arguments can be reversed.

[6] Although most authors referred to in the main text use examples with one argument *wa*-phrase and one adverbial *wa*-phrase, the generalisation holds also of two argument *wa*-phrases. Kuno (1973: 48) cites examples with two argument *wa*-phrases.

(22) a. sono inu-wa BILL-WA moo sudeni kyonen kandeiru.
that dog-wa Bill-wa already last.year bite-perf.
b. BILL$_i$-WA sono inu-wa moo sudeni kyonen t_i kandeiru.
Bill-wa that dog-wa already last.year bite-perf.
'That dog has already bitten Bill last year.'

Given the constraint in (13), it should be impossible for both of the *wa*-phrases in (22a) or (22b) to be topics. More specifically, only the left-most *wa*-phrase in each example should display the characteristics we identified to be of topic *wa*-phrases in Sections 2 and 3.

The prediction is borne out. Firstly, in the discourse contexts that force a *wa*-phrase to be a topic, discussed in Section 2, the relevant *wa*-phrase must appear clause-initially. Thus, the request in (3), repeated below as (23), forces *sono inu* 'that dog' to be interpreted as the topic in the following utterance. As shown in (24), *sono inu-wa* 'that dog-wa' must precede the other *wa*-phrase *Bill-wa*.

(23) sono inu-nituite osiete-kudasai
that dog-about tell-please
'Tell me about that dog.'

(24) a. sono inu-wa BILL-WA moo sudeni kyonen kandeiru. (=(22a))
that dog-wa Bill-wa already last.year bite-perf.
b. #BILL$_i$-WA sono inu-wa moo sudeni kyonen t_i kandeiru.(=(22b))
Bill-wa that dog-wa already last.year bite-perf.

Similarly, in the context provided in (25)/(26), where *Bill-wa* is interpreted as a contrastive topic, it must occupy clause-initial position, as illustrated by the contrast in (26).

(25) sono inu-wa/ga John-o kanda no?
that dog-wa/nom John-acc bit Q
'Did that dog bite John?'

(26) Hmm, John-wa doo-ka sira-nai-kedo,
well, John-wa how-whether know-not-but,
'Well, I don't know about John, but...'
a. #sono inu-wa BILL-WA moo sudeni kyonen kandeiru. (=(22a))
that dog-wa Bill-wa already last.year bite-perf.
b. BILL$_i$-WA sono inu-wa moo sudeni kyonen t_i kandeiru. (=22b))
Bill-wa that dog-wa already last.year bite-perf.

While providing support for the constraint in (13), the above observations also suggest that the non-clause-initial *wa*-phrases in the felicitous (24a) and (26b) must

be something other than topics. For reasons discussed in Section 2, I believe that the unstressed *wa*-phrase s*ono inu-wa* 'that dog-wa' in (26b) is a discourse anaphoric item. On the other hand, the use of *wa* on a stressed object in-situ to indicate contrast, as on *Bill* in (24a), is widely discussed in the literature (Kuroda 1965, 2003, Kuno 1973, Saito 1985, Hoji 1985, Hara 2006, Tomioka 2007, Heycock to appear, among others). Particularly remarkable are the facts that the particle forces a contrastive interpretation on the object and it can project to generate VP-contrast. Thus, (24a) has the implicature that the dog has not yet bitten someone else, or has not yet committed any other violent act. I will not discuss the properties of this contrastive *wa* here. For the purpose of demonstrating that there can be only one topic per clause, it suffices to show that a stressed *wa*-phrase in non-clause-initial position is not a topic. If it is to be interpreted as a topic, it must appear in clause-initial position, as in the examples in (26) and also (12).[7]

Data involving relative clauses further confirm the claim that there can be only one topic per clause. The prediction is that an unstressed *wa*-phrase cannot be interpreted as an argument inside a relative clause if it follows a contrastive *wa*-phrase, but it can be if it precedes the contrastive *wa*-phrase. If it follows the contrastive *wa*-phrase, it cannot be a topic, but must be a discourse anaphoric *wa*-phrase, which does not have the privilege to be associated with a position inside a relative clause. The following example shows this is true. *Sono onnanoko* 'that girl' is intended to be construed as the subject of *katteiru* 'have' in the relative clause, but the sentence is not acceptable.

(27) #BILL$_k$-WA sono onnanoko$_i$-wa [$_{NP}$ [$_{TP}$ pro$_i$ e_j katteiru] inu$_j$]-ga kinoo
Bill-wa that girl-wa have dog-nom yesterday
t_k kande-simtta.
bite-closed
'As for Bill$_j$, and as for that girl$_i$, the dog that she$_i$ has bit him$_j$.'

[7] Some authors do in fact refrain from using the term 'contrastive topic' when referring to non-clause-initial stressed *wa*-phrases. Instead, they call such use of *wa* as 'contrastive *wa*' (Kuno 1973, Hara 2006, Kuroda 2005, among others).

On the other hand, if *Bill-wa* stays in-situ, allowing *sono onnanoko-wa* 'that girl-wa' to occupy clause-initial position, the sentence becomes acceptable:[8]

(28) sono onnanoko$_i$-wa [$_{NP}$ [$_{TP}$ pro$_i$ e_j katteiru] inu$_j$]-ga BILL-WA kinoo
that girl-wa have dog-nom Bill-wa yesterday
kande simtta.
bite-closed

The data considered in this section show clearly that there can be only one topic in a clause, lending further support to the claim that a topic must appear in clause-initial position.

5 How much information structure is in syntax?

In this section, I discuss implications of the above findings for a theory of the interface between syntax and information structure. Neeleman & van de Koot (to appear) view the schema in (1) as templates that constrain the mapping between syntax and information structure. If the topic moves in the syntax, the constituent that is the sister to the moved topic is interpreted as its comment, while if the focus moves, the constituent that is the sister to the moved focus is interpreted as its background. The point is illustrated in (29). These mapping rules are not associated with particular functional projections or positions in the syntactic structure. An item that is to be interpreted as focus, for instance, may undergo movement to an adjoined position internally to IP. Following Neeleman & van de Koot, I will call this the flexible approach.

---

[8] It is interesting to note that if the contrastive *wa*-phrase, *BILL-WA*, is fronted to a position following *sono onnanoko-wa* 'that girl-wa', as in (i), the sentence is infelicitous. The discussion in the main text suggests that this sentence should in fact be acceptable with *sono onnanoko-wa* being interpreted as an aboutness topic and *Bill-wa* as simply contrastive as in (22a)/(24a).

(i) #sono onnanoko$_i$-wa BILL$_k$-WA [$_{NP}$ [$_{TP}$ pro$_i$ e_j katteiru] inu$_j$]-ga kinoo t_k kande simtta.
that girl-wa Bill-wa have dog-nom yesterday bite-closed

Considering that movement requires motivation, it seems reasonable to assume that a *wa*-phrase moves in order to be interpreted as a topic, by moving to TopP a la Rizzi (1997), for instance. In (i), *Bill-wa* has undergone movement, and should therefore be a topic, but it is not in clause-initial position, as required by (13), hence the unacceptability. See also Section 2 for motivation for topic displacement. Some speakers find (28) marginal, but report a clear contrast between (i) and (28).

(29) a. XP_i $[_{YP} \quad t_i \quad]$ b. XP_i $[_{YP} \quad t_i \quad]$ (Syntax)
| |
Topic comment Focus background (Information Structure)

There is an alternative approach to the interface between syntax and information structure. It is widely known as the cartographic approach, where functional projections associated with interpretations such as topic and focus are projected in a rigid order in the CP-domain of a clause (Rizzi 1997, 2004, also Watanabe 2003 for Japanese). Items that are to be interpreted as topic or focus bear syntactic topic- and focus-features, respectively, and move to the specifier positions of TopicP and FocusP, where the features are checked by the functional heads. Typically, Topic Phrase is projected recursively in pre-Focus as well as post-Focus projection, as shown in (30). Some researchers argue that these discourse-related functional projections are also projected in the IP-domain (e.g., Belletti 2004, Grewendorf 2005, Paul 2006, cf. also Meinunger 2000)

(30) ... TopP* FocP TopP* ...

On the cartographic approach, discourse anaphoric items are very often treated as topics. Thus, they also undergo movement to the specifier position of a TopP, where they are licensed (Rizzi 1997, 2003, Belletti 2004, Grewendorf 2005, Meinunger 2000, among others). Considering that I have argued that the Japanese particle *wa* can mark topics as well as discourse anaphoric items, data such as (2), which shows that a *wa*-phrase can precede or follow a fronted focus, may at first sight appear to give support to a clausal structure like (30). However, as we saw in Sections 2-3, discourse anaphoric *wa*-phrases have a distinct syntactic distribution from topic *wa*-phrases, suggesting strongly that they are not licensed in the same manner in the syntax. Moreover, we also saw in Section 4 that there can only be one topic per clause.

It is true that discourse anaphoric items are often topics and topics are often discourse anaphoric. However, there are reasons to believe that this two-way association does not hold. Reinhart (1981), for instance, provides several examples illustrating the point. There is also some evidence that post-focal items in Italian that Rizzi argues occupy the lower SpecTopP positions in (30) are not topics. Vallduví (1992) and Samek-Lodovici (2006) argue that they are right-dislocated. Right-dislocated items are indeed usually discourse anaphoric, but do not behave like topics in that they cannot introduce new topics or be contrastive topics (Lambrecht 1994). Moreover, based on observations involving Weak Crossover effects, Benincá and Poletto (2004) claim that post-focal items are in fact foci. Whatever the correct analysis of post-focal items in Italian, it seems that they do

not show the same behaviour as pre-focal topics, raising doubt as to whether they should be analysed as 'topics'.

One may suggest that the post-focal Topic Phrases can perhaps be considered functional projections associated with discourse anaphoricity instead, bearing the label Disc.Ana.P, for instance. However, this move seems highly undesirable from a theoretical point of view. On the cartographic approach, particles such as *wa* have been taken as evidence for the existence of a Topic projection, *wa* being a morphological realisation of the Topic head. Admitting that *wa* can mark topic as well as discourse anaphoric items amounts to a claim that appears circular, namely that *wa* is a morphological realisation of Top^{o} only when the phrase to which it is attached is interpreted as a topic, but it is a morphological realisation of Disc.Ana^{o} if it marks a discourse anaphoric item.[9]

By contrast, the idea that the particle *wa* marks topics as well as discourse anaphoric items is more easily accommodated on the flexible approach. Under this approach, nothing forces the particle to be directly associated with a topic interpretation. What *wa* marks is a separate issue from the syntactic representation of a sentence containing a topic. A displaced item is interpreted as a topic by virtue of its sister constituent being interpreted as the comment by the discourse.

One may wonder then why *wa* marks topic at all, if topic can be identified by other means. Here, I speculate that this is to do with the fact that without the particle, it would be difficult to distinguish topicalisation from other kinds structures. Japanese does not have the phonological correlates of what Jackendoff (1972) calls A-accent and B-accent which are found in languages such as English to distinguish the interpretation of items bearing such accents: A-accent indicates focus, while B-accent (contrastive) topic (Hara 2006 and references therein).[10] Consequently, a sentence containing an object aboutness topic has the same intonation as a sentence in which the object has undergone A-scrambling to a position in front of the subject (Ishihara 2001), as in (31). The pre-verbal item, the subject *John-ga*, bears the main stress in both. Similarly, a sentence in which an object is interpreted as a contrastive topic has the same intonation as a sentence with a fronted accusative object, which is interpreted as a contrastive focus, as demonstrated in (32). Here, the main stress falls on the object, with the rest of the sentence deaccented (Ishihara 2001, Tomioka 2007).

---

[9] See Neeleman & van de Koot (to appear) for further arguments against positing functional projections for discourse anaphoric items.

[10] Japanese does have what is known as prominence lending rise, which has some pragmatic effects (Oshima in press). However, it does not appear to systematically distinguish topic from focus (cf. Hayashishita 2007).

(31) a. sono hon-wa John-ga yonda.
that book-wa John-nom read
'Speaking of that book, John read it.'
b. sono hon-o John-ga yonda.
that book-acc John-nom read
'John read that book.'

(32) a. SONO HON-WA John-ga yonda.
that book-wa John-nom read
'John read that book(, but not others).'
b. SONO HON-O John-ga yonda.
that book-acc John-nom read
'It is that book that John read.'

Thus, while the claim that the particle *wa* can mark topics as well as discourse anaphoric items does not argue directly against a cartographic approach to the interface between syntax and information structure, it seems to fit less comfortably than in an approach where the particle is not associated with a particular functional projection in the syntax.

6 Subject – object asymmetry

Before concluding, I would like to point out a puzzling contrast between subject and object *wa*-phrases. It is widely observed that an object *wa*-phrase in-situ must bear heavy stress and be interpreted as contrastive (e.g., Saito 1985, Watanabe 2003, Tomioka 2007). Thus, in the following discourse, it is not possible to mark the object with *wa*, as in (34a), although, just like the subject in (2), it is mentioned in the preceding question (33). The discourse anaphoric object must appear instead with the accusative case marker *o*, as in (34b) (Fiengo & McClure2002, Heycock to appear). The reply in (34a) is not completely infelicitous, but has an additional implicature that the book in question is being contrasted with another book.

(33) Dare-ga sono hon-o katta no?
Who-nom that book-acc bought Q
'Who bought that book?'

(34) a. #JOHN-GA sono hon-wa kinoo katta.
John-nom that book-wa yesterday bought
b. JOHN-GA sono hon-o kinoo katta.
John-nom that book-acc yesterday bought

A discourse anaphoric subject, on the other hand, is awkward at best with the nominative case marker *ga* and strongly favours being marked by *wa* (Kuno 1973, Tomioka 2007):

(35) sono inu-ga dare-o kanda no? (=(2a))
that dog-nom who-acc bit Q

(36) sono inu-#ga/wa JOHN-O kanda.
that dog-nom/wa John-acc bit

It is not the case that an object *wa*-phrase cannot be construed as discourse anaphoric. If the subject is a *wa*-phrase, contrastive or non-contrastive, the interpretation of an object *wa*-phrase in-situ as simply discourse anaphoric becomes much more acceptable, as shown below. The example in (38) is uttered in response to the request in (37), where the subject is already marked with *wa*. Here, the object *sono hon* 'that book', which is also mentioned in the request, can be marked with *wa* without giving rise to a contrast (Kawamura to apear). The context in (39)/(40), where the subject is a contrastive topic, illustrates the same point.

(37) Mary-wa tosyokan-de sono hon-o karita no?
Mary-wa library-at that book-acc borrowed Q
'Did Mary borrow that book in the library?'

(38) Ie, Mary-wa sono hon-wa kekkyoku honya-de KAIMASITA.
No, Mary-wa that book-wa in.the.end book.shop-at bought
'No, Mary bought the book in the end at the bookshop.

(39) John-wa sono hon-o kekkyoku ka-eta no?
John-wa that book-acc in.the.end buy-could Q
'Did John manage to buy that book?'

(40) Hmm, John-wa doo-ka sira-nai-kedo,
well, John-wa how-whether know-not-but
BILL-WA sono hon-wa denwa-de tyuumon-simasita.
Bill-wa that book-wa phone-by ordered
'Well, I don't know about John, but Bill ordered the book by phone.'

At present, I have no insightful explanation for the contrast between subject and object, illustrated in (33)-(36), or for the parasitic nature of object *wa*-phrases shown in (37)-(40). Following Tomioka (2007), I speculate that an account at least of the former observation may be found in the differences in the interpretations available to nominative subjects and accusative objects. A nominative subject disallows a discourse anaphoric interpretation, while an accusative object allows it.

A discourse anaphoric subject must therefore be marked with *wa*, while a discourse anaphoric object need not be.[11] One must of course still explain why there is such a contrast between nominative subjects and accusative objects. I will leave investigation of this contrast as well as the parasitic nature of object *wa*-phrases for future research.

7 Conclusion

In this paper, I have attempted to show that the particle *wa* does not only mark topic. It can also mark discourse anaphoric items. There is clear evidence that discourse anaphoric *wa*-phrases have a distinct syntactic distribution from topic *wa*-phrases, suggesting that the two types of *wa*-phrases should not be treated alike in the syntax. I claimed further that a topic must always appear in clause-initial position, but a discourse anaphoric *wa*-phrase need not. The prediction that followed from this claim that there can be no more than one topic per clause, was also shown to be supported by data involving multiple *wa*-phrases.

References

Belletti, A. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.) *The Structure of CP and IP: The Cartography of Syntactic Structures, Volume 2*. Oxford: OUP. 16-51.

Benincà, P., & C. Poletto. 2004. Topic, focus, and V2: defining the CP sublayers. In L.Rizzi (ed.) *The Structure of CP and IP: The Cartography of Syntactic Structures, Volume 2*. Oxford: OUP. 52-75.

Büring, D. 1997. *The Meaning of Topic and Focus*. London: Routledge.

Fiengo, R., & W. McClure. 2002. On how to use *-wa*. *Journal of East Asian Linguistics* 11, 5–41.

Grewendorf, G. 2005. The discourse configurationality of scrambling. In: J. Sabel and M. Saito (eds.) *The Free Word Order Phenomenon; Its Syntactic Sources and Diversity*. Berlin: Mouton de Gruyter. 75-135.

Hajičová, E., B. H. Partee & P. Sgall 1998. *Topic-focus Articulation, Tripartite Structures, and Semantic Content*. Kluwer: Dordrecht.

Hara, Yurie 2006. *Grammar of Knowledge Representation: Japanese Discourse Items at Interfaces*. PhD dissertation. University of Delaaware.

Hayashishita, J.-R. 2007. On the nature of the inverse scope. Poster presentation at On Linguistics Interfaces Conference. University of Ulster, Belfast, 1-3 June 2007.

[11] Tomioka claims that the possibility of *wa*-marking the subject is determined by what interpretations are available to the subject if it is marked by the nominative case marker *ga*. For instance, a nominative subject must be interpreted as indefinite in a matrix clause, while it need not be in an embedded clause (Kuno 1973). Consequently, he argues, the subject in a matrix clause must be marked by *wa* if it is to be interpreted as definite, but not in an embedded clause.

Heycock, C. to appear. Japanese -wa, -ga, and Information Structure. In Mamoru Saito & Shigeru Miyagawa (ed.) *Handbook of Japanese Linguistics*.

Hoji, Hajime 1985. *Logical Form Constraints and Configurational Structures in Japanese*. PhD dissertation, University of Washington.

Ishihara, Shinichi. 2001. Stress, focus and scrambling in Japanese. In *MIT Working Paper in Linguistics 39*, 142-175.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Kawamura, Michihiko. forthcoming. Topical contrast and 'contrastive topics' in Japanese. Proceedings of Workshop Information Structure in Context. IMS, Stuttgart, 15-17 November 2002.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, Mass.: MIT Press.

Kuroda, S.-Y. 1965. *Generative Grammatical Studies in the Japanese Language*. PhD dissertation, MIT.

Kuroda, S.-Y. 1988. Whether we agree or not: a comparative syntax of English and Japanese. In William J. Poser (ed.), *Papers from the Second International Workshop on Japanese Syntax*. CSLI, Stanford. pp. 103-143. [Also in *Linguisticae Investigationes* 12, 1-47. And reprinted in Kuroda, S.-Y. 1992. *Japanese Syntax and Semantics: Collected Papers*. Kluwer, Dordrecht.]

Kuroda, S.-Y. 1992 Judgment Forms and Sentence Forms. In *Japanese Syntax and Semantics: Collected Papers*. Kluwer, Dordrecht, 13-77.

Kuroda, S.-Y. 2005. Focusing on the matter of topic: A study of *wa* and *ga* in Japanese. *Journal of East Asian Linguistics*, 14:1–58.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. CUP.

Li, Charles N. & Sandra A. Thompson 1976. Subject and topic: A new typology of language. In Li, Charles N. (ed.) *Subject and Topic*. Academic Press, New York, 457-489.

Meinunger, André 2000. *Syntactic Aspects of Topic and Comment*. Amsterdam: John Benjamins.

Nakanishi, Kimiko. 2001. Prosody and Information Structure in Japanese: A Case Study of Topic Marker *wa*. In Noriko Akatsuka and Susan Strauss (eds.) *Japanese / Korean Linguistics*10, CSLI: Stanford, 434-447.

Neeleman, A. & T. Reinhart 1998. Scrambling and the PF interface. In M. Butt and W. Gueder (eds.) *The Projection of Arguments*. CSLI: Stanford. 309-353.

Neeleman, A & H. van de Koot to appear. Dutch scrambling and the Nature of Discourse Templates. *Journal of Comparative Germanic Linguistics*.

Oshima, David. In press. Morphological vs. phonological contrastive topic marking. In *Proceedings of CLS* 41.

Paul, Waltraud. 2006. What the topic is (not) about: the case of Mandarin Chinese, Ms. CRLAO, EHESS, Paris.

Perlmutter, David. 1972. Evidence for shadow pronouns in French. In P. M. Perantueau, J. N. Levi & G. C. Phares (eds.), *The Chicago Which Hunt*. Chicago Linguistic Society: Chicago. 73-105.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence Topics. *Philosophica* 27: 53-94.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.) *Elements of Grammar*. Dordrecht: Kluwer.

Rizzi, Luigi. 2004. Locality and left periphery. In: Adriana Belletti (ed.) *Structures and Beyond. The Cartography of Syntactic Structures, Vol. 3*. Oxford: Oxford University Press.

Samek-Lodovici, V. 2006. When right dislocation meets the left-periphery. *Lingua* 116: 836-873.

Sakai, Hiromu. 1994. Complex NP constraint and case conversion in Japanese. In Nakamura, M. (Ed.), Current Topics in *English and Japanese: Proceedings of the Second English/Japanese*

Comparative Syntax Workshop, Tohoku University, Sendai, Japan, 1994. Hituzi Syoboo, Tokyo. 179-203.
Saito, M. 1985. *Some Asymmetries in Japanese and their Theoretical Implications*. PhD diss, MIT.
Saito, M. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1. 69–118.
Tada, H. 1993. *A/A-bar Partition in Derivation*. PhD diss, MIT.
Tomioka, Satoshi. 2007. The Japanese existential possession: A case study of pragmatic disambiguation. *Lingua* 117, 881-902.
Vallduví, Enric. 1992. *The Informational Component*. New York: Garland.
Watanabe, Akira 2003. Wh and operator constructions in Japanese. *Lingua*, 113. 519–558.

Semantics and Pragmatics

*Generativity, Relevance and the Problem of Polysemy**

INGRID LOSSIUS FALKUM

Abstract

In this paper, I discuss the generative account of word meaning and interpretation of words in context proposed by Pustejovsky (1995). The goal of the generative lexicon theory (GLT) is to provide a formal account of 'logical polysemy', involving cases where a basic word meaning gives rise to different meanings in different contexts. My main claim is that the GLT fails to account for the flexibility of the processes involved in the modulation of lexical meaning in context, and thus makes a range of wrong predictions with regard to the derivation of compositional interpretations. I argue that the range of phenomena falling under the label of 'logical polysemy' on Pustejovsky's theory are better given a pragmatic account within the relevance-theoretic framework (Sperber & Wilson 1995, Carston 2002), which not only provides a simpler, more unified account, but also avoids the problems of misinterpretation associated with GLT.

1 Introduction: The Problem of Polysemy

There are several different ways in which a word can have more than one meaning. There is the traditional distinction between *homonymy*, involving cases where two distinct lexical items accidentally have the same linguistic form (e.g. the English form *bank*, which may refer to a financial institution or a riverside), and *polysemy*, referring to cases where a single lexical item has several (related) senses. Also, any word, whether monosemous, homonymous or polysemous, may have its meaning further modulated on a particular occasion of use by the linguistic or non-linguistic context in which it occurs. However, the distinctions between these cases of a word being associated with more than one meaning are not clear-cut. For instance, the question of where to draw the distinction between homonymy and polysemy remains a source of theoretical discussion in the lexical semantics literature. Similarly, the distinction between polysemy and cases of pragmatic modulation in context is not always easy to draw, and there is a question concerning what kind of phenomena fall under the label of 'polysemy'. Consider (1)-(4):

*I am grateful to Robyn Carston for her helpful comments and suggestions on previous drafts of this paper.

(1) a. *The newspaper* fired its editor.
b. Mary spilled coffee on *the newspaper*.
c. *The newspaper* revealed surprising details of the trial.
(2) Mary *began* a book.
(3) a. John *baked* a potato.
b. John *baked* a cake.
(4) a. Every chef needs a *good* knife.
b. 'The Kite Runner' is a *good* book.

In (1), *newspaper* has different meanings in (1a), where it refers to the publisher of the newspaper, (1b), where it denotes a physical object, and (1c), where it refers to the newspaper as an information type. The verb *began* in (2a) takes an NP as its complement, thereby leaving the VP underspecified with regards to the kind of activity associated with it (e.g. reading, writing, binding, etc.). The verb *bake* has different interpretations in (3a), where it has a change of state sense, and (3b), where it has a creative sense, while the adjective *good* expresses different properties in (4a) and (4b).

These kinds of examples have been given different treatments in the literature. The noun *newspaper* in (1) has been analysed as a *densely metonymous* term (Nunberg 1996: 126) with a number of interdefined denotations (e.g. organisation, edition or publication type, physical object or token, etc.), licensing a "widespread predicate transfer from the properties of one of its denotations to another". The verb *began* in (2) has been analysed as having a derived sense meaning 'began reading', where the process specified by the VP is made apparent from context or the other lexical items in the sentence (Langacker 1984: 185). The verb *bake* in (3) has been argued to be systematically ambiguous between a creative and a change of state sense (Atkins et al. 1988), while the adjective *good* in (4) has been analysed as an incomplete one-place predicate of individuals (Szabó 2001: 133-134), which is associated with a set of individuals only when additional information is provided (e.g. under one completion we get the set of individuals/objects which are knives that are good for cutting, under another we get the set of books that are good to read, etc.). Szabó postulates a variable in the lexical representation of *good* (and for many other adjectives too), which is fixed by the context in which the word is used.

Although the examples given in (1)-(4) arguably exhibit quite different characteristics, Pustejovsky (1995) has suggested treating them all as instances of a single phenomenon labelled 'logical polysemy', defined as the case where a word has two or more meanings, but where "there is no change in lexical category, and the multiple senses of the word have overlapping, dependent, or shared meanings" (ibid. 28). In the case of (1), Pustejovsky takes the polysemy of *newspaper* to be encoded directly into the semantics of the word. In (3), the different senses of *bake* are determined by the semantics of the complement noun, while the meaning of the

adjective *good* in (4) is assumed to be dependent on the semantics of the noun it modifies. Example (2), however, stands out as different from the previous cases, as the polysemy of the verb *begin* is viewed as related to the different complement types it may select for. Pustejovsky's notion of polysemy is thus a broad one.

With the goal of providing a formal account of examples such as (1)-(4), Pustejovsky (1991, 1995, 1998a, 1998b) develops a generative theory of word meaning structured around a representation and treatment of cases of 'logical polysemy'. The theory is designed to provide an account of how words can have different meanings in different linguistic contexts and how new senses can emerge as a result of the linguistic compositionality process, while at the same time limiting the number of entries in the lexicon. He rejects the view of lexical sense organisation assumed by so-called *SEL*s (*Sense Enumeration Lexicons*), according to which polysemy can be accounted for in terms of multiple listings of senses in the lexicon (where these senses are stored either in separate lexical entries or as multiple listings under a single lexical entry), as he believes that these accounts fail to explain the nature of the phenomenon, and miss the important generalisations that can be made on the basis of the available data. Instead, he proposes that the lexicon should reflect the following two assumptions: (1) the meaning of a lexical item cannot be divorced from the structure that carries it; and (2) word meaning should mirror our non-linguistic conceptual organising principles. The result is a lexicon consisting of complex lexical entries, over which a set of generative operations may apply in order to yield compositional interpretations such as the ones discussed for (1)-(4).

My aim in this paper is to discuss and assess the generative account of word meaning and interpretation of words in context proposed by Pustejovsky (1991, 1995, 1998a, 1998b). My main point is that this theory, while providing an interesting perspective on the relation between word meaning and compositional interpretations, nevertheless fails to account for the flexibility of the processes of sense modulation shown in (1)-(4), and makes a range of wrong predictions with regard to the derivation of compositional interpretations. I will argue that it is doubtful that the complex lexical representations assumed by the generative lexicon theory actually succeed in picking out the right denotations in the world. I will also argue, along with previous authors, that the theory does not provide a proper distinction between word meaning (what exactly is encoded by a word) and world knowledge. Finally, I will try to show that the range of phenomena falling under the label of 'logical polysemy' on Pustejovsky's theory are better given a pragmatic account within the relevance-theoretic framework (Sperber & Wilson 1995, Carston 2002), without having to posit complex lexical entries or any arbitrary cut-off point between aspects of general knowledge that are part of the lexicon and those that are part of the general knowledge system.

In the next section, the central tenets of the generative lexicon theory (henceforth GLT) will be presented. Section 3 discusses some objections to the theory, while section 4 presents a reanalysis of the data from the viewpoint of relevance theory.

2 The Generative Lexicon Theory (GLT)

The *generative lexicon* is viewed as a system consisting of (at least) four levels of linguistic representation (Pustejovsky 1995: 61). These are (1) the *Argument structure*, which includes the specification of number and type of logical arguments; (2) the *Event structure*, which is a definition of the event type of an expression, e.g. a state, process, or transition; (3) the *Qualia structure*, involving the essential attributes of an object or event as defined by the lexical item; and (4) the *Lexical inheritance structure*, describing the way in which a lexical item is globally related to other concepts in the lexicon. The semantics of a lexical item α is viewed as a structure consisting of these four components, which are connected by a set of generative devices, referred to as *type coercion*, *co-composition* and *selective binding*, that allow for the compositional interpretation of words in context.

2.1 Complex Lexical Entries

2.1.1 Argument and Event Structures. In GLT, the Argument structure of a word is taken to be the minimal specification of its semantics. The arguments for a lexical item α are represented as a list structure where the type of argument is directly encoded in the argument structure (ARGSTR), in the following way (Pustejovsky 1995: 67):

(5)
$$\begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \ldots \\ \text{ARG2} = \ldots \\ \text{D-ARG1} = \ldots \\ \text{S-ARG1} = \ldots \end{bmatrix} \\ \ldots \end{bmatrix}$$

As the representation in (5) shows, GLT posits different types of arguments for lexical items: (1) *True arguments* (ARG1 and ARG2) are those that must be syntactically realised; (2) *Default arguments* (D-ARG1) are not necessarily expressed syntactically, but are necessary for the logical well-formedness of the sentence; (3) *Shadow arguments* (S-ARG1) also refer to semantic content that is not necessarily expressed, but which is expressible under certain circumstances. The

different types of arguments are exemplified by the argument structures for the verbs *butter* and *build* in (6) and (7) (ibid.):

(6)
$$\begin{bmatrix} \textbf{butter} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \textbf{human} \\ \text{ARG}_2 = \textbf{phys\_object} \\ \text{S-ARG}_1 = \textbf{butter} \end{bmatrix} \\ \ldots \end{bmatrix}$$

(7)
$$\begin{bmatrix} \textbf{build} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \textbf{animate\_individual} \\ \text{ARG}_2 = \textbf{artifact} \\ \text{D-ARG}_1 = \textbf{material} \end{bmatrix} \\ \ldots \end{bmatrix}$$

The representations in (6) and (7) show that the verb *butter* takes a human as its subject, a physical object as its complement, as well as a shadow argument, *butter*, which is the material used in performing the buttering act, (e.g. $John_{ARG1}$ buttered *his* $sandwich_{ARG2}$ (*with* $butter_{S\text{-}ARG1}$)). Similarly, the verb *build* takes an animate individual and an artefact as its subject and complement respectively, but also takes a 'default argument', which is the material involved in the building, which may be explicit or implicit (e.g. $Mary_{ARG1}$ built *a* $house_{ARG2}$ (*out of* $bricks_{D\text{-}ARG1}$)).

The *Event structure* of a lexical item or a phrase defines the type of event associated with that lexical item or phrase. Events are classified as *processes, states* or *transitions,* and may also have subevental structures associated with them. These subevents may have one of the following three restrictions on their temporal ordering: '$<_\alpha$' describes a subevental structure involving a development process and a resulting state (e.g. *build*), '$< o_\alpha$' describes two simultaneous subevents, where one starts before the other (e.g. *walk*, where the movement of the legs brings about the movement of the body), and 'o_α' describes two completely simultaneous subevents (e.g. *accompany*, the person accompanying and the one being accompanied). The relative prominence of the subevents is also specified, where the *head* of an event is defined as "the most prominent subevent of the event structure of a predicate, which contributes to the "focus" of the interpretation" (Pustejovsky 1995: 72). The event structure for the verb *build* is given in (8):

(8)
$$\begin{bmatrix} \textbf{build} \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = \textbf{process} \\ E_2 = \textbf{state} \\ \text{RESTR} = <_\alpha \\ \text{HEAD} = E_1 \end{bmatrix} \\ \ldots \end{bmatrix}$$

This event structure includes two subevents, E_1 and E_2, of which the former is the head (i.e. the event that figures as the most prominent in interpretation), and specifies that E_1 must precede E_2.

2.1.2 Qualia Structure. The Qualia structure of a lexical item is the distinctive feature of Pustejovsky's theory, and is supposed to be "the set of properties or events associated with a lexical item which best explain what that word means" (Pustejovsky 1995:77). The qualia consist in a specification of four basic roles: the constitutive, formal, telic and agentive roles, each of which is taken to provide a distinct but essential component to uniquely determine the meaning of a word.

The *constitutive* role of a lexical item captures the relation between an object and its constituents, or proper parts (this may include a specification of material, weight, and of parts and component elements). The *formal* role specifies what distinguishes the object within a larger domain (and may include information about the orientation, magnitude, shape, dimensionality, colour and position of the object). The *telic role* defines the purpose and function of the object, if this may be specified, while the *agentive* role describes the factors involved in the origin or 'bringing about' of an object (including distinctions such as e.g. natural kind/artefact). The decompositional semantics of a lexical item thus includes a specification of these four aspects of meaning. (9) presents a schematic representation of the qualia structure for a given lexical item (Pustejovsky 1998a: 295):

(9)
$$\begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = x \\ \ldots \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \textbf{what } x \textbf{ is made of} \\ \text{FORMAL} = \textbf{what } x \textbf{ is} \\ \text{TELIC} = \textbf{function of } x \\ \text{AGENTIVE} = \textbf{how } x \textbf{ came into being} \end{bmatrix} \end{bmatrix}$$

The representation in (10) is a minimal semantic description for the noun *novel*, including values for each of the qualia roles (adapted from Pustejovsky 1991: 427, 1995: 78):

(10)
$$\begin{bmatrix} \textbf{novel} \\ \ldots \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \textbf{narrative(x)} \\ \text{FORMAL} = \textbf{book(x)} \\ \text{TELIC} = \textbf{read(e,y,x)} \\ \text{AGENT} = \textbf{write(e}'\textbf{,z,x)} \end{bmatrix} \end{bmatrix}$$

This representation shows how the noun *novel* is supposed to encode information about particular properties and activities associated with it (in this representation, *x* is a distinguished variable which represents the object itself). A novel is a narrative, which has the form of a book; its purpose is to be read and it comes into being by a process of writing (*e* and *e´* refer to the event types of reading and writing, which are both transitions).[1] Pustejovsky assumes that a qualia structure may be defined for all types of lexical items, but that not all lexical items need carry a value for each role. For instance, the lexical representation associated with the verb *break* is taken to be the following (Pustejovsky 1995: 80):

(11)
$$\begin{bmatrix} \textbf{break} \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = \textbf{e}_1\textbf{:process} \\ E_2 = \textbf{e}_2\textbf{:state} \\ \text{RESTR} = <_\alpha \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{FORMAL} = \textbf{broken(e}_2\textbf{,y)} \\ \text{AGENTIVE} = \textbf{break\_act(e}_1\textbf{,x,y)} \\ \ldots \end{bmatrix} \end{bmatrix}$$

The event structure encoded by the causative verb in (11) includes two subevents: a process followed by a resulting state. These subevents are assumed to map directly

[1] However, there is more than this to the qualia structure for *novel*, since this noun is taken to be a subtype of *book*, which is represented as a *Lexical Conceptual Paradigm* (on this notion, see section 2.1.2.1), being polysemous between a physical object and a textual information sense. The noun *novel* inherits this polysemy from *book*, but its lexical representation is more informative in the sense that it includes a specification of the sort of textual information associated with it (i.e. a narrative) (see Pustejovsky 1995: 156).

into the formal and agentive qualia roles, where the resulting state corresponds to the formal role and the causing event to the agentive role.

2.1.2.1 Lexical Conceptual Paradigms. So how does Pustejovsky use this idea of complex lexical entries to account for the examples of polysemy we set out in the introduction? Let us consider again the example in (1), repeated here as (12):

(12) a. *The newspaper* fired its editor.
b. Mary spilled coffee on *the newspaper*.
c. *The newspaper* revealed surprising details of the trial.

In this case, where *newspaper* is used to denote an organisation (12a), a physical object (12b), and an information type (12c), GLT takes the polysemy to be encoded directly into the lexical representation of the object. The notion of a *Lexical Conceptual Paradigm* (lcp) provides a means of characterising a lexical item as a *meta-entry* (Pustejovsky 1995, 1998b). Consider the uses of the noun *book* in (13):

(13) a. *The book* is sitting on the coffee table.
b. Mary found *the book* interesting.
c. Paul has started writing *his second book*.

This noun is analysed as being polysemous between a physical object sense (13a), a textual information sense (13b), as well as the combination of the two senses (13c), all of which are captured in the meta-entry for *book*. The lcp acts as a type constructor, which creates a complex type for a term α, carrying the polysemous senses σ_1 and σ_2, on the basis of the following rule (Pustejovsky 1995: 93):

(14) $\underline{\alpha: \sigma_1 \quad \alpha: \sigma_2}$
$\text{lcp}(\alpha): \sigma_1 . \sigma_2$

The lcp of a polysemous lexical item may then be represented as in (15), consisting of the complex type (a Cartesian product), as well as the two base senses (ibid.):

(15) $\text{lcp} = \{\sigma_1 . \sigma_2, \sigma_1 , \sigma_2\}$

Pustejovsky thus suggests that three senses of a lexical item, clustered in one single lexical entry, may be constructed from two base types. The lexical representation for *book* may thus look like (16) (Pustejovsky 1995: 101):

(16)
$$\begin{bmatrix} \textbf{book} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textbf{x:information} \\ \text{ARG2} = \textbf{y:phys\_obj} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \textbf{information.phys\_obj\_lcp} \\ \text{FORMAL} = \textbf{hold(y,x)} \\ \text{TELIC} = \textbf{read(e,w,x.y)} \\ \text{AGENT} = \textbf{write(e´,v,x.y)} \end{bmatrix} \end{bmatrix}$$

The same rule can be used to give a lexical representation for the senses of *newspaper* (Pustejovsky 1995: 156):

(17)
$$\begin{bmatrix} \textbf{newspaper} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textbf{x:org} \\ \text{ARG2} = \textbf{y:info . physobj} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \textbf{org.info.physobj\_lcp} \\ \text{FORMAL} = \textbf{y} \\ \text{TELIC} = \textbf{read}(e_2,\textbf{w,y}) \\ \text{AGENT} = \textbf{publish}(e_1,\textbf{x,y}) \end{bmatrix} \end{bmatrix}$$

In this way, the different senses of *newspaper* will be encoded in its lexical representation as part of a lexical conceptual paradigm in the qualia structure (where the 'info.physobj' sense is itself an lcp). However, as Pustejovsky himself notes, the way in which the noun *newspaper* denotes is different from that of *book*, since the former cannot be used to denote the combination of the senses stored in the lcp, while the latter can, cf. (13c). I will return to this issue in section 3.2.

Nunberg (1996: 126), who analyses *newspaper* as a densely metonymous term, notes that a problem with these kinds of metonymies is that it is often hard to decide which use of the word is 'basic' and which uses are 'derived', and, as a corollary, to know in which direction the transfer operates. The GLT seems to solve this problem by letting all senses be encoded as a lexical conceptual paradigm in the qualia structure for the word; hence there is no need to assume that one sense is more 'basic' than the other senses.

2.2 Generative Mechanisms

Now let us consider the remaining cases of 'logical polysemy' that we looked at in the introduction. The examples in (2)-(4) are different from the previous case in

that their relevant interpretations are assumed to be conditioned by a set of generative operations, described as *type coercion*, *co-composition*, and *selective binding* (Pustejovsky 1995), whose application is determined by linguistic context.

2.2.1 Type Coercion. The semantic transformation of type coercion is defined as a "semantic operation that converts an argument to the type that is expected by a function, where it would otherwise result in a type error" (Pustejovsky 1995: 111). This generative mechanism accounts for examples of the sort in (2), repeated here as (18a). Compare (18a-c):

(18) a. Mary began *a book*.
b. Mary began *reading a book*.
c. Mary began *to read a book*.

In (18a) the verb *begin* selects an NP as its complement (specifying an object), while in (18b) and (18c) its verbal complements are both of the semantic type 'event'. Instead of assuming that there are different lexical entries for *begin*, that is, one for each complement type the verb may select for, this verb is treated as having a single lexical entry, with the following event and argument structures (cf. Pustejovsky 1995: 116):

(19)
$$\begin{bmatrix} \textbf{begin} \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = \textbf{transition} \\ E_2 = \textbf{transition} \\ \text{RESTR} = < o_\alpha \end{bmatrix} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textbf{x: human} \\ \text{ARG2} = \textbf{e}_2 \end{bmatrix} \end{bmatrix}$$

The lexical representation in (19) shows that *begin* selects for an event as its complement type (cf. ARG_2). In cases where this requirement is not directly satisfied by the surface syntactic structure, as in (18a) above, coercion applies in order to change the type of the complement NP into an event. We saw that the qualia structure for the noun *book* in (16) above included a specification of two events in the agentive and telic roles (namely those of reading and writing), hence the denotation of the NP *a book* is coerced into an event denotation consistent with the interpretations 'Mary began to write a book' and 'Mary began to read a book' respectively.

2.2.2 Co-Composition. The generative operation referred to as *co-composition* (Pustejovsky 1991, 1995: 122), accounts for the example in (3), repeated below as (20). This mechanism operates over structures that allow for more than one function application.

(20) a. John *baked* the potato.
b. John *baked* the cake.

As we have seen, *bake* has different meanings in (20a), where it has a change of state sense, and (20b), where it has a creation sense. Rather than treating these two senses of *bake* as separate lexical entries (or, alternatively, as different senses stored under a single lexical entry), Pustejovsky assumes only one sense for *bake*, a change-of-state reading, and that any other readings are derived through the generative mechanism of co-composition, which applies at the level of the entire verb phrase. This mechanism allows for the information carried by a complement to act on the governing verb, by taking the verb as argument and shifting its event type. Consider the lexical representation for the verb *bake* (Pustejovsky 1995: 123):

(21)
$$\begin{bmatrix} \textbf{bake} \\ \text{EVENTSTR} = \begin{bmatrix} E_1 = \textbf{e}_1\textbf{:process} \\ \text{HEAD} = \textbf{e}_1 \end{bmatrix} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textbf{1} & \begin{bmatrix} \textbf{animate\_ind} \\ \text{FORMAL} = \textbf{physobj} \end{bmatrix} \\ \text{ARG2} = \textbf{2} & \begin{bmatrix} \textbf{mass} \\ \text{FORMAL} = \textbf{physobj} \end{bmatrix} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \textbf{state\_change\_lcp} \\ \text{AGENTIVE} = \textbf{bake\_act}(\textbf{e}_1\textbf{,1,2}) \end{bmatrix} \end{bmatrix}$$

Here *bake* is specified as having a change of state denotation. When combined with the complement *potato*, whose agentive role is specified as 'natural kind', the change of state interpretation of *bake* remains unchanged. Pustejovsky sees this as a case where the noun does not, on its own, contain a specification of an event structure and hence the process of co-composition does not apply. The semantics of the complement *cake* in (20b) does, however, shift the core meaning of *bake* from a change-of-state to a creation sense. This is a result of the lexical representation associated with *cake*, given in (22) (Pustejovsky 1995: 123):

(22)
$$\begin{bmatrix} \textbf{cake} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = \textbf{x:food\_ind} \\ \text{D-ARG1} = \textbf{y:mass} \end{bmatrix} \\ \text{QUALIA} = \begin{bmatrix} \text{CONST} = \textbf{y} \\ \text{FORMAL} = \textbf{x} \\ \text{TELIC} = \textbf{eat}(e_2,z,x) \\ \text{AGENTIVE} = \textbf{bake\_act}(e_1,w,y) \end{bmatrix} \end{bmatrix}$$

The qualia structure for *cake* specifies that there is an event associated with this object's coming into being, namely the process of baking. In this case, the agentive role of the complement noun makes reference to the very process denoted by the verb phrase, a relation referred to as *co-specification* (Pustejovsky 1991, 1995). The 'creation' sense of *bake* thus arises from the semantics of *cake* by virtue of its being an artefact (i.e. originating out of the act of baking). On this account, therefore, it is not the verb *bake* itself that is polysemous, but the difference in meaning between *bake a cake* and *bake a potato* is the result of principles of semantic composition.[2] A verb such as *use*, which is said to be underspecified with respect to the activity being performed, is analysed along similar lines (examples from Pustejovsky 1995):

(23) a. John used *the new knife* on the turkey.
b. Mary has used *soft contact lenses* since college.
c. This car uses *unleaded gasoline*.
d. My wife uses *the subway* every day.

In these cases, it is the object NPs, by virtue of their qualia structures, which contribute towards specifying the interpretation of *use*. In (23a) *use* refers to the action of *cutting*, in (23b) to *wearing* the contact lenses, and similarly in (23c) and (23d) the actions expressed by *use* are determined by the kind of event associated with the object NPs. The telic roles of the objects thus play an important role in determining which action is expressed by the verb; it is assumed that our

[2] If the lexical entry for *potato* is taken to be the same as that for *cake* except for the agentive role, which could be specified as **grow(e,w,y)**, then *potato* would, just like *cake*, have an event structure associated with it. Pustejovsky (1991: 422) acknowledges that "relative to the process of *growing*, the noun *potato* does denote an event". He does not, however, make it clear whether he thinks that *potato* acquires this event structure when combined with the verb *grow* (e.g. *Mary grew the potato*), in which case it would be the verb that is modifying the meaning of its complement (and not vice versa), or if this event structure is somehow cancelled when *potato* is combined with a verb such as *bake*.

understanding of sentences such as the ones in (23) is facilitated by default interpretations of properties and activities associated with the objects. In syntactic terms this means that the semantics of a governing expression (in this case the head of VP) is determined by the semantic content of the expression it governs (in (23) by the object NPs).

2.2.3 Selective Binding. The third generative mechanism presented in Pustejovsky (1991, 1995, 1998a, 1998b) is that of selective binding, which is described as a process whereby an adjective "is able to make available a selective interpretation of an event expression contained in the qualia for the head noun" (Pustejovsky 1995: 128). This is the process that is assumed to take place for an evaluative adjective such as good in (4), repeated here as (24):

(24) a. Every chef needs a *good* knife.
b. 'The Kite Runner' is a *good* book.

In (24a), it is assumed that *good* selectively modifies the event description given by the telic role of the noun *knife* (i.e. the action of cutting), so as to yield the interpretation 'a knife that cuts well'. Similarly, in (24b), the adjective acts as an event predicate for the function of the object to yield the interpretation 'a good read', rather than modifying the entire denotation of the noun.

3 Problems with the Generative Lexicon Theory

Pustejovsky's theory of word meaning takes the semantics of a lexical item to involve a specification of argument, event, qualia and lexical inheritance structure.[3] It is assumed that verbs, nouns and adjectives may all function actively in determining the denotation of a phrase, as a result of the application of the generative operations of type-coercion, co-composition and selective binding. The theory thus seems to fulfil its goal of accounting for polysemy without having to assume multiple listings of word meanings stored in the lexicon, save in the case of

[3] The notion of lexical inheritance structure is not given much attention in Pustejovsky (1995). Pustejovsky (1991: 433) distinguishes between two inheritance mechanisms for representing the conceptual relations in the lexicon: *fixed* inheritance, involving e.g. the hypernym/hyponym relation (e.g. *book/novel*), and *projective* inheritance, which is assumed to operate "*generatively* from the qualia structure of a lexical item to create a relational structure for ad hoc categories." (ibid.). This mechanism is supposed to account for a relation such as that between e.g. *escape* and *prisoner* in a sentence like *The prisoner escaped last night*, and for the information conveyed by this sentence having a more prototypical character than that of e.g. *The prisoner ate dinner last night*.

words encoding a lexical conceptual paradigm, where two of the meanings will be listed in the qualia structure. The theory also makes some interesting observations with regard to the interaction between word meanings in linguistic context. However, I believe that, ultimately, the theory is unworkable; first of all because it makes a range of wrong predictions, and, second, because the correct predictions that it makes are better treated pragmatically. The details of such a pragmatic account will be described in the next section. In this section, I will consider three central objections to the GLT, concerning (1) the way certain inferences seem to come out as strictly necessary entailments as a result of the generative operations over lexical entries; (2) the (in)ability of the complex lexical representations to pick out the right denotations in the world, and (3) the extent to which world knowledge is assumed to be built into the lexicon. The two latter issues are in fact closely related, as they both concern the kind of information stored as the semantic component of the lexical representation of a word in the GLT.

3.1 Necessary Inferences?

An important objection against the GLT, which has been advanced by a number of authors (e.g. Blutner 2002, Fodor & Lepore 1998), is the way certain inferences seem to come out as strictly necessary entailments as a result of the application of the generative operations over lexical entries. Consider the use of *bake* in (25), which is Pustejovsky's prime example of the process of co-composition:

(25) Mary *baked* the pizza.

Here Pustejovsky's account would predict a creative reading of *bake* (*pizza* being an artefact in much the same way as *cake* is), but a non-creative reading, on which Mary is simply heating up a frozen pizza, would be just as likely to be the preferred reading as the creative one. The inference via co-composition from *Mary baked the pizza* to 'Mary baked the pizza (creatively)' is therefore defeasible, hence not strictly necessary. The same point applies to inferences involving type coercion and selective binding. For example, *want a cigarette* is type coerced into 'want to smoke a cigarette' (Fodor & Lepore 1998: 274), but it could also be interpreted as e.g. 'want to hold a cigarette' in a context where someone has recently quit smoking but wants to hold a cigarette to keep the craving away (or as 'want to smell a cigarette'). By selective binding, *a fast book* yields 'a book that can be read in a short time', but in a context of book-sellers, *fast books* could be interpreted as the ones that sell quickly, and *slow books* as the ones that stay on the shelves for a long time (and there are numerous other interpretive possibilities).

Blutner (2002), following Fodor and Lepore (1998), points out that there is also a problem related to the restrictiveness of the coercion mechanism, in that the

account incorrectly predicts that *begin a car* should mean 'begin to drive a car' and *begin a thermometer* should mean 'begin to measure the temperature'. This problem also extends to cases of selective binding, where the account would predict that e.g. *a fast cake* should be interpreted as 'a cake that is fast to eat' (cf. the telic role for *cake* given above). Copestake and Briscoe (1996), who combine an HPSG-like approach to syntax with Pustejovsky's notion of qualia structure in their account of polysemy, try to solve this problem by treating compositional interpretations from the qualia as defeasible. An NP such as e.g. *fast typist* is normally interpreted as someone who types rapidly, but by allowing for specific (linguistic) contexts to suggest other interpretations that are not inherent to the qualia of the lexical item, the NP *fast typist* might also be used to refer to a typist who is e.g. running quickly. In such cases of non-default interpretations, the linguistic context is expected to explicitly give the exceptional component used in the interpretation.

However, the move to (defeasible) default interpretations does not seem to solve the problem. Consider the type coercion mechanism proposed by Pustejovsky again. Even if we assume that the interpretations generated by this mechanism are default but defeasible, we still cannot prevent it from computing the incorrect readings described above. Also, the coercion mechanism fails to indicate which is the default interpretation in cases where there is more than one possible interpretation compatible with the selection requirement of a verb, such as in e.g. example (18) above, where *Mary began a book* is compatible with both the interpretation 'Mary began to write a book' and 'Mary began to read a book'. If we assume that there are default (but defeasible) compositional interpretations associated with a qualia structure, which interpretation of *Mary began a book* should we take to be the default one, and why? Both possible interpretations may be generated by the process of type coercion, acting upon information specified by the qualia structure for *book*. Assuming that, say, 'Mary began to read a book' is the default interpretation, does this mean that the information specified by the telic role (which involves 'read') of a noun is more 'default' that that specified by the agentive role (which involves 'write')? In these cases, therefore, it seems that the coercion mechanism just generates another underspecified structure (i.e. two possible event interpretations), and it would be a task for pragmatics to choose between them.

Also, the move to defeasible default interpretations raises another kind of question related to whether we want our semantics to include defeasibility when we have already got it as part of our pragmatic system. No reason is given for not simply leaving it to pragmatics, using information from the immediate context as well as general world knowledge, to construct the intended interpretation.

3.2 Adequate Lexical Semantic Representations?

In the GLT, the notion of 'qualia structure', and the four types of roles (constitutive, formal, telic, and agentive roles) seem to apply mainly to the description of the semantics of nouns, and to nouns denoting concrete objects in particular. In fact, with only a few exceptions, the nouns discussed in Pustejovsky (1995) all denote physical objects. Pustejovsky maintains that a qualia structure can be defined for all types of lexical items, but that not all lexical items need carry a value for each role. It is, however, hard to see how the four types of roles associated with a qualia structure are supposed to apply to e.g. prepositions, connectives, etc. Even for lexical categories such as verbs, adjectives, and adverbs, it seems like a category mistake to try to describe their semantic content in terms of a qualia structure, as defined in (9). Does it make sense to speak of e.g. what an action/process is made of (i.e. its constitutive role) or what the function of an adjective is (i.e. its telic role)? Furthermore, there seems to be a problem with regard to Pustejovsky's specifications of argument structure for a number of lexical items. The lexical representation he gives for e.g. *build* in (7) specifies that this verb takes an animate individual as its subject. However, this makes it unclear how he would accommodate examples such as e.g. *The army built the wall in a day*, *The cognitive system builds and stores representations in memory* into this analysis. Similarly, the subject of the verb *begin* in (19) is specified as 'human', which seems to rule out examples such as e.g. *The engine began to make a strange noise*, etc. There seems, in fact, to be exceptions to all of Pustejovsky's uses of 'animate', 'human', etc., as they figure as part of his lexical entries.

One may also question GLT's notion of so-called 'default' and 'shadow' arguments, as arguments which are not necessarily syntactically expressed, but nevertheless encoded as part of the semantic representation for a lexical item. The question is whether the information contained by such 'arguments' should be regarded as linguistic. In the case of *John buttered his sandwich* (involving the shadow argument *butter*) and *Mary built a house* (involving a default argument specified as 'material') it is clearly a matter of metaphysical necessity that the bread has to have been buttered with some kind of material (e.g. butter, margarine, peanut butter, chocolate butter, etc.) and that the house has to have been built out of some material or other (e.g. bricks, wood, stone, corrugated iron, etc.), but this fact alone does not entail that they are semantically necessary. Consider an alternative view based on the concept of an 'unarticulated constituent', that is, a constituent of the statement made (by uttering a sentence in a given context) which does not correspond to any linguistic entity in the sentence. Recanati (2002: 305-306) makes a distinction between *metaphysical* and *communicational* unarticulated constituents. For something to count as a communicational unarticulated

constituent, it must be part of "what the speaker means by his or her utterance" (ibid. 306), while a metaphysical unarticulated constituent is subject to no such requirement. In the event of a speaker uttering *I've had a large breakfast* as a response to the question *Are you hungry?*, it is assumed that the hearer has to identify the temporal location of the breakfast event (as the day of the utterance) in order to be able to derive the intended implicated meaning (implicature) that the speaker is not hungry, hence this would count as an unarticulated constituent in the communicational sense. A case of a metaphysical unarticulated constituent would be the location of the dancing in an utterance of *Juliet danced*; obviously there is, as a matter of metaphysical necessity, a place at which Juliet danced but comprehension of the proposition expressed by the utterance does not require the recovery of a location constituent. The GLT 'shadow' and 'default' arguments posited in the case of *John buttered his sandwich* and *Mary built the house* would, on this account, be cases of metaphysical unarticulated constituents, since (in most instances, at least) the hearer will be able to understand these utterances without having to identify the material with which the sandwich was buttered or out of which the house was built. While a case could be made for there being some covert linguistic entity corresponding to the communicational constituents (see Stanley 2000), there is no such case for the metaphysical unarticulated constituents, which suggests that Pustejovsky's shadow and default arguments should not be regarded as linguistic, but rather as part of our general knowledge about the world.

The idea of lexical conceptual paradigms as a way to account for metonymous senses of a word also involves some problems. In cases like *book* and *newspaper*, it seems that Pustejovsky's account is not as different from the sense enumeration lexicon approach as he makes out, in that the two basic senses forming the input to the lcp operator are both listed in the lexical representations for these words. Furthermore, the fact that the type constructor generates a sense for *newspaper* for which there is no corresponding denotation (i.e. the 'org.info.physobj' sense) does not seem to favour Pustejovsky's analysis. We would not be happy with a generative grammar that generated ungrammatical sentences; similarly, a generative lexicon that generates impossible senses is unsatisfactory.

Finally, consider the lexical representation for the noun *cake* given in (22) above. If we were to provide a representation of the noun *pizza* discussed in 3.1 along the same lines, it is actually hard to see how this semantic representation would differ from the one given for *cake*, as all the properties and activities associated with *cake* in (22) may also be said of *pizza* (and possibly of other artefacts of similar type). How, then, are these lexical representations supposed to pick out the right denotations in the world? It seems that Pustejovsky's lexical entries do not specify sufficient conditions for the application of a concept. Furthermore, as pointed out above in the discussion of *begin a book*, some of the features generated are not necessary ones. Thus, there are three kinds of failure with regard to denotations: (1)

sometimes the lexical representations do not distinguish between what are obviously distinct denotations (e.g. *pizza*, *cake*); (2) sometimes they predict a narrower denotation than is in fact the case (e.g. *begin a book, enjoy a book*); and (3) sometimes they simply predict a wrong denotation (e.g. *begin a car*, *begin a thermometer*).

The problem of giving necessary and sufficient conditions for the application of a concept is well-known within philosophy and cognitive science, and Fodor (1970, 1981, 1998) and Fodor et al. (1980) have provided convincing arguments against the possibility of a definitional analysis of word meaning. Given this 'undefinability' of most words (including nouns), it seems likely that, in many cases, it will be difficult to give any precise values for e.g. the qualia roles (or for any of the other levels of linguistic representation proposed by Pustejovsky) associated with a lexical item. Try, for instance, to provide qualia structures for abstract nouns such as e.g. *semantics*, *freedom*, *justice*, etc.

3.3 World Knowledge and the Lexicon

Another problematic issue regarding the lexical representations assumed by the GLT is the way it presupposes world knowledge being built into the lexicon. This is one of Fodor and Lepore's (1998) main objections to the theory; that it lacks a distinction between what the language tells us and what we know about the world. Fodor and Lepore argue that even though it may seem at first glance that Pustejovsky's theory is able to account for the differences in interpretation of *bake* in *bake a cake* and *bake a potato* solely by distinguishing artefacts from natural kinds, this is clearly not sufficient. Although e.g. knives and trolley cars are artefacts, *bake a knife* and *bake a trolley car* resist a creative reading just as much as *bake a potato* does. They further claim that even *bake a cake* is ambiguous, since it is perfectly possible to bake a (pre-existing) cake by putting it in the oven and non-creatively baking it. Instead, they argue, the reason we have an intuition that *bake* is lexically ambiguous in the case of (20) above, is because the verb is genuinely polysemous between the two readings: The verb *bake* has a single lexical entry, but denotes two different processes (the creative and the change of state sense), and it is our knowledge about the world that tells us which reading of *bake* to pick out in the case of *bake a potato*, and that makes *bake a knife* sound funny. On their view, it is not the case that the lexicon encodes or generates this information.

Fodor and Lepore (1998: 281) further claim that Pustejovsky's account offers no convincing cases of the meaning of a governing expression being modulated by the lexical content of the expression that it governs. In the case of the verb *use*, they note that it is perfectly possible to use things that do not have uses (e.g. a rock to break a window, snow to make a snowball, etc.) and that this presents a serious

problem for Pustejovsky's theory, as there may, in fact, be many cases in which a verb makes a demand on an argument that the lexical entry for that argument does not satisfy (and we do not want to say that the telic role for e.g. *rock* includes the function *break*, i.e. that this function is part of the semantics of the noun). This problem also extends to cases of selective binding, where the lexical representation of a governing noun may, in many cases, not contain a specification of a telic role at all, in which case there would be no event description for the adjective to selectively modify. Szabó (2001: 132) points out that the meaning of the sentence *This is a good pebble* depends on the context of its use (the pebble could be good e.g. for playing marbles, breaking a window, etc.), and thus concludes that the meaning of a noun *N* in a phrase '*a* is a good *N*' is often not sufficient in order to decide in which way *a* is said to be good. Moreover, Fodor and Lepore argue that it is perfectly possible to understand a sentence such as e.g. *Sarah likes to use the subway* and to wonder at the same time what she likes using it for (she might like to use the subway for sightseeing, shelter, etc.). The inference from *likes to use the subway* to 'likes to travel on the subway' is made possible by what we know about subways but this information should not be regarded as contained as part of the lexical entry for *subway*. On their view, therefore, Pustejovsky's theory conflates linguistic knowledge and world knowledge, and furthermore offers no way of constraining the amount of world knowledge entering into the lexical representation of a given lexical item.

The same point can be made with regard to the analysis of *bake the pizza* in terms of co-composition: we certainly do not want to say that the lexical representation for *pizza* actually contains some kind of distinction between frozen and home-made pizzas as part of its qualia structure, which is one way in which Pustejovsky could account for the two interpretations of *bake the pizza* if he wants to maintain his analysis of *bake* as dependent on the semantics of the complement noun. However, this distinction is clearly part of our world knowledge, and contextual factors (including extra-linguistic ones) will indicate which interpretation to select or construct in each case.

4 A Relevance-Theoretic Approach

The problems discussed in the previous section illustrate the limitations of relying on linguistic context alone in the derivation of compositional interpretations. An important motivation for a re-analysis of Pustejovsky's examples of 'logical polysemy' in pragmatic terms is the fact that the GLT, although incorporating a considerable amount of world knowledge into the lexicon, still leaves a lot of work for pragmatics to do in finding the interpretation intended by a speaker. First of all, pragmatics is needed to correct the cases of wrong interpretations predicted by the

theory. Second, when there is more than one possible default interpretation associated with a qualia structure, it would be entirely a matter of pragmatics to choose which is the correct one (if any of them) in a given context. Žegarac (2006: 1707) argues that a pragmatic theory which is capable of doing the considerable amount of work that a GLT-based account of lexical meaning leaves for it to do can also do that part of the interpretive work that a GLT does adequately. I follow him in this and would add that a wholly pragmatic account is to be preferred, not only because it would give a simpler, more unified account overall (with a much leaner semantics), but also if it can be shown to avoid the problems of misinterpretation discussed in the previous section. I believe that relevance theory (Sperber & Wilson 1986/1995, Carston 2002) provides a pragmatic theory that does meet these requirements.

Relevance-theoretic pragmatics is based on a definition of relevance and two principles: the Cognitive and Communicative Principles of Relevance. Relevance is a potential property of any input to cognitive processes, and may be assessed in terms of processing effort and cognitive effects. An input is relevant to an individual when it yields a positive cognitive effect as a result of being processed in a context of available assumptions (Wilson & Sperber 2004). Positive cognitive effects include contextual implications (that is, conclusions that are deducible from input and context together), strengthening and/or elimination of existing assumptions. Other things being equal, the greater the cognitive effects of an input to an individual who processes it, and the smaller the processing effort required to derive these effects, the greater the relevance of that input to that individual at that time. The two principles, given in (25) and (26), are proposed on the basis of this definition:

(26) **The Cognitive Principle of Relevance**
Human cognition tends to be geared to the maximisation of relevance. (Sperber & Wilson 1995: 260)

(27) **The Communicative Principle of Relevance**
Every act of ostensive communication communicates a presumption of its own optimal relevance. (ibid.)

Following the definition of relevance, the Cognitive Principle of Relevance predicts that hearers will aim at deriving as many cognitive effects as possible for as little processing effort as possible, while it follows from the Communicative Principle of Relevance that the speaker, by the very act of addressing the hearer, is communicating that her utterance is the most relevant one she could have produced given her own preferences and abilities, and is at least relevant enough to be worth the hearer's processing effort (i.e. it is 'optimally' relevant). The hearer's task,

then, is to find an interpretation of the speaker's utterance that is consistent with this presumption, i.e. the first one he accesses which yields sufficient cognitive effects for no gratuitous processing effort. (For more detail, see Wilson & Sperber 2004.)

4.1 Linguistic Meaning and Explicature

The relevance-theoretic account assumes that communication is essentially an inferential process. Inferential communication is taken to involve the application of general-purpose inference rules, applying to any type of information which is represented conceptually (Sperber & Wilson 1986/1995: 176). On this view, linguistic communication is but one way in which communication may obtain, but it is special in the sense that it can achieve a degree of precision and complexity which is rarely achieved in non-linguistic communication. Verbal communication is taken to involve two kinds of communication/comprehension process: one is coding and decoding, the other ostension and (non-demonstrative) inference. The first one provides input to the pragmatic inferential process, which is seen as the main part of the comprehension process.

Now let us focus in a bit more detail on the relevance-theoretic notion of the input to the inferential process, that is, the view of linguistic meaning assumed by the theory. The general idea is that the linguistic meaning of an utterance greatly underdetermines its interpretation: it not only underdetermines the implicit meaning of the utterance (i.e. its implicature(s)), but also the proposition expressed by it, i.e. its *explicit* content (Carston 2002: 29). On this view, the semantic representation encoded by a sentence is taken to be an incomplete logical form, which is, at best, a fragmentary representation of a thought. In order to recover the complete (truth-evaluable) proposition the speaker has expressed from the linguistically decoded semantic representation, it must be inferentially enriched (by disambiguation, assignment of referents, free enrichment) in a context of available assumptions. The resulting proposition is an *explicature*, which is defined as follows (Carston 2002: 377):

> [An explicature is] an ostensively communicated assumption which is inferentially developed from one of the incomplete conceptual representations (logical forms) encoded by the utterance.

It is assumed that, at every stage of the process of developing the linguistic meaning into a complete proposition, the hearer will choose the solution involving the least processing effort, and this solution will be abandoned only if it does not yield an interpretation consistent with his expectations of relevance (Sperber & Wilson 1986/1995: 185). This view of linguistic meaning, on which semantic

representations are incomplete logical forms that must be contextually developed, departs radically from the one assumed by GLT, where a lot of the information which is taken to be pragmatically inferred on the relevance-theoretic account, is encoded in the lexicon.

4.2 A Re-analysis

4.2.1 Free Enrichment. Now let us return to our examples of 'logical polysemy'. I will leave aside the metonymy example in (1) for a moment, and focus on the one in (2) (repeated here as (28)):

(28) Mary *began* a book.

Recall that Pustejovsky analyses this as a case of type coercion, where the type of the complement is changed into an event consistent with the interpretations 'Mary began reading a book' and 'Mary began writing a book', as a result of the events of reading and writing being encoded as part of the qualia structure for *book*. On the relevance-theoretic account, of course, no such information is assumed to be encoded by this word, and the relevant event associated with *begin a book* will be supplied entirely on pragmatic grounds. This would count as a case of 'free' pragmatic enrichment (e.g. Carston 2002: 323), where a conceptual constituent which is not articulated in the linguistic form of the utterance is contextually derived, as a result of the hearer's trying to figure out which proposition the speaker has explicitly expressed.[4] Let us imagine a context where John, Mary and Sue have been given the job of cleaning their late grandmother's house before it is advertised for sale. The bookshelves haven't been dusted for years, and, because they are completely covered in dust, each book has to be dusted individually. After they have been cleaning for a while, John asks if they should all take a break and go for coffee, and Sue replies (29):

(29) Let's wait for a couple of minutes. Mary just began a book.

In this context, the interpretation according to which the explicit content of the last part of Sue's utterance is 'Mary just began [DUSTING] a book' would clearly be the most relevant one. It is the most accessible interpretation, that is, the one requiring

[4] However, if it turns out that the semantics of the verb *begin* comes with some kind of parameter indicating that its complement is an event or activity, then this will be a case of saturation (rather than free enrichment). Either way, though, the recovery of the 'reading', 'writing' or other specific book-involving activity is a matter for context and pragmatic inference.

the least processing effort on the part of the hearer, and it is also one that offers an adequate explanation for why Sue thinks they should wait before leaving for coffee (it functions as a premise for the conclusion given in the first part of Sue's utterance). John derives this interpretation according to the relevance-theoretic comprehension heuristic (Wilson & Sperber 2004), which is applied automatically to verbal input. According to this procedure, the hearer (1) takes the decoded linguistic meaning, follows a path of least effort in mutually adjusting explicit content, contextual assumptions and contextual implications, and (2) stops when the interpretation he arrives at satisfies his expectations of relevance. At no stage in the interpretation process, therefore, is it assumed that a 'default' interpretation (according to which Mary began e.g. reading a book) is computed and then cancelled by context.[5] It doesn't take much imagination to think of contexts in which the correct (intended and easily retrieved) interpretation is 'began binding a book', 'began mending a book', 'began ripping up a book', and so on.

However, there is not much doubt that Pustejovsky is right in assuming that certain interpretations come more readily to mind than others in a 'null' context. In the absence of real-life contextual constraints, the utterance in (29) would probably most often be interpreted as meaning that Mary began reading a book. Sperber and Wilson (1986/1995: 185) see this as a case of hearers favouring the least effort-consuming conceivable interpretation. A person reading a book may be regarded as a stereotypical event, which may be stored in encyclopaedic memory as a chunk and accessed as a single unit of information. Retrieving this information from encyclopaedic memory during the interpretation of (29) would require little processing effort, while deriving the interpretation according to which Mary began e.g. dusting a book would involve accessing several units of information and hence be more costly in processing terms (see discussion in Sperber & Wilson 1986/1995: 186). In this way, relevance theory provides a natural explanation for the fact that certain interpretations are often favoured over others, without being committed to the view that these interpretations are always computed first as a result of default inferences associated with lexical items.

4.2.2 Ad Hoc Concept Construction. In order to account for Pustejovsky's remaining cases of 'logical polysemy', we need to take a closer look at the relevance-theoretic account of concepts and lexical-pragmatic processes. Not surprisingly, relevance-theorists take a radically different stance on the question of

[5] The different predictions made by GLT and relevance theory with regard to the derivation of compositional interpretations in cases like (29) should be experimentally testable. Lexical priming experiments (cf. e.g. Rubio Fernández 2007) may provide evidence either for the view that a so-called default interpretation is accessed first and then cancelled by context (GLT), or for the view that only the relevant interpretation is derived (RT).

word meaning from the decompositional view assumed by the GLT, in that they do not believe that it is generally possible to provide definitional word meanings (i.e. components of meaning which are individually necessary and jointly sufficient to capture the encoded content of a word). On the relevance-theoretic account, therefore, most words encode simple, unanalysable, atomic concepts in the sense of Fodor (1981, 1998). The word cat, for instance, encodes the atomic concept CAT. Such atomic concepts are represented in our 'language of thought', and consist of an address, or entry, in memory that may give access to three kinds of information: (1) the logical properties of the concept (e.g. a one-way meaning postulate, such as CAT ⇒ ANIMAL OF A CERTAIN KIND);[6] (2) a set of assumptions, or encyclopaedic information, about the denotation of the concept (e.g. 'cats have four legs', 'cats purr', various kinds of culture-specific information related to the concept, etc.), and (3) the lexical (i.e. phonological and syntactic) information related to the linguistic form that encodes the concept (Sperber & Wilson 1995: 85-93). On this view, lexical interpretation typically involves constructing an ad hoc concept, or an occasion-specific sense, which may be narrower (e.g. *bird* used to express 'bird of a specific sort, e.g. woodland bird') or broader (e.g. *wizard* used metaphorically to express 'a person who is very skilled in a particular field or activity') than the linguistically-specified denotation. The hearer takes the encoded concept and its associated logical and encyclopaedic entries as input, and uses the relevance-theoretic comprehension heuristic to derive warranted conclusions about the speaker's meaning (Wilson 2003, Wilson & Carston 2007), a process which includes a mutual adjustment of explicit content, contextual assumptions and contextual implications, guided and constrained by expectations of relevance. In many cases, an ad hoc concept is constructed as a result of a process of 'backward inference', where there is an inference from expected contextual implications to the explicitly expressed proposition containing the ad hoc concept (For more detail, see Carston & Powell 2006, Wilson & Sperber 2002, Wilson & Carston 2006). Concept narrowing and broadening are thus seen as the outcome of a single interpretive process, which adjusts the interpretation of individual words (Carston 1997, 2002, Wilson & Carston 2007).

Now let us consider again the example involving the verb *bake*, and see how the difference in meaning between *bake a cake* and *bake a potato* can be analysed in relevance-theoretic terms. The example is repeated below as (30):

---

[6] In espousing logical entries for concepts, relevance theory stands with Fodor in his earlier manifestation when he supported atomism (no definitions) together with meaning postulates (e.g. Fodor et al. 1980, Fodor 1981), rather than with his more recent position which eschews any logical/ encyclopaedic distinction (Fodor 1998).

(30) a. John *baked* a potato.
b. John *baked* a cake.

As we have seen, Pustejovsky analyses the verb *bake* as having a single change of state denotation, which is changed into the creation sense as a result of the process of co-composition, whose application is dependent on the semantics of the complement noun (e.g. natural kind/artefact). Etymologically, the basic sense of *bake* is assumed to be 'cook by dry heat',[7] and I will therefore follow Pustejovsky in considering the change of state denotation to be the one that is indeed linguistically specified by this word. However, in what follows I will try to show that, contrary to Pustejovsky's view, the change in meaning of the verb in (30b) is a pragmatic matter rather than a lexical semantic operation internal to the language system.

If we assume that the linguistically-specified denotation of the concept BAKE is the change of state meaning, then the concept communicated by use of *bake* in (30b) may be analysed as a case of pragmatic narrowing, since 'creation' in fact entails a 'change of state' (but not vice versa). In this case, the hearer will construct the ad hoc concept BAKE*, which will have a more restricted denotation than the encoded concept, i.e. picking out a subset consisting of those instances of the process of cooking by dry heat which bring something into existence (cake, pizza, bread, etc). The hearer arrives at this interpretation as a result of applying the relevance-theoretic comprehension heuristic. The decoded meaning of the sentence in (30b) will contain the concepts BAKE and CAKE, both of which will give access to a range of encyclopaedic information associated with their denotations. The encyclopaedic entry for CAKE will contain assumptions about the way in which cakes come into being, i.e. as a result of a 'creative' process of baking. The interpretation according to which BAKE communicates the creation sense BAKE*, would in this case be the one that is the most accessible to the hearer. Similarly, in (30a), where *bake* is used to communicate its encoded concept (BAKE), an important step in the hearer's accepting the encoded change of state meaning will be his accessing of the encyclopaedic entry for the concept POTATO, where he would find information to warrant this interpretation of *bake*.[8]

Some may argue that the difference between this analysis of *bake* and the one given by Pustejovsky is really just a matter of whether one takes the information

---

[7] *The Oxford Dictionary of English Etymology*. 1966/1985. Oxford: Clarendon Press.

[8] It would be interesting to do more research on the etymology of the verb *bake* in Germanic languages: in a language like Norwegian, for instance, the creative sense of the word is more likely to be the basic one, while the change of state sense is only accessible when the verb is combined with specific complements (such as e.g. *potato*). Such cases would induce a broadening of the linguistically-specified denotation of *bake*.

associated with the complement nouns to be linguistically encoded or part of the encyclopaedic information associated with the encoded concept. However, the pragmatic analysis has a great advantage over the semantic one, in that it allows for the necessary flexibility in lexical interpretation (in a given context, *bake* in *bake a cake* could just as well be used to convey the change of state meaning), while on the semantic approach one is committed to assuming that there is a single interpretation for *bake a cake*, which can be overridden only in the case of explicit contextual evidence pointing to a different interpretation.

However, a possible objection to the analysis of the creative sense of *bake* as an instance of on-line ad hoc concept construction could be that this sense may in fact be lexicalised. On this view, *bake* would be a case of genuine polysemy, as argued by Fodor and Lepore (1998), and there would be a single lexical entry for *bake*, with two senses, each denoting different but semantically related processes. This may well be so (at least for some people), in which case the pragmatic analysis just given can be seen as providing an account of the semantic relation between these two senses: assuming that the 'change of state' sense is the basic one, the 'creation' sense may have become lexicalised as a result of frequent narrowing. (For further discussion of the polysemy issue, see Wilson & Carston 2006.)

Turning to Pustejovsky's examples of selective binding, these can be straightforwardly analysed as cases of lexical narrowing on the relevance-theoretic account. Consider again the case of *good*, repeated below as (31):

(31) a. Every chef needs a *good* knife.
b. 'The Kite Runner' is a *good* book.

This adjective may be seen as encoding a very broad concept, which, on most occasions of use, will have to be pragmatically adjusted in order to arrive at the interpretation that was intended by a speaker. In (31a), the concept communicated by *good* would be the ad hoc concept GOOD*, conveying 'good for cutting', which would have a much narrower denotation than the one linguistically-encoded by the word. In (31b), a different ad hoc concept, GOOD**, would be communicated by this word, conveying that 'The Kite Runner' is a 'good read'.

There are, however, many other ways in which a book can be good. A *good book* can be one that is easily read, one that is entertaining, one of a high academic or intellectual quality, one that is good to kill flies with, etc. Similarly, it is not difficult to imagine contexts in which *good knife* would be given entirely different interpretations than the one in (31a). Consider (32)-(33):

(32) To become a member of Billy's exclusive gang you had to have a *good* knife.[9]
(33) This is a *good* knife for people with wrist pain.

In (32), the communicated meaning of *good* could be 'good for stabbing people' or 'good for threatening people with', while in (33), it could be 'good for holding'. Again, the relevance-theoretic approach predicts that *good* may communicate different occasion-specific senses in these cases, by assuming that lexical interpretation is a matter of adjusting the interpretation of individual words in accordance with one's context-specific expectations of relevance. The variety of possible interpretations of *good* in the above examples also strongly suggests that Pustejovsky's analysis of the meaning of *good* as dependent on the function of the object it modifies is inadequate.[10]

4.2.3 'The Problem of Metonymy'. Finally, let us consider our metonymy example in (1), repeated below as (34). The word newspaper has different denotations in (34a-c):

(34) a. *The newspaper* fired its editor. (organisation)
b. Mary spilled coffee on *the newspaper*. (physical object)
c. *The newspaper* revealed surprising details about the trial. (information type)

On Pustejovsky's account, the different senses of *newspaper* were taken to be encoded directly into its lexical representation as a lexical conceptual paradigm, an analysis which involved the generation of a sense for *newspaper* for which there was no corresponding denotation (as *newspaper* cannot simultaneously denote an organisation, an information type and a physical object), a clearly unsatisfactory outcome. However, as Wilson and Carston (2007: 253) concede, the question of how to analyse cases of metonymy within the relevance-theoretic framework (in terms of ad hoc concept construction) is not straightforwardly answerable.[11] They suggest, following Nunberg's (1996, 2004) account, that at least some cases of metonymy could be analysed as involving lexical broadening. Nunberg argues that

---

[9] I owe this example to Robyn Carston.

[10] It is worth mentioning that the flexibility lacking in Pustejovsky's account is allowed for in an approach like Szabó's (2001), which postulates a variable in the semantic representation of *good*, inducing a pragmatic operation of saturation.

[11] Papafragou (1996) gives a preliminary relevance-theoretic account of metonymy, where metonymy is seen as a variety of interpretive use of language as defined by Sperber and Wilson (1986/1995: 228-229).

in many metonymical utterances (such as e.g. an utterance of *I'm parked out back*) the meaning transfer applies to the conventional meaning of the predicate (i.e. to *parked out back* and not to the subject *I*), and such cases can be analysed in terms of lexical broadening on the relevance-theoretic account. However, in (34) above, it is clearly the noun *newspaper* to which the alternation in meaning applies, and this case does not seem straightforwardly analysable in terms of lexical narrowing or broadening. Wilson and Carston (2007: 253-254) note that the notion of metonymy is intuitively harder to grasp than that of e.g. metaphor, and there is not always agreement on which cases are instances of metonymy and which are not. I believe it is a genuine question whether the different senses of *newspaper* should be regarded as semantic (that is, as encoded in the lexicon), in which case the noun could be regarded as being polysemous in the traditional sense (and associated with a set of conventionalised metonymic senses), or pragmatic, in which case one of the senses should be taken to be the encoded one and the other senses pragmatically derived. In such a case, however, it is not obvious which of the meanings in (34a-c) should be taken to be the basic one, although the physical object sense in (34b) would be a likely candidate. (For further discussion of the semantics/pragmatics issue, see Nunberg 2004: 351-354).[12] I will not choose between these possible analyses here, but will simply acknowledge that this is an issue which remains to be adequately explained within the relevance-theoretic framework.

5. Conclusion

In this paper I have discussed Pustejovsky's generative theory of word meaning and compositional interpretation, and its treatment of cases of 'logical polysemy'. I hope to have shown that the range of phenomena falling under this label in Pustejovsky's theory are better given a pragmatic account, and that the data can be reanalysed in terms of the independently-motivated pragmatic processes of free enrichment and ad hoc concept construction within the relevance-theoretic framework. This analysis allows for, in fact predicts, the flexibility that we find in the derivation of compositional interpretations.

[12] However, whether the metonymy of *newspaper* is regarded as a semantic or as a pragmatic issue, an adequate account of metonymy should have something to say about the distinction between clearly pragmatic cases, such as e.g. *The ham sandwich left without paying*, and other more predictable/systematic cases, such as e.g. *newspaper, book*, etc.

References

Atkins, B. T. & Kegl, J. & Levin, B. 1988. Anatomy of a verb entry: from linguistic theory to lexicographic practice. *International Journal of Lexicography* 1 (2), 84-126.

Blutner, R. 2002. Lexical semantics and pragmatics. *Linguistische Berichte* 10, 27-58.

Carston, R. 1997. Enrichment and loosening: complementary processes in deriving the proposition expressed? *Linguistische Berichte* 8, 103-127.

Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. & Powell, G. 2006. Relevance theory – new directions and developments. In *The Oxford Handbook of Philosophy of Language*, Lepore, E. & Smith, B. (eds.), 341-360. Oxford: Oxford University Press.

Copestake, A. & Briscoe, T. 1996. Semi-productive polysemy and sense extension. In *Lexical Semantics: The Problem of Polysemy*. Pustejovsky, J. & Boguraev, B. (eds.), 15-67. Oxford: Clarendon Press.

Fodor, J. A. 1970. Three reasons for not deriving "kill" from "cause to die". *Linguistic Inquiry* 1, 429-438.

Fodor, J. A. 1981. The present status of the innateness controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science*, 257-316. Brighton, Sussex: Harvester Press.

Fodor, J. A. 1998. *Concepts: Where Cognitive Science Went Wrong.* Oxford: Clarendon Press.

Fodor, J. A., Garrett, M. F., Walker, E. C. T. & Parkes, C. H. 1980. Against definitions. *Cognition* 8, 263-367.

Fodor J. A. & Lepore, E. 1998. The emptiness of the lexicon: reflections on James Pustejovsky's 'The generative lexicon' *Linguistic Inquiry* 29, 269-288.

Langacker, R. 1984. Active zones. *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 172-188.

Nunberg, G. 1996. Transfers of meaning. In *Lexical Semantics: The Problem of Polysemy*. Pustejovsky, J. & Boguraev, B. (eds.), 109-132. Oxford: Clarendon Press.

Nunberg, G. 2004. The pragmatics of deferred interpretation. In *The Handbook of Pragmatics*, Horn, L. & Ward, G. (eds.), 344-364. Oxford: Blackwell.

Papafragou, A. 1996. On metonymy. *Lingua* 99, 169-195.

Pustejovsky, J. 1991. The generative lexicon. *Computational Linguistics* 17 (4), 409-441.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, Massachusetts: The MIT Press.

Pustejovsky, J. 1998a. Generativity and explanation in semantics. Reply to Fodor and Lepore. *Linguistic Inquiry* 29, 289-311.

Pustejovsky, J. 1998b. The semantics of lexical underspecification. *Folia Linguistica* 32 (3-4), 323-347.

Recanati, F. 2002. Unarticulated constituents. *Linguistics and Philosophy* 25, 299-345.

Rubio Fernández, P. June, 16 2007. Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *Journal of Semantics* doi:10.1093/jos/ffm006.

Sperber, D. & Wilson, D. 1995 [1986]. *Relevance: Communication and Cognition*. Oxford: Blackwell. [2nd edition].

Stanley, J. 2000. Context and logical form. *Linguistics and Philosophy* 23, 391-434.

Szabó, Z. G. 2001. Adjectives in context. In *Perspectives on Semantics, Pragmatics and Discourse: A Festschrift for Ferenc Kiefer*, Kenesei, I. & Harnish, R. M. (eds.), 119-146. Amsterdam: John Benjamins.

Wilson, D. 2003. Relevance theory and lexical pragmatics. *Italia Journal of Linguistics/Rivista di Linguistica* 15 (2): 273-291. [Special Issue on Pragmatics and the Lexicon].

Wilson, D. & Carston, R. 2006. Metaphor, relevance and the 'emergent property' issue. *Mind & Language* 21 (3), 404-433.

Wilson, D. & Carston, R. 2007. A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts. In *Pragmatics*, Burton-Roberts, N. (ed), 230-259. Palgrave.

Wilson, D. & Sperber, D. 2002. Truthfulness and relevance. *Mind* 111, 538-632.

Wilson, D. & Sperber, D. 2004. Relevance theory. In *The Handbook of Pragmatics*, Horn, L. R. & Ward, G. (eds.), 607-632. Oxford: Blackwell.

Žegarac, V. 2006. 'Believing in': a pragmatic account. *Lingua* 116 (10), 1703-1721. [*Language in Mind: A Tribute to Neil Smith on the Occasion of his Retirement*].

Subsentential utterances, ellipsis, and pragmatic enrichment[*]

ALISON HALL

Abstract

It is argued that genuinely subsentential phrases can be used to perform speech acts with truth conditions. Attempts to assimilate this phenomenon to syntactic ellipsis (sluicing, gapping, etc.) are discussed, and are rejected on the grounds that any implementation of this idea will involve a redundant level of representation in natural language that plays no role in the interpretation process, and therefore be less economical than a pragmatic enrichment account. An argument against the latter kind of approach from the indeterminacy of content is discussed, then it is shown how a pragmatic account can accommodate this indeterminacy and turn it into an advantage through consideration of the role of processing effort in inferential comprehension.

1 Introduction

1.1 Background: The determinants of truth-conditional content

This paper addresses the status of what appear to be discourse-initial subsentential phrases, such as "The second man from the left" to pick someone out of a line-up, or "From Greece" to indicate the provenance of an item. The aim is to show that these are genuine cases of just words or (subsentential) phrases used to perform speech acts, and that they therefore provide evidence for the reality of the disputed pragmatic process of free enrichment.

The wider issue forming the background to this discussion of subsentential speech is the ongoing debate about whether all constituents of the truth-conditional content of an utterance can be traced to the encoded linguistic meaning (logical form). There are two broadly opposing views on this issue. One view, the more semantically oriented, represented by Stanley (2000, 2002, 2005a); King and Stanley (2005); Stanley and Szabo (2000); Taylor (2001), among others, is that all determinants of truth conditions are indeed traceable to logical form (or parameters of the lexical semantics). Apart from disambiguation, pragmatic contributions to truth-conditional content are limited to saturation – assigning values where the linguistic form calls for them. These authors take the object of semantic theory to be the intuitive truth-conditional content of an utterance (Stanley and Szabo 2000:

[*] Many thanks to Robyn Carston for many helpful discussions of subsententials and comments on earlier drafts of the paper. This work is supported by an AHRC doctoral award.

240; King and Stanley 2005: 141), as opposed to a minimal semantic content more or less isomorphic with overt (i.e. pronounced) elements. This means that, to account for the pragmatic effects of quantifier domain restriction and other kinds of nominal restriction, adjectival modification, provision of a location value with weather verbs, and so on, they are forced to posit either extensive hidden structure in the logical forms of sentences – syntactic entities such as covert variables attached to the relevant overt lexical item – or parameters in the semantics of the expressions. The alternative approach, defended by Recanati (2002); Carston (2002, 2004), etc., is that of 'contextualism', which maintains that pragmatics can have a far more pervasive effect on truth conditions than merely supplying values which are linguistically mandated. According to the contextualist view, at least some of the above kinds of pragmatic effects, and many others, are not cases of saturation; instead, a pragmatically motivated process of free enrichment – 'free' from linguistic control but, obviously, tightly constrained by pragmatic considerations – provides 'unarticulated constituents' of truth-conditional content (where 'unarticulated' means not just unpronounced, but not articulated at any level of linguistic representation – i.e. not traceable to a covert variable or parameter). Free enrichment occurs on pragmatic grounds, where the result of decoding, disambiguation, and saturation would not be a proposition that the speaker intends to express.

Advocates of the semantic approach claim that it enables a clear, systematic account of how hearers grasp the truth conditions of utterances, in contrast to the seemingly unconstrained and unpredictive pragmatic enrichment account. Its detractors point to the implausibility of the proliferation of hidden linguistic structure or parameters required by the syntactic/semantic account, and are optimistic that further investigation will show that the pragmatic mechanisms that are independently necessary for (particularized) conversational implicature also operate to constrain optional pragmatic contributions to truth-conditional content.

1.2 'Subsentential' utterances and the free enrichment debate

We can utter what appear to be isolated noun phrases, prepositional phrases, etc. such as those in (1)a-(7)a – mostly familiar from papers by Stainton (e.g. 2004), Stanley (2000), and Botterell (2005) – and thereby communicate propositional contents; some possibilities are given in (1)b-(7)b:

(1) a. [Uttered while indicating a man across the room]
John's father.
b. THAT IS JOHN'S FATHER. [1]

(2) a. [Pointing at a restaurant table]
Reserved.
b. THAT TABLE IS RESERVED.

(3) a. The editor.
b. THAT WOMAN IS THE EDITOR OF MODE MAGAZINE.

(4) a. [Uttered by someone who has been robbed and is trying to pick the robber out of a police line-up]
The second man from the right.
b. THE SECOND MAN FROM THE RIGHT ROBBED ME.

(5) a. [Uttered by a used-car salesman]
Only 10,000 miles. Like new.
b. THIS CAR HAS ONLY BEEN DRIVEN 10,000 MILES. IT'S LIKE NEW.

(6) a. Nice dress.
b. YOU ARE WEARING A NICE DRESS.

(7) a. Typical.
b. THAT BEHAVIOUR IS TYPICAL OF HIM.

The consensus is that the propositions in (b) are the propositions expressed by (truth conditional contents of) utterances of the (a) expressions, rather than mere implicatures. As Stainton (e.g. 2004) has pointed out, one could clearly lie with (1)-(5), for example, which shouldn't be possible if all they communicate is implicatures, and these propositions are the starting-points for inference to implicatures – (6) can be used ironically, for instance; (7) could implicate disapproval.

Since (1)-(7) can be uttered without prior discourse, they do not immediately look like varieties of syntactic ellipsis. (So as to not prejudge the issue, I will label them 'fragments' for now, rather than subsententials). The received opinion on ellipsis is that it is a grammatical operation, with the unpronounced material being reconstructed algorithmically, internal to the language faculty; it is not inferred according to principles of pragmatics or general reasoning. Hence the need for an overt linguistic antecedent to serve as a licenser for ellipsis, as found with recognized forms of ellipsis such as gapping, VP-ellipsis, and sluicing. So (1)-(7) do not (at least prima facie) seem to be syntactically elliptical sentences. If appearances reflect reality, then all that (1)-(7) encode are bare NPs/DPs, and so on,

[1] These are rough indications of the propositional contents, which will, of course, not contain names, pronouns, demonstratives, or referentially used definite descriptions, needing reference assignment; the references will be fixed, and the natural-language indexicals etc. in (1b)-(7b) just represent the concepts of objects and properties that are part of the propositions.

with the semantics of phrases, yet their semantics can be combined with non-linguistic information to express propositions. It follows that the propositions expressed must contain unarticulated constituents, and such utterances would, then, provide obvious and extensive evidence of the reality of free enrichment. To maintain, as the semanticist wants to, that the linguistic form and the truth conditions of an utterance are structurally isomorphic requires denying that there are any genuine cases of subsentences being used to perform speech acts (or at least to express truth conditions). Since it is agreed that utterances of (1)-(7) and many examples like them do have truth conditions, the semanticist must show that they are structurally complete sentences, with some kind of ellipsis having applied.

In the next section, I discuss the recent defence of this 'sententialist' approach by Stanley (2000), and then some variations on it by Merchant (2004; 2006), and Ludlow (2005), who all propose to account for most of the data in question by treating it as syntactic ellipsis. This is seen as having the advantages of preserving the standard, systematic mapping between logical form and truth conditions, and of being more parsimonious than a pragmatic enrichment account by virtue of making use of the independently motivated constraints and operations governing the production/interpretation of more familiar varieties of ellipsis, therefore avoiding the need to introduce extra, allegedly ad hoc, pragmatic machinery. In response, Stainton (2006a; 2006b), Barton (2006), and others have argued that the interpretation of discourse-initial fragments (or at least most of them) is unlike that of recognized kinds of ellipsis. Building on their arguments, I develop the general 'subsententialist' case further by examining how, on these ellipsis proposals, the comprehension systems integrate the contextual information required to go from the pronounced fragment to a full-fledged proposition. I argue that this process cannot take place entirely in natural language, but must at least partly involve representations in some other medium – Language of Thought, or Mentalese representations. Any attempt to rescue the ellipsis story for these discourse-initial cases requires that the full sentence/proposition be represented in natural language at some stage of processing, and I aim to show that all such attempts suffer from problems of redundancy by necessitating superfluous levels of representation that play no role in the interpretation process.

Having argued that data such as (1)-(7) really are subsentential phrases, in section 3 I turn to a more positive defence of the contextualist approach which sees the interpretation of these utterances as involving pragmatic enrichment. It has been suggested (Stanley 2000) that the use of a truly (i.e. non-elliptical) subsentential phrase does not count as performing a genuine linguistic speech act, because the utterance lacks determinate content and/or illocutionary force. In response, first I argue, following Stainton (2006b etc) and Clapp (2005), that this requirement is unrealistically strict, and would wrongly entail that much of speech, including many fully sentential, grammatical utterances, would not count as linguistic speech

acts. Second, I sketch an account of how this indeterminacy is an advantageous feature of subsentential speech, once proper consideration is given to the role of processing effort in inferential comprehension.

2 Ellipsis

2.1 Fragments and varieties of ellipsis

A number of semanticists who want to preserve the traditional equation of linguistic form and truth conditions have claimed that utterances such as (1)-(7), when used to perform speech acts, are actually structurally complete sentences to which syntactic ellipsis has applied, leaving just the pronounced fragment. This section discusses some recent defences of this sententialist approach (Stanley 2000; Merchant 2004, 2006; Ludlow 2005) and the response from those who believe there are genuine subsentential speech acts (e.g. Stainton 2006a,b).

The sententialist claims that apparent subsententials such as (1)-(7) are in fact elliptical sentences, hence similar to phenomena such as direct, immediate answers to questions (8), sluicing (9), gapping (10), and VP-ellipsis (11):

(8) A: Where are you from?
B: Italy.
(9) He left. I wonder why.
(10) Linda speaks French and Jane German.
(11) Linda speaks French but Jane doesn't.

(8)-(11) are widely accepted as cases of syntactic ellipsis: the apparent fragment is syntactically a full sentence, and the elided material can be recovered algorithmically by processes internal to the language faculty, essentially just copying material into the ellipsis site that is syntactically identical to material present in the antecedent (subject to syntactic locality constraints).

A feature of accepted types of ellipsis is that they can't occur discourse-initially, or in isolation: they need explicit linguistic antecedents as licensers. Gapping, for example, while often easily comprehensible, is ungrammatical in isolation or without the right type of linguistic antecedent:

(12) A: Does anyone speak French or German?
B: *Yes, Jane German.

(1)-(7) don't have explicit antecedents, so don't look to be the same phenomenon. They are perfectly natural when uttered discourse-initially: The utterance of "John's father" doesn't require an explicit prior question; all that is needed is that

an appropriate person be salient of whom can be predicated the property of being John's father.

However, Stanley (2000: 401-9) argues that, despite appearances, many such fragments, including (1)-(5) in section 1.2, are only discourse-initial in an implausibly wide sense of discourse-initial. They don't have explicit linguistic antecedents but, according to Stanley, it would be a mistake to conclude that they don't have linguistic antecedents at all: explicitly mentioning a linguistic antecedent, he says, is not the only method of raising linguistic expressions to salience. The felicitous use of these apparently discourse-initial utterances depends on something in the non-linguistic context raising to salience some linguistic expression to serve as the antecedent for the ellipsis: for example, they would generally be infelicitous without a preceding ostensive stimulus (such as a demonstration) to draw attention to some object, person or situation, as in (1). They therefore aren't discourse-initial in any relevant sense, claims Stanley, because the prior context necessary for such utterances to be acceptable will have made linguistic antecedents salient to serve as licensers for ellipsis. For (1), the implicit question "Who is that man?" is salient, so this and the utterance "John's father" function similarly to the overt question-answer pair in (8) above: The implicit question makes the linguistic expression "That man is…" available as the restoration of the deleted material in the 'reply'.

If this is correct, it removes an objection to treating (1) as syntactic ellipsis. However, it would be premature to conclude that what is going on here really is ellipsis. From the fact that an utterance requires prior linguistic context (in Stanley's extended sense of what can constitute linguistic context), it does not follow that the utterance is elliptical, because needing prior linguistic context is not exclusive to elliptical sentences. The felicitous use of many overtly fully sentential utterances also would depend on this kind of context being available, for example to complete the definite description in (13) (assume it is attributively used)[2] to give THE AUTHOR OF THAT BOOK:

(13) [Speaker points at a book]
The author's going to be signing copies later.

[2] Judging from what is said about context in various places in Stanley's work, it seems that he is drawing a distinction between two roles of context in interpreting these fragments. On the one hand, there is the role of context in making linguistic expressions salient. This is what Stanley must have in mind in his discussion of fragments, with non-linguistic context making linguistic expressions salient to serve as antecedents for ellipsis. On the other hand, since such natural-language expressions cannot serve as the content where singular reference is involved, context has a different role – that of saturation (assigning referents to the natural-language indexicals which have been made salient by the non-linguistic context). King and Stanley (2005: 130) mention this second role of context in connection with the referential use of definite descriptions.

Regarding example (1), the only kind of situation that Stanley could treat as genuinely discourse-initial would be where no attention is established on a referent, and there is no demonstration of the entity or property being referred to, in order to make a linguistic indexical salient, but in such a situation, an utterance of (13) would be just as infelicitous as (1). So (1) needing prior context is no argument for it being a case of ellipsis.

This argument can be pushed further by considering (6)a ("Nice dress"). Stanley's original discussion of this example was as follows:

> In this case, it is fairly clear that an assertion has been made, whose content is a singular proposition about the object in question, to the effect that it is a nice dress. However, it is intuitively plausible to suppose, in this case, that the speaker simply intended her utterance to be shorthand for "that is a nice dress". (Stanley 2000: 409)

This 'shorthand' proposal is not developed any further by Stanley, but Elugardo and Stainton (2004: 448-54) consider several different conceptions of what the interpretation of 'shorthand' would involve. To qualify as a genuine alternative to free enrichment, shorthand must involve some kind of encoded link between a word/phrase and a propositional meaning; however, this leads to an implausible multiplication of linguistic elements: either multiple lexical ambiguity, or multiple conventions of use associated to an expression. (See Elugardo and Stainton's paper for the details of the argument.)

The shorthand strategy, then, looks to be a non-starter, and Stanley himself would, reportedly, no longer appeal to it[3]. However, this means that there are cases which, as he acknowledges, "can occur discourse initially, are clearly uttered with assertoric force, and have determinate unique propositional contents" (Stanley, ibid). As I will discuss later, there seems, pace Stanley, that there is some indeterminacy about the propositional content (and possibly also about the illocutionary force) of (6)a, in which case it may be open to the sententialist to deny that it constitutes a proper speech act – a strategy which is considered in section 3). However, my guess is that the sententialist would treat such cases as ellipsis, since I think most people would agree that (6)a is used in this context to express truth conditions (there is a strong intuition that if the speaker in fact believes the dress to be hideous, her utterance of (6)a is false).

The problem for the sententialist, now that the examples previously dismissed as shorthand will have to be treated as ellipsis, is that the notions of 'discourse-initial' and 'prior context' will have to be restricted and broadened, respectively, even further than they were in Stanley's (2000) discussion of ellipsis. Since an utterance

[3] Stanley, personal communication to Stainton, cited in Stainton (2006b: 147).

of "Nice dress" can be used without the speaker and hearer having previously attended to the dress or been involved in any form of discourse, the 'not discourse-initial' requirement that is satisfied by more familiar examples of ellipsis will have to be weakened, and what we end up with is that an utterance can count as non-discourse-initial, and therefore be a genuine case of syntactic ellipsis, as long as the context makes salient enough any descriptive material required to interpret the utterance. This further modification, though, collapses any distinction with sentential speech: If the required context is inaccessible, because no appropriate linguistic expression is salient, then many fully sentential utterances (e.g. probably just about all sentences containing a non-referentially used quantifier) will be uninterpretable (see footnote 2 on the two roles of context Stanley is assuming). Thus the fact that a given subsentential utterance needs some 'linguistic context' to license it provides no way of choosing between this account and the pragmatic enrichment one. The choice between the two approaches will have to be made on other grounds, and, in the next subsection, I argue in more detail against some specific ellipsis accounts and in favour of pragmatic enrichment.

A further reason why, from the fact that an apparently subsentential utterance is felicitous and interpretable, it does not automatically follow that it is elliptical (or, for that matter, shorthand) is that to draw such a conclusion would be simply to stipulate that ellipsis (or shorthand) is present, and would be ruling out free enrichment without any argument. Given that Stanley does not exclude the possibility of the existence of free enrichment (contingent on a satisfactory future account of it), that conclusion would need arguing for. There is no evidence that this is ellipsis, other than that no other option is left after dismissing the possibility that it's free enrichment. Yet as Stanley (2005b) says, the semanticist position (that there are no strong pragmatic effects on truth conditions) is an empirical hypothesis, in advance of detailed inquiry. Since one would expect an empirical hypothesis to be falsifiable, data that threatens to falsify it cannot be dismissed on the grounds that to do so is necessary to maintain the truth of the hypothesis. There must be some independent evidence to support the claim that what we see here is ellipsis, and, as I discuss in the rest of this section, such evidence is completely lacking.

The main difficulty in treating discourse-initial[4] phrases like (1)-(7) as elliptical is that, if this is ellipsis, then it is not much like any of the familiar varieties, and its dissimilarity reinforces the impression that the only reason to try and assimilate it to more usual kinds of ellipsis is to save the semanticist (sententialist) position. The generally accepted view of ellipsis is that the exact natural-language source must be

[4] Given the discussion of different conceptions of 'discourse-initial' on the previous couple of pages, I should make it clear at this point that, from now on, I will be using 'discourse-initial' as shorthand for 'occurring without any explicit linguistic antecedent', and not in Stanley's more restrictive sense of occurring without any context which could raise linguistic antecedents to salience.

identifiable: as Barton (2006) puts it, "An ellipsis account presumes a full sentential source from which syntactically and semantically identifiable material has been deleted". Identification of the unique deleted material is possible with sluicing, gapping, VP-ellipsis, pseudogapping, and question-answer pairs, as is to be expected if these constructions are just ordinary natural-language sentences, some elements of which go unpronounced, and it follows from the fact that ellipsis does not occur discourse-initially: it is the explicit antecedent that provides the linguistic expression to complete the fragment. But, as has been repeatedly pointed out (Stainton 2006b; Barton 2006, etc.), in the discourse-initial cases, we can't identify the allegedly deleted linguistic material. For instance, candidates for the unpronounced elements in (1) might include "He is", "That is", "That man is", "The man who just came in is", "has just come in", "The man in the pink tie is", "is the man near the door", "is over there", and so on. That we can't say what sentence was allegedly uttered, or what sentence the hearer recovered, suggests that there wasn't a sentence uttered: the rest of the proposition that we understand is not from a linguistic source. But for the syntactic ellipsis account to hold, there needs to have been deletion of a linguistic expression: if not, then part of truth-conditional content is not traceable to the logical form of what was uttered, so has been supplied through free enrichment.

2.2 Ellipsis accounts of discourse-initial fragments

Having established that discourse-initial fragments are not very naturally assimilated to syntactic ellipsis, in this subsection I consider the various proposals given by Stanley (2000), Merchant (2004) and Ludlow (2005), in order to examine in more detail what an ellipsis account would involve and show why a pragmatic account is needed instead.

As argued in section 2.1, since there is no linguistic antecedent to license ellipsis, and therefore no unique identifiable linguistic material that is recoverable, it is problematic to analyse the discourse-initial fragments in (1)-(7) as varieties of syntactic ellipsis. The sententialist may reply that recovery of the exact intended linguistic material is not essential: discourse-initial fragments were not previously considered, and maybe recoverability is not a necessary feature of ellipsis, but just something that happens to be displayed by most forms of ellipsis. An essential feature (and, it could be argued, the defining feature) of syntactic ellipsis is that reconstruction of the missing material is an algorithmic, dumb process carried out entirely internal to the language faculty. If, as Stanley (2000) claims, context makes available a linguistic expression to serve as antecedent for the ellipsis, then it would be expected that the linguistic expression could vary between hearers, especially where the role of the allegedly elided expression is to refer to an object or person – as is often the case with the fragments under discussion – since any number of

descriptions or indexicals might serve equally well. Whether this is plausible depends on what the correct wider conception is of the relationship between language and thought representations in the interpretation of these fragments, which is the issue that I turn to now.

The two general approaches to the analysis of these fragments line up (approximately) with different views on what the medium of thought is. A commonly held conception, defended by Fodor (1975), and adopted by the advocates of a pragmatic enrichment analysis, is that thought takes place in a distinct medium from natural language, and that the integration of material from different sources also takes place in this distinct Language of Thought (LoT). The other view, which seems more compatible with the ellipsis account, is that thought takes place (mostly) in natural language. In what follows, first I explain the LoT idea in a little more detail and discuss the competing approaches to fragments with regard to this view. I then discuss varieties of what is to some extent a 'thinking in natural language' view, and try to show that on no conception of the relation between language and thought in the interpretation of these fragments is the ellipsis account really plausible.

The view of mental architecture and processing shared by those defending a subsententialist account is along the general lines suggested by Fodor (1983). The mind is divided into a central system (or systems; questions of its/their internal architecture can be set aside here), and various peripheral input/output systems. The central system deals with representations in a conceptual format (a Language of Thought), and is where inferential processes (including pragmatic processing) take place. The peripheral systems are the perceptual (input) systems and a linguistic de-/encoding (input/output) system. The perceptual systems and linguistic decoding deliver to the central system schematic representations, in a conceptual format, which carry information about the world, and it is in the central system that information from different sources – the perceptual systems, memory, inference – is integrated. Stainton (1994; 2006b) sketches an account, situated within this picture of cognition, of how what are genuinely only subsentential phrases could be used to express propositions; this goes roughly as follows. The linguistic input is decoded into a conceptual representation that is delivered to the central system, which is where pragmatic inference occurs; the same happens with inputs to the other perceptual systems, e.g. vision. Information stored in encyclopaedic memory or inferred from stored assumptions is also available in the central system, in the same conceptual format, so representations derived from the various perceptual and language systems can be integrated by the central system with information from inference and memory. Integration, suggests Stainton, is performed by function-argument application: the speaker utters either (i) a word or phrase whose content is an argument to some propositional function, and context provides the function, or (ii) a word/phrase whose content is the propositional function, and context provides

the argument. Applying the propositional function to the argument results in the proposition expressed. For example, in the case of (2) above ("Reserved"), the central system will receive from the visual system a concept of the table pointed at (i.e. the argument); the input from linguistic decoding is a property concept (the propositional function), and the two inputs are concatenated in the language of thought (Stainton 2006b: 156).

Stanley's ellipsis account of the 'subsentential' data does not fit very comfortably into this picture of the mind. According to Stanley, the linguistic expression needed to complete the sentence uttered is made salient by the object to which it refers, or by the demonstration of that object (taking as an example (1), "John's father", uttered while glancing at a man across the room). Since it is difficult to see how a demonstration of an object in order to make salient a linguistic indexical (which needs to be assigned the object as content), could avoid simultaneously activating in the hearer a concept of the object, it follows that a conceptual representation of the object must inevitably be tokened even on the ellipsis account. A representation provided by the uttered phrase will also be available in the same format, since perceptual systems translate their input into conceptual representations. And the two LoT representations are concatenated to form a proposition (as on Stainton's account, described above).

On the assumption that such informational integration does take place in the language of thought, then if noticing the object does for some reason activate a natural-language description of it (something that many people would find implausible anyway[5]), or an indexical or demonstrative that is used to refer to it (even more implausible, since these natural-language expressions are too coarse-grained to track objects in the world to which they are used to refer), that would be incidental: the integration takes place in LoT, and natural language plays no further role once the language faculty has delivered its subsentential input to the central system. Construction of the proposition expressed is a matter of free enrichment, which is a process of adding further concepts to a conceptual representation decoded from an utterance. All the required information is present already in a conceptual format and needs to also be integrated in the central systems if it is to be of any use, since this is the format it needs to be in to play a role in thought. Going

[5] As in the following quote from Elugardo and Stainton (2003: 277): "…the idea that to understand less-than-sentential speech one must recover an ordinary natural-language expression that picks out the element supplied by the environment is no more plausible than the idea that whenever someone notices an object, she tokens a natural-language expression that refers to it. It is highly implausible to suppose that, when someone looks at her desk and sees the objects on it, recognizing their features, a constant flurry of English sentences runs through her head. But then why suppose that when one notices an object *being discussed*, and considers its properties, one must token a singular term in English that refers to it?"

through an additional phase of integration in or translation into natural language is, therefore, unnecessarily multiplying representations.

The ellipsis story, requiring that a natural-language representation of the entire sentence be constructed at some point in processing, would initially seem less objectionable if either natural language were the medium of thought, or inputs from different modalities are integrated in natural language. Both options have their adherents: the first is probably taken by some sententialists; the second is the view of Carruthers (2002).

There are quite a few good arguments for natural language not being the medium of thought (Fodor 1975; 1987: appendix), including:

- We probably don't want to say that any humans (e.g. infants, wild children, aphasics) who haven't (yet) acquired a natural language, or have lost their linguistic abilities, can't think.
- We couldn't learn natural language if we didn't have a LoT: identifying speaker's meaning in vocabulary acquisition requires a 'target' against which to form hypotheses, i.e. we have a concept of an object (e.g.), and have to work out what the object is called.
- Linguistic underdeterminacy: sentences are not fine-grained enough to distinguish different thoughts; words are not fine-grained enough to distinguish different concepts (especially the kind of referring expressions – words such as "he" and "that" – posited as involved in the elliptical sentences being discussed here).

It's less than perfectly clear what the sententialists' views are on LoT: Stanley (2000) in his comments about context and interpretation seems to be doing without it; the others, who I discuss near the end of this section, are Merchant (2004), who seems to believe in it, and Ludlow, who thinks that conceptual structure is superfluous (2003: note 17). The Fodorian story described above, with the central system processing and integrating LoT representations received from the perceptual systems and from memory or inference, fits far more naturally with an enrichment account of subsententials than with an ellipsis account. If, on the other hand, that story turns out to be not entirely correct, and natural language is the medium for at least some thought (or just for the integration of contents from different modalities), then it might seem that a serious problem of redundancy raised by LoT for the ellipsis account would disappear, since the representation of the full sentence in natural language would appear to be necessary.

However, there are two related objections to this line of response which mean that it ultimately suffers from the same problem of multiplying levels of representation. The first is that natural language does not include de re individual

concepts[6,7] – it only includes descriptions, indexicals etc. used to refer to individuals (persons or objects); not concepts of them. Because the interpretation of subsententials, with their fairly minimal encoding, requires context to do a lot of work, the identification of the speaker's meaning is often highly dependent on the presence of salient entities in the immediate perceptual environment to serve as the discourse topics. These entities are therefore known to the hearer by acquaintance and he represents them by means of de re concepts, rather than by description. It is a feature of most (perhaps all) of the subsentential cases that part of what is omitted, what needs supplying to construct a full proposition, is a de re concept. So the thoughts that are recovered in interpreting subsententials cannot, in the usual case, be entirely in natural language, but must be, at best, an amalgam of e.g a natural-language predicate and some individual concept in another medium. As on the LoT story sketched above, the contextual salience of the entity (which one presumably has to perceive, before forming a natural-language description of it, or tokening a natural-language indexical or demonstrative to refer to it) means that the individual concept that represents it will be available prior to any representation of it in natural language. So, construction of any natural-language description/indexical is, again, not part of the interpretation process, and would only occur post-hoc.

The second objection concerns any potential cases where the unpronounced part of the recovered proposition does not include a de re individual concept. We might accept that thought occurs partly in natural language and that therefore the proposition recovered in such cases could consist entirely of natural-language material, with the result that, at least for these cases, this natural-language sentence would not be redundant. However, the ellipsis analysis still does not gain much support. The reason is that there is still no evidence that the unpronounced part of the thought was encoded in the uttered fragment, rather than being inferred – since the resulting thought would be the same, whether achieved through the grammatical reconstruction of ellipsed material, or through pragmatic inference. Unless one is antecedently irrevocably committed to the sententialist thesis, there is no reason to think that syntactic ellipsis is involved here. It might appear that there is little to choose between the two accounts for these cases – simpler encoded meaning plus more pragmatic work, versus more complex encoding with automatic reconstruction – but, as I will argue in section 3, considerations of the processing effort involved in decoding and inferential comprehension strongly favour the pragmatic view, even for these examples.

---

[6] I use the term "individual concept" to mean a mental representation of an individual entity, following Powell (2003). Similar to Recanati (1993), Powell "identifies individual concepts with dossiers containing information all of which is taken by the holder of the concept to be satisfied by the same individual" (Powell ibid: 21).

[7] King and Stanley (2005: 130), who seem inclined to the thought-in-natural-language view, also recognize that natural language does not include these entities.

Finally in this section, I will consider a type of ellipsis account which seems to be enjoying some popularity lately, as, when applied to discourse-initial fragments, it promises to explain away the problematic fact that they are unlike other types of ellipsis. These accounts acknowledge that the missing elements of truth-conditional content are not uniquely identifiable, and leave a lot of work to pragmatics. The idea is that what has been deleted, and gets recovered grammatically, is just the syntactic completion of the sentence, and this consists of deictic elements (both pronouns and underspecified verbs, which are never pronounced). The exact semantic content recovered (truth conditions) is therefore predicted to vary between hearers, but what was encoded was still a full (indexical) sentence[8].

Two variants of such an account are given by Merchant (2004) and Ludlow (2005). Merchant acknowledges that discourse-initial fragments do not obey the same conditions as sluicing, VP-ellipsis, etc, conceding that there is no linguistic antecedent to act as a controller for the ellipsis. He does not want to claim that the more fine-grained natural language material that would correspond to the complete proposition expressed can be made salient enough by the discourse context. He recognizes that such a claim would be likely to commit him to the view, which he isn't happy with, "that perception and thought be conducted for these purposes in language itself, in Chomsky's 'narrow language faculty', and not entirely in the language of thought/semantic representations". However, since he is reluctant to countenance a non-standard mapping (i.e. a lack of isomorphism) between linguistic form and truth conditions, he still wants to analyse the discourse-initial fragments as syntactic ellipsis. His suggestion is that, when the elided elements are "[$_{VP}$ *do it*]" or "[$_{IP}$ *this/that* [$_{I'}$ *is* t]]", then the ellipsis can function without a linguistic antecedent, as long as it is clear from the context what "it" and "this/that" refer to (Merchant 2004: 725). The former might cover cases like (4) (picking someone out of a line-up); the latter would apply to (1)-(3) and (6) (where the pronounced fragment predicates something of an object or person, an individual concept of which is the content of the unpronounced deictic).

The immediately obvious problem here is that these two options ("do it" and "this/that is" ellipsis) are insufficient to cover anywhere near all the linguistically discourse-initial examples. Consider (5), said of a used car, or (14), said to a small child who looks like spilling its glass of milk (Stainton's example), or (15)B (due to Robyn Carston):

(5) Only 10,000 miles. Like new.
(14) Both hands.

[8] Whether this *syntactic* material should be consciously identifiable or not is uncertain: It seems to consist of some kinds of deictics which are not merely silent counterparts of overt indexicals – so it would possibly be reasonable to maintain of this kind of thing that we needn't be able to report what the completion of the sentence was.

(15) A: Big house.
B: Four kids.

The first sentence of (5) obviously does not mean "This is only 10,000 miles", or "Only 10,000 miles do it"; (14)-(15) are similarly impossible to account for with only these two candidate completions; see Stainton (2006a: 108-9) for a number of other examples. So the list of options for the elided portion of discourse-initial fragments needs to be expanded.

An account that has a better chance of covering all the possible interpretations of such fragments is that of Ludlow (2005), who proposes that cases of apparently subsentential speech are really full sentences, and the unpronounced material is formed from some combination of the following deictic elements: PRO in subject position; an unpronounced light verb such as "have", "do", "be"; OBJ in object position and DET in determiner position. However, as Stainton (2006b: 126-7) notes, this idea remains at best underdescribed, since in the sketchy form given by Ludlow, it wildly overgenerates elliptical discourse-initial expressions. Nothing is said about where these deictic elements can and cannot occur: To mention two examples given by Stainton, what prevents the subsentence "PRO already bought OBJ" (with the phonological form "already bought") being a well-formed English sentence meaning "Fiona already bought some jam"; and why can these silent deictics not appear in ordinary sentences so that an utterance of "John tall" is a grammatical sentence?

Even if the proposal were more fully described, though, when applied to discourse-initial fragments, it would be subject to more or less the same objection that I argued is faced by the more usual idea of ellipsis as deletion of the syntax and the semantic content: that its only potential merit is to preserve the equation of logical form and truth conditions, and that this rather questionable benefit anyway incurs an unjustifiable cost in introducing extra representations. First, as Carston (2002: 155) says, such a structure with indexicals needing saturation is largely redundant, since the conceptual material necessary to saturate it would be highly activated anyway. The incorporation of this material into the proposition expressed, whether by saturation or free enrichment (concatenation with the decoded subsentential logical form), would involve negligible processing effort. The free enrichment account is thus preferable because of the extra processing effort entailed by the ellipsis account for decoding the silent deictic elements.

A further objection to these 'unpronounced deictic' accounts arises from the fact that pragmatic inference would have to be involved not only in assigning content to the recovered deictics, but also in choosing the correct logical form, prior to reference assignment. Merchant (2004) pretends that this pre-semantic inference can be idealized away, since on his account, there is a choice of precisely two logical forms. His probable reasoning is that, since one of them can occur only in

subject position and the other only in predicate position, they are sometimes mutually exclusive anyway as completions of fragments, so the disambiguation doesn't place undue strain on pragmatics and can be forgotten about. Some people might agree with Merchant (as Stainton 2006a: 101 seems to) that such an idealization is harmless, provided the number of choices is strictly limited to around two. However, as Stainton demonstrates (see the examples on the last page, plus Stainton 2006a) and Ludlow's account also recognizes, the number of choices cannot be restricted to just two, but will need to include other light verbs and deictic elements – besides, there is no reason (other than the desire to keep the number of options down, and with it the amount of pragmatic work involved) why these other silent elements should not feature in Merchant's account as well: if "this" and "that" are allowed, why not "there", "here", "him", and so on; since "is" and "do" are required, what principled reason would there be for excluding other light verbs? In that case, then, when faced with a discourse-initial fragment, the hearer has a large number of possible sentential logical forms to choose among. On Merchant's and Ludlow's accounts, if the fully sentential expression containing the deictic is to play any useful role, the hearer would *first* decide which of the possible deictics is the correct completion of the logical form, and only *then* go about assigning the intended content to those deictics. The sequence of interpretive steps assumed by such accounts is clearly illogical, since the disambiguation required to work out which logical form is being used is dependent on the hearer working out the content that forms the proposition expressed. So we have returned to the familiar objection to ellipsis accounts of these cases: The allegedly reconstructed natural-language material serves no purpose, and simply adds an extra, completely superfluous stage to the comprehension process.

3 Subsententials and indeterminacy

Section 2 argued against treating discourse-initial fragments as syntactic ellipsis. However, if the sententialist maintains that some version of an ellipsis story could be shown to hold for some of the data, there is a further set of cases which even he is unlikely to try and subsume under such an account. To avoid the conclusion that free enrichment is involved in interpreting them, another response that has been suggested is to deny that some fragments are used to perform genuine linguistic speech acts.

According to Stanley, "Linguistic speech acts must determinately be made with the relevant sort of force. They must also express determinate contents" (2000: 407). The example he discusses is (16):

(16) (Thirsty man staggers up to water vendor)
Water!

Stanley concludes that an utterance of (16) is not a proper speech act as it lacks determinate content (the truth conditions are not determinately that THE SPEAKER WANTS WATER, as opposed to a number of other options, such as that THE ADDRESSEE SHOULD GIVE THE SPEAKER WATER) and illocutionary force (is (16) an assertion, order, request, …?). While such utterances can undoubtedly be used as vehicles of communication, beyond the initial decoding it is general (as opposed to linguistic) knowledge that is brought to bear in interpreting them; Stanley likens this sort of communication to a kick under the table, a tap on the shoulder, or a frown. All that these utterances and non-linguistic gestures communicate is implicatures: they don't have propositions expressed, so the question of free enrichment contributing to truth conditions does not arise.

In response, Stainton (2006b) and Elugardo and Stainton (2004) argue that the requirement that, to be a genuine speech act, something must have determinate content and force, is far too strong, as, on such a criterion, many cases of fully sentential speech would, contrary to everyone's intuitions, not count as performing speech acts. Two of Stainton's examples are given here:

(17) (Looking out at Grand Canyon)
That's beautiful.
(18) You must turn in your report before you leave today.

Of (17), Stainton asks whether there must be a determinate referent for "That" – a particular object or collection of objects – for the utterance to count as an assertion. Intuitively, this is not required, in which case we have a speech act where there is no determinate content. Similarly, (18) is undoubtedly a speech act, though we might not be certain what force it has – e.g. whether it is an order, or an assertion of policy, or a request. If the 'determinacy' criterion is applied consistently, then much of verbal communication – whether sentential or subsentential, even if judged grammatical, and if grammatically and semantically complex – will not count as performing linguistic speech acts. Interestingly, Stanley and Szabo (2000: 237-8) themselves acknowledge that in quantifier domain restriction, the context does not provide the unique descriptive material that specifies the domain, and this is the reason they give for not treating this phenomenon as syntactic ellipsis: discussing the quantifier "every", they note that "there are very few cases where there is a single plausible candidate for the role of the domain restricting predicate", whereas "In cases of syntactic ellipsis, there is a unique phrase recoverable from the context". Since in the usual (i.e. non-referential) cases of quantifier domain restriction, the descriptive material (the domain-restricting predicate) constitutes

the content, Stanley and Szabo are agreeing that there is indeterminacy of content here (and the same will apply to non-referential uses of definite and indefinite descriptions). Yet Stanley apparently fails to notice the implications for this of his determinacy criterion: having admitted that sentences needing domain restriction often do not have determinate contents, yet accepting that they are used to perform speech acts, it follows that the (in)determinacy of content should not be used as a criterion to decide whether a given utterance constitutes a genuine linguistic speech act, and the decision should be made on other grounds.

Intuitions, which Stanley et al profess to take seriously as the core data for a semantic theory to explain, are that the domain restriction cases and many (apparent) subsententials do express truth conditions. In this regard, Clapp (2005) also argues that the determinacy criterion and intuitions pull in different directions. In cases where intuitions are that an utterance is used to express truth conditions, but where it is indeterminate exactly what those truth conditions are, then applying the determinacy criterion strictly would require one to accept that speaker-hearer's intuitions about truth conditions must be wrong. But if Stanley were to claim that intuitions about truth conditions are often wrong, it would undermine his entire project of trying to use semantic theory to account for them: It would render otiose his syntactic strategy (the appeals to syntactic ellipsis and shorthand to account for subsententials, and the positing of hidden indexicals to account for quantifier domain restriction and other effects of context in sentential utterances (since any cases of quantifier domain restriction etc. that do have determinate contents could be analysed as elliptical)). If one is claiming that people are often mistaken about the truth conditions of their utterances, then rather than modifying linguistic theory with hidden indexicals and novel forms of syntactic ellipsis in order to account for these intuitions, one could instead become a semantic minimalist à la Cappelen and Lepore (2005) or Borg (2004). As Clapp says, **all** fragments that are not **obvious** cases of syntactic ellipsis (i.e. the cases such as (1)-(7), for which it needs to be argued that they are elliptical, precisely because the exact deleted syntactic and semantic material is not identifiable, i.e. their content is indeterminate) will not qualify as expressing truth conditions, so would not need accounting for by semantic theory. To allow for the indeterminacy displayed by (17) and (18), and quantifier domain restriction, and many other cases, while still respecting speaker-hearer intuitions that such utterances perform speech acts (with truth conditions), the determinacy criterion will have to be loosened to an extent that it will actually exclude very few (allegedly) subsentential utterances. The majority, then, would have to be analysed as syntactically elliptical – an option which I hope has now been discounted.

Returning now to the subsentential cases like (16) (Thirsty man: "Water!"), which probably does exhibit a greater degree of indeterminacy than most other sentential or subsentential utterances, I'll consider how these should be treated:

whether Stanley is correct to deny that they perform speech acts, or whether they should be treated as speech acts, expressing truth conditions, despite their obvious indeterminacy.

I agree with Stanley that we can single out a category of utterances that do not express propositions. Likely examples that fall into this category are (19) and (20):

(19) (Mother to husband upon realizing they've mislaid their three-month-old) The baby!
(20) (Uttered by someone who's gone out and remembers she left the oven on) The oven!

With these utterances, it may well be the case that the speaker does not intend to explicitly express any proposition, and the hearer does not assign any truth conditions to the utterance. It hardly makes sense to ask whether the truth conditions/proposition expressed by (19) is THE BABY IS NOT HERE, or WE HAVE FORGOTTEN THE BABY, or WE SHOULD FIND THE BABY, and so on. (19) and (20) probably function simply as devices to focus the hearer's attention on, e.g., the said baby, and therefore may fall in with non-verbal ostensive gestures, such as demonstrations, which by themselves would not be taken to express truth conditions.

The case of "Water!", however, seems different. Intuitions are that the speaker has *said* something: her aim is not to focus the hearer's attention on water, but rather to express a proposition. And the interpretation that the hearer ends up with includes a proposition that is a development of the encoded phrase. Although there is likely to be variation between hearers in what they take the proposition expressed to be, this variation will be restricted to a clear range – including THE SPEAKER WANTS WATER; THE HEARER SHOULD GIVE THE SPEAKER WATER; etc. Unlike (19) and (20) above, it at least seems meaningful to ask the question of what proposition was expressed, which indicates that we have intuitions that there exists such an entity, and the suggestion that I will develop in more detail shortly is that the proposition that the hearer constructs, developing the encoded meaning, is what should count as the proposition expressed. "Water!" is probably a 'directive' of some sort, rather than an assertion, which complicates the question of how to judge what its truth conditions are. However, intuitions are much clearer with other cases. Recall the example which Stanley (2000) originally treated as shorthand:

(6) Nice dress.

As I said in discussing this example in section 2.1, an utterance of (6) is intuitively false if the speaker does not in fact believe that the dress in question is nice. So speakers and hearers undoubtedly take the utterance of (6) to have truth

conditions (and to perform a successful speech act), although there may be some indeterminacy about illocutionary force ((6) may be an assertion, but it is also possible that it is an exclamative), and there is certainly a fair amount of indeterminacy about what the truth-conditional content is (THAT IS A NICE DRESS; YOU ARE WEARING A NICE DRESS; YOU BOUGHT A NICE DRESS; or maybe WHAT A NICE DRESS).

So the contextualist is likely to say that an utterance such as (6) or (16) does express a proposition, and so includes unarticulated constituents, despite the indeterminacy. In what follows I sketch a contextualist account of the interpretation of subsententials using Relevance Theory (Sperber and Wilson 1986/1995). According to RT, all pragmatic processes operate according to the same principles, no matter at what level of representation their results are seen. Consider pragmatic processes that do not contribute to the proposition expressed: in cases of conversational implicature, non-verbal or paralinguistic communication, interjections, and so on, it is implausible that satisfying the speaker's communicative intention involves the hearer recovering exactly the thought content that the speaker had in mind (and in many cases, it is anyway unlikely that the speaker had any very specific content in mind). Everyone would agree that such communication can be successful while typically incorporating a great deal of indeterminacy about the exact set of propositions that the speaker intends the hearer to construct. It is also generally accepted that (virtually) every utterance requires some degree of pragmatic inference to arrive at the proposition expressed.

Following Sperber and Wilson, the domain of pragmatics is the class of ostensive stimuli, whether verbal or non-verbal, and all such stimuli are interpreted by a single pragmatics system employing the same pragmatic principles. By virtue simply of not being decoding, but rather hypothesis formation and confirmation, all pragmatic inference involves some leeway for divergence between the thought that the speaker has in mind and the thought that the hearer infers: strict duplication of thoughts is anyway an unrealistic requirement and not necessary for successful communication, but, depending on the accuracy demanded in a given discourse situation, any of a range of propositions might be near enough. Given this and the fact that linguistic meaning virtually always underdetermines the proposition expressed anyway, there is no motivation for singling out the proposition expressed as necessarily having to meet a higher standard of determinacy than any other communication, and no requirement that a single unique content be grasped by both speaker and hearer.

A considerable degree of indeterminacy about the proposition expressed is, then, acceptable from the RT and contextualist point of view. A sketch of the RT account and justification for this, concentrating on subsententials, is as follows. A speaker, judging what information will be manifest to the hearer, can have some more or less precise expectations about what interpretation the hearer can construct from a

given utterance. So a speaker who utters a subsentential phrase, having available the linguistic resources to be more explicit, can be assumed not to have any one particular proposition in mind that she expects the hearer to recover; instead, a number of propositions will be compatible with the speaker's communicative intention, and so this intention will be satisfied if the hearer recovers any one of these propositions. Whichever of these propositions the hearer constructs, it will inevitably have some constituents supplied by free pragmatic enrichment, since the subsentential linguistic input did not encode a propositional schema.

From the contextualist (and particularly RT) point of view, then, the indeterminacy about truth conditions[9] that is the result of interpreting many subsentential utterances (and many utterances in general) is no disadvantage, and is to be expected given reasonable assumptions about what is required for successful communication. Far from being problematic, the acceptance of indeterminacy in fact is tied up with a significant advantage for any pragmatic theory that, as it needs to if it aims at being a theory of the actual processes and principles used in utterance comprehension, takes account of the interaction of the processing effort expended and the effects achieved. If as much equivalence as possible between the thoughts of speaker and hearer were the aim, then the greater amount of encoding, the better, since less work is left to pragmatic inference. However, decoding linguistic expressions requires processing effort, and further effort will be needed to saturate, disambiguate, enrich and otherwise modulate the linguistic logical form. So it is clear that on occasions where, for example, the context makes uniquely salient an object that the speaker wants to refer to – say when speaker and hearer are both looking at a particular restaurant table, and this fact is mutually manifest to them – the hearer will be entertaining a representation of the object, in the conceptual format in which it is available for integration with representations from other (e.g. linguistic) sources. In that case, it may well be that, as Carston (2002: 154-6) suggests, it is more effort for the hearer to decode the fully sentential utterance "That table is reserved" and assign reference to the demonstrative description, than to interpret just the subsentential utterance "Reserved" and work out that this property is being predicated of the table. This working out would, on an RT account of the comprehension procedure, involve minimal effort: according to RT, the fact that a speaker, by addressing a hearer, is demanding some processing effort from him, licenses a particular comprehension procedure for interpreting ostensive stimuli:

[9] Note that the propositions resulting from pragmatic processing and forming the truth-conditional content will themselves be determinate. What is indeterminate is exactly what falls under the speaker's communicative intention.

> Relevance theory comprehension procedure: Test interpretive hypotheses in order of accessibility; stop when you find an interpretation that meets your expectations of relevance.

The speaker, wanting to get her message across, and being able to predict to some extent what information is accessible to the hearer, will try to shape her utterance to ensure that the hearer constructs an interpretation that satisfies her communicative intention, and to minimize the risk of misunderstanding. The tacit knowledge that interlocutors have this ability means that the first interpretive hypothesis to occur to the hearer has a high degree of plausibility simply by virtue of occurring first, since its occurrence should have been predictable by the speaker. In the context described above for the utterance of "Reserved", a concept of the table will be the first argument tested as something that could combine with the predicated property, since it is mutually manifest to speaker and hearer that the table is highly salient. The resulting interpretation, that the table is reserved, is relevant enough, so is accepted, and the comprehension procedure does not go on to consider further hypotheses about the proposition expressed. The efficiency of this comprehension strategy is even clearer in the case of (1), the utterance of "John's father". Carston (2002: 155) points out that, given the context in which the speaker's demonstration makes the referent salient, the hearer may have available to him any one of a number of different representations of the referent (THE MAN WHO IS STANDING NEXT TO THE DOOR IS X; X HAS JUST WALKED IN; etc), and, on the subsentential account, the decoded phrase can simply slot in to whichever of these conceptual representations is most salient in the hearer's mind.

Although Stanley's "Water!" example is much less determinate than most other alleged subsententials, I don't see any problem in analysing this as also expressing a proposition. It doesn't seem to fall in with Stanley's examples of non-verbal communication: a kick under the table or a tap on the shoulder are simply devices for getting attention; there is no intuition that anything is decoded from these actions which serves as a constituent of whatever is communicated. The thirsty man is a rather different matter from just attention-getting, as he can be assumed to have been trying to communicate a proposition with his utterance of "Water!". And the hearer will, in interpreting the utterance, recover a proposition, and get some information to the effect that the man wants some water: there is a clear range of candidates for the proposition expressed which would satisfy the speaker's communicative intention. To relegate this to the status of a mere implicature would be unintuitive, and the only reason for claiming that it can only be implicated, and not the proposition explicitly expressed, can be its high degree of indeterminacy. So to claim that the utterance has no truth conditions is to say that hearers' intuitions are mistaken. In which case, this theory faces the conundrum posed by Clapp (2005) that, if our intuitions about the scope of semantic theory are wrong, it

makes no sense to modify grammar or semantic theory in order to render our intuitions correct. RT, on the other hand, accepts that, on occasion, there may be quite a lot of indeterminacy in the proposition expressed (and has to accept this, given how much pragmatic work is supposed to happen between logical form and proposition expressed), so does not face the problem of having to draw an arbitrary cut-off point beyond which an utterance has too indeterminate a content to count as expressing a proposition/having truth conditions.

4 Concluding remarks

In the last section, I've sketched a relevance-based pragmatic account that accommodates subsententials and indeterminacy quite neatly. There are several aspects of the above picture that I suspect the truth-conditional semanticist would not be enthusiastic about, so, in this concluding section, I will consider briefly whether he would be justified in rejecting it. First, how determinate does a proposition have to be to count as the proposition expressed? It seems to be accepted that much of linguistic communication suffers from the meaning-intention problem (Schiffer 1995; see also Wettstein 1981): for cases of quantifier domain restriction, propositional attitude reports, and so on, no facts about either the context or the speaker's intentions can identify a unique proposition expressed. But the construal of this as a problem assumes that there is some abstract interpersonally or metaphysically determined entity that is 'the proposition expressed'. It is agreed that hearers cannot actually recover such an entity, even assuming that it has some reality. So there is no sense in considering this abstract entity the object of explanation of a theory that aims to account for how hearers really interpret utterances online, which is what Stanley (2005a) states that semantic theory should do, and what his approach has to do if it is being presented as an alternative to a pragmatic enrichment account. What must be explained, then, is how hearers grasp the proposition expressed/truth conditions that they actually do grasp, and this is a determinate proposition, with truth conditions – the issue of whether it is determinately *the* thing that the speaker had in mind becomes unimportant. Moreover, Stanley agrees that truth-conditional semantics cannot account for the proposition that the hearer does recover from an utterance of a subsentential phrase: it is *pragmatically* developed into a full proposition. To avoid an obviously question-begging argument (if it's indeterminate, it's not a speech act; if it's determinate, it can't be free enrichment so it's elliptical), there would have to be some evidence that the result of such pragmatic development is inevitably too indeterminate to count as a speech act. But I predict that no such evidence would be forthcoming: after all, there are undoubtedly cases where the results of optional pragmatic inference are determinate and there is practically no freedom for the

hearer to construct a different interpretation: take the case of scalar implicatures where "some" implicates "not all", or indirect answers to "yes/no" questions, as simple examples. This is a further reason why, from the fact that a given apparently subsentential utterance has determinate content, it cannot be concluded that, because pragmatic processes are inherently too imprecise to have succeeded in arriving at this particular content, the utterance is syntactically elliptical. Indeterminacy has to be allowed in assigning values for indexicals ("here", "now", and "there" being obvious examples where there can be considerable leeway in the exact values for locations or times that the hearers assign), quantifier domains, etc. So I see no principled justification for allowing indeterminacy resulting from saturation to be part of the proposition expressed, but excluding any indeterminacy that cannot be traced to saturation, if intuitions are that these pragmatically supplied elements contribute to truth conditions[10]: This is obviously an idealization to make a small part of communication partially tractable for a semantic theory, but cannot serve as a theory of utterance interpretation.

References

Barton, E. 2006. Towards a nonsentential analysis in generative grammar. In Progovac et al (eds), 11-31.

Borg, E. 2004. *Minimal Semantics*. Oxford: Oxford University Press.

Botterell, A. 2005. Knowledge by acquaintance and meaning in isolation. In Elugardo, R. and Stainton, R. (eds), 165-84.

Cappelen, H. and Lepore, E. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.

Carruthers, P. 2002. The cognitive functions of language. *Behavioral and Brain Sciences*, 25:6: 1-18.

Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. 2004. Explicature and semantics. In Davis, S. and Gillon, B. (eds) *Semantics: A Reader*, 817-845. Oxford: Oxford University Press.

Clapp, L. 2005. On the interpretation and performance of nonsentential assertion. In Elugardo, R. and Stainton, R. (eds), 109-30.

Elugardo, R. and Stainton, R. 2003. Grasping objects and contents. In Barber, A. (ed.) *Epistemology of Language*, 257-302. Oxford: Oxford University Press.

[10] The justification that the truth-conditional semanticist would give is, of course, that the only pragmatic contributions to truth-conditional content must result from saturation if we are to maintain (a version of) the principle of compositionality for this level of content, so that the truth conditions of an utterance result from combining the values of (only) the constituents of the sentence. The arguments in section 2 against ellipsis accounts showed that the level of intuitive truth-conditional content of these fragments, which is the propositional level that the sententialist accounts aim to explain, does not meet the principle of compositionality.

Elugardo, R. and Stainton, R. 2004. Shorthand, syntactic ellipsis, and the pragmatic determinants of what is said. *Mind and Language* 19: 442-71.
Elugardo, R. and Stainton, R. (eds) 2005. *Ellipsis and Non-sentential Speech*. New York: Springer.
Fodor, J. 1975. *The Language of Thought*. New York: Thomas Crowell.
Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
King, J. and Stanley, J. 2005. Semantics, pragmatics, and the role of semantic content. In Szabo, Z. (ed.) *Semantics vs. Pragmatics*, 111-64. Oxford: Oxford University Press.
Ludlow, P. 2003. Referential semantics for I-languages? In Hornstein, N. and Antony, L. (eds) *Chomsky and his Critics*, 140-161. Oxford: Blackwell.
Ludlow, P. 2005. A note on alleged cases of non-sentential assertion. In Elugardo, R. and Stainton, R. (eds), 95-108.
Merchant, J., 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27, 661–738.
Merchant, J. 2006. Small structures: A sententialist perspective. In Progovac et al (eds), 73-91.
Powell, G. 2003. *Language, Thought and Reference*. PhD dissertation, University of London.
Progovac, L., Paesani, K., Casielles, E. and Barton, E. (eds) 2006. *The Syntax of Nonsententials: Multidisciplinary Perspectives*. Amsterdam and Philadelphia: John Benjamins.
Recanati, F. 1993. *Direct Reference: From Language to Thought*. Oxford: Blackwell.
Recanati, F. 2002. Unarticulated constituents. *Linguistics and Philosophy* 25: 299-345.
Schiffer, S. 1992. Belief ascription. *Journal of Philosophy* 89: 499-521.
Sperber, D. and Wilson, D. 1986/1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.
Stainton, R. 1994. Using non-sentences: an application of relevance theory. *Pragmatics and Cognition* 2: 269-84.
Stainton, R. 2004. The pragmatics of non-sentences. In Horn, L. and Ward, G. (eds.) *The Handbook of Pragmatics*, 266-87. Oxford: Blackwell
Stainton, R. 2006a. Neither fragments nor ellipsis. In Progovac et al (eds), 93-116.
Stainton, R. 2006b. *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford: Oxford University Press.
Stanley, J. 2000. Context and logical form. *Linguistics and Philosophy* 23: 391-434.
Stanley, J. 2002. Making it articulated. *Mind and Language* 17 1&2: 149-68.
Stanley, J. 2005a. Semantics in context. In Preyer, G and Peter, G. (eds) *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, 221-54. Oxford: Oxford University Press.
Stanley, J. 2005b. Review of François Recanati's Literal Meaning. *Notre Dame Philosophical Reviews*.
Stanley, J. and Szabo, Z. 2000. On quantifier domain restriction. *Mind and Language* 15: 219-61.
Taylor, K. 2001. Sex, breakfast, and descriptus interruptus. *Synthese* 128: 45-61.
Wettstein, H. 1981. Demonstrative reference and definite descriptions. *Philosophical Studies* 40: 241-57.

Plurals, possibilities, and conjunctive disjunction[*]

NATHAN KLINEDINST

Abstract

Sentences with disjunction in the scope of a possibility modal often convey something stronger than predicted under the standard semantics for modals and disjunction, roughly paraphrasable in terms of a wide scope *con*junction. *You may have beer or wine*, for example, is naturally understood as conveying that you may have beer **and** that you may have wine. This 'puzzle of free choice permission' (Ross, 1941) has spurred a wide array of revisions, often radical, to standard assumptions about disjunction, modality, scalar implicature, or some combination thereof. A deeper puzzle is that, while there is good evidence that the conjunctive effect is due to SCALAR IMPLICATURES, standard and well-motivated assumption about the latter predict precisely that it should **not** arise.

In this paper I observe that plural existential quantifiers – but not singulars – pattern with possibility modals in giving rise to an analogous conjunctive effect, and that identical analytical puzzles arise. These patterns remain mysterious on otherwise plausible revisionist accounts. I explore the possibility that the pattern is in fact entirely revealing: possibility modals behave like plural existentials because they are. I suggest a unified account of the conjunctive effect as due to an 'embedded implicature' triggered by a DISTRIBUTIVE OPERATOR (which distributes over the parts of the plurality introduced by the plural existential/possibility modal). This implicature is a subcase of those generally triggered by universal quantifiers over disjunction.

[*]This paper is based on a talk given at *Sinn und Bedeutung 10*, a highly condensed and simplifed version of Chapters 1 and 2 of my Ph.D Thesis (*Plurality and Possibility*, UCLA, 2006). The latter should thus be consulted for important details omitted or glossed over herein, and is preferable as a citation. I gratefully acknowledge the help of Philippe Schlenker and Danny Fox, and support from the following grant: ACI Systèmes Complexes en SHS ('Implicatures, Sémantique Dynamique et Thèorie du Choix Rationnel'), CNRS/Institut Jean-Nicod, 2003-2005. Thanks also to Benjamin Spector, Denis Bonnay, and Emmanuel Chemla for insightful comments.

1 Introduction

1.1 An old puzzle

The information conveyed by a sentence with *or* in the scope of *may*, *might*, *can*/*could*, etc. is often stronger than expected under the standard treatment of such expressions as (equivalent to) $\Diamond$-operators of modal logic, and of *or* as (equivalent to) boolean disjunction. For example, (1) naturally conveys that *both* drinking beer and drinking wine are permissible options for the hearer – (1b), but is predicted under the latter assumptions to be true even if only one of them is permissible – (1a).[1,2]

(1) You may have beer or wine.
   a. $\Diamond(\mathrm{B} \vee \mathrm{W})$ *n.b.* $=\Diamond\mathrm{B} \vee \Diamond\mathrm{W}$
      (1a) is True in w iff $\exists \mathrm{w}'$
   b. $\Diamond\mathrm{B} \wedge \Diamond\mathrm{W}$

(Throughout the paper we follow the practice of using the first letter of the first contentful word of a disjunct as its translation in the language of propositional modal logic).

The same puzzle crops up across modalities ((2a)-(2c) allow for identical strengthenings), but for historical reasons – it was first noticed in work on deontic logic (Ross, 1941) – it is known to as 'the puzzle of free choice permission'.[3]

(2) a. Jenny may/might be a doctor or a lawyer. (*epistemic*)
   b. Jenny could/might have been a doctor or a lawyer. (*metaphysical?*)
   c. Jenny can outsmart a doctor or a lawyer. (*ability*)

In each case, rather than (just) $\Diamond(D \vee L)$, the logically stronger $\Diamond D \wedge \Diamond L$ is naturally understood (using the corresponding, appropriate accessibility relation for '$\Diamond$'). That the same conjunctive effect is observed across kinds of modalities casts doubt on the possibility that the effect with (1) is due solely to something special about *the act of* granting permission. This point is made

[1] Since (1a) only requires that there exist a possible world/state of affairs in which the laws/rules (of the actual state of affairs w*) hold, and in which the disjunction holds, i.e. that

(i) $\exists \mathrm{w}(\mathrm{wRw}^* \wedge (B \text{ is in True in w} \vee W \text{ is true in w}))$ *where* R *is the relation of deontic accessibility*

[2] (1b) does not in fact fully characterize the meaning actually conveyed, a point to which we return, but is sufficient to frame the basic puzzle.

[3] We follow tradition in using deontic examples – though the account developed applies in full generality across modalities – and stick to the somewhat misleading terminology.

more directly by the fact that the conjunctive effects arises equally where the function of (1) is to report permission (*according to this rulebook...*), rather than grant it.

At the same time, there are good reasons to assume that the explanation for the puzzle is not that the standard truth conditions are (drastically) wrong, but rather that the strengthening is pragmatic in nature. In particular, the strengthened meaning bears all of the hallmarks of being due to scalar implictures. Far from explaining away the puzzle, however, this observation deepens it; standard and well-founded assumptions about scalar implicature predict precisely that such a strengthening should be *im*possible. Before turning to an elaboration of the latter two claims in Section 2, we show that the puzzle generalizes beyond possibility modals, arising for plural existential quantification over the domain of individuals and times, and sketch an account that builds crucially on the latter observation.

1.2 A new puzzle

Plural existential quantifiers in both the individual and temporal domain can give rise to conjunctive effects for *or* in their scope, completely parallel to the case of possibility modals. Singular existential quantifiers systematically cannot. The following, for example, are naturally understood as conveying the conjunctive (b) meanings, rather than the expected disjunctive ones ((a)):

(3) (The air in the train was extremely stuffy...) Some passenger**s** became nauseous or had trouble breathing.

a. $[\exists X\text{: } P(X)]\ (N(X) \vee T(X))$

b. $[\exists X\text{: } P(X)]\ N(X) \wedge [\exists X\text{: } P(X)]\ T(X)$
'Some passengers became nauseous, and some passengers had trouble breathing.'

(4) (his year at work has been very difficult...) Sometime**s**/at time**s** John broke down into tears or was too tired to continue.

a. $[\exists T]$ (J-cries at T $\vee$ J-too-tired at T)

b. $[\exists T]$ J-sick at T $\wedge$ $[\exists T]$ J-too-tired at T
'Sometimes John broke down into tears, and sometimes he was too tired to continue.'

(The variables 'T' and 'X' range over pluralities (which can be thought of as sets of atomic individuals), and the predicate letters in the above should be read

as expressing 'distributive' properties, i.e. properties that hold of a plurality p iff they hold of each (atomic) individual that is part of p (i.e. each member of the set)).

Similar to the case of possibility modals, the (a) meanings are too weak because they are consistent with one of the disjuncts failing to be satisfied be anything in the domain of the quantifier – for example, in the case of (3), with it being the case that no passenger got sick.[4] Similarly, the conjunctive effect is not predicted straightforwardly even under the more realistic assumption (to be adopted later) that the disjoined predicates in (3) and (4) are simple predicates of atomic individuals, with distributivity handled by an intermediary operator (see Schwarzschild (1996) for a summary of motivations for this assumption). For example,

(5) $[\exists X\text{: } P(X)]\ [\forall x\text{: } x \text{ is-part-of } X]\ (N(x) \vee T(x))$

can be witnessed (made true by the existence of) a "homogeneous" plurality of passengers, i.e. one containing only passengers who got nauseous (or only passengers who had trouble breathing).

A conjunctive effect is impossible with the singular counterparts of (3) and (4); only the (a) meanings – here, in accordance with expectation – are attested:

(6) (The air in the train was extremely stuffy…) Some passenger/a passenger became nauseous or had trouble breathing.

a. $[\exists x\text{: } P(x)]\ (N(x) \vee T(x))$

b. $*[\exists x\text{: } P(x)]\ N(x) \wedge [\exists x\text{: } P(x)]\ T(x)$
'Some passengers became nauseous, and some passenger had trouble breathing.'

(7) (his year at work has been very difficult…) Once/at at least one time, John broke down into tears or was too tired to continue.

a. $[\exists t]$ (J-cries at t $\vee$ J-too-tired at t)

b. $*[\exists t]$ J-sick at t $\wedge$ $[\exists t]$ J-too-tired at t
'At least once John broke down into tears, and at least once he was too tired to continue.'[5]

[4]The (b) paraphrases given here are in fact not entirely correct, a point which is later addressed in full, but the difference is irrelevant for present purposes.

[5]I henceforth focus attention on the examples with individual quantifiers, ignoring temporal cases, though the proposal developed applies in full generality to both.

The conjunctive phenomenon observed with plural existentials is on the face of it exactly the same as that observed with possibility modals, and, as we'll see, the analytical problem it poses is identical: the strengthening seems to be due to scalar implicatures, but it (apparently) can't be. Adding to the mystery is the fact that possibility modals do not pattern with singular existentials. We might expect precisely the opposite, since on the standard assumption they have the same semantics, possibility modals being singular existential quantifers over possible worlds. The observation of the asymmetry between plural existentials/possibility modals and singular existentials, previously unobserved or ignored in the literature, turns out to be the key to unravelling the puzzle of free choice permission.

1.3 Towards one solution

Summarizing, we are apparently faced with 3 problems: explaining (i) how the conjunctive strengthenings arise for possibility modals, (ii) how they arise for plural existentials, and (iii) why they don't with singular existentials (outside the modal domain). The goal of this paper is to dissolve them in turn.

(i) will be assimilated into nothing more than a sub-case of (ii), as we propose that possibility modal *are* in fact plural existential quantifiers, over possible worlds. From this it follows that (iii) isn't a well-defined problem in the first place: there is nothing mysterious a priori about why possibility modals don't behave like singular existentials, since they aren't. The lack of conjunctive strengthenings with singulars itself is shown to be unsurprising, for exactly the reason that there seemed to be a puzzle about possibility modals in the first place: otherwise motivated assumptions about scalar implicature straightforwardly rule them out.

What is in need of explanation from this perspective is just (ii): what is special about *plurals*. We observe that *overtly* distributive plural existential quantification yields a conjunctive effect, just as found with (3), (4) (as with these, the effect doesn't follow from the semantics: see the discussion under ex. (5)):

(8) Some of the boys each wrote a poem or novel.
$\rightsquigarrow$ Some of the boys wrote a poem, and some of the boys wrote a novel.[6]

The proposal developed here has two parts. The first is that distributivity is a

[6] '$\rightsquigarrow$' is used here informally for 'conveys', and later in a technical sense for '(all else being equal,) leads to the scalar implicature that'.

necessary condition for the conjunctive effect with *or* in the scope of existentials – modal or non-modal. The second is that it is an implicature calculated based on the distributive operator (overt or covert) and *or* that is responsible – in particular, an *embedded* implicature added within the scope of the existential. To see how the idea works, observe that a sentence like

(9) Each of them got nauseous or had trouble breathing

implicates that

(10) At least one of them did each of the things.

It is shown that this can be given simple explanation as due to scalar implicatures, derived by comparison to the simpler, stronger sentences with one disjunct eliminated, *Each of them got nauseous/had trouble breathing*. Now we observe that the nuclear scope of (3) is essentially a sentence like (9), where *them* is bound by the existential *some passengers*. By effectively adding the implicature ((10)) which arises for (9) in general to its *embedded* instance in (3) – and binding its pronoun to the existential – we straightforwardly derive the conjunctive effect. (cf. 'There are some passengers such that: each of them N or T, and at least one of them N and at least one of them T').

The second part of the proposal is of course independent from the first, but is shown in Section 3 to have make some desirable predictions.

I will *not* give a knockdown argument that possibility modals are plurals, offer only considerations of conceptual naturalness, and some preliminary empirical and conceptual arguments (Section 5). I also show that the assumption is very weak (Section 4.1), and hence harmless (but in proportion difficult to prove).

Before developing the proposal in some further detail, we turn to motivating its most basic facet. Why should we want an account of free choice permission (and conjunctive effects with plurals) as due to scalar implicatures, and where do existing (scalar implicature) based accounts go wrong?

2 The Implicature Paradox

As noted by Alonso-Ovalle (2005), the conjunctive effect with *or* under possibility modals has the hallmarks of strengthened meanings due to scalar implicatures: it (very strongly tends to) disappear in downward entailing contexts, and

can be 'cancelled'. We begin with the first point, and return to the latter below; consider:

(11) You may not have beer or wine.

The overwhelmingly natural 'reading' is exactly the one that is expected under standard assumptions – in light of the independent fact that negation always scopes above deontic modal auxiliaries in English (e.g. von Fintel 2006): $\neg\Diamond$(B $\vee$ W), =$\neg\Diamond$B $\wedge$ $\neg\Diamond$W (no beer, no wine!). The same point is not as easy to establish with plural existentals – as least with *some* – since they are so-called 'postive polarity items', resisting appearance in DE contexts in which implicatures routinely disappear. It clearly holds for their negative polarity variants, however:

(12) John doubts that any students drank beer or wine.

a. $\neq$'John doubts this: some students drank beer, and some students drank wine.'

For many people *some* can appear unproblematically in the scope of weakly downward entailing operators, (i.e. non-antiadditive ones) like *at most n people* (Szabolcsi, 2004), allowing the point to be established more directly:

(13) At most three people sent some friends a card or a letter.

a. $\neq$'At most three people are such that they sent some friends a card, and some friends a letter.' (i.e. possibly a fourth sent some friends a card, but no friends a letter, or vice versa)

b. ='At most three people sent a card or a letter to some friends.'
entails: no more than three people sent cards to friends, and no more than three sent letters

There is little hope that the conjunctive effect can be derived in a standard way as scalar implicatures, however. Consider the robust empirical generalization, (14), witnessed by the examples following it:

(14) For any embedding X (possibly null), if *XA*/*XB* is logically stronger than *X(A* $\vee$ *B)*, *X(A* $\vee$ *B)* has among its implicatures that $\neg$KXA and $\neg$KXB (where *K* means 'the speaker knows/is certain that)

(15) Alex drank beer or wine. =B(a) $\vee$ W(a)

a. $\rightsquigarrow$The speaker isn't certain that Alex drank beer, and isn't certain

that Alex drank wine

(16) Most students drank beer or wine. =[Most x: S(x)]: B(x) ∨ W(x)

a. ⇝The speaker isn't certain that most students drank beer, and isn't certain that most students drank wine

(17) Every student drank beer or wine. =[∀x: S(x)]: B(x) ∨ W(x)

a. ⇝The speaker isn't certain that every student drank beer, and isn't certain that every student drank wine

(18) John must clean his room or take out the trash. =□(C ∨ T)

a. ⇝¬□C, ¬□T (=◇¬C, ◇¬T)

This generalization has a fairly straightforward explanation in classical Gricean pragmatics as a kind of Quantity/scalar implicature (as well as in neo-Gricean system, e.g. Sauerland 2004). If the speaker knew that XA/XB, saying so would have been more informative (and briefer, no less). So as long as knowing the truth of these more informative statements is relevant to the purposes of the conversation, and the speaker is assumed to be cooperative, a hearer is licensed to infer that the speaker wasn't in a position to assert *XA/XB*, i.e. doesn't know that XA/XB. As is often observed, these implicatures tend, as a matter of empirical fact, to strengthen to K¬XA/XB, where certain conditions are met. So for example (17) typically conveys in addition to its literal content that (the speaker knows that) not every student drank beer (¬[∀x: S(x)] B(x)), and that (the speaker knows that) not every student drank wine (¬[∀x: S(x)] W(x)) – which is to say, that there were both wine drinking and beer drinking students.

Of course possibility modals and plural existentials are precisely embedding operators that fall under the antecedent of the conditional in (14). For example *John may have beer* is logically stronger than *John may have beer or wine*, and *Some passengers got nauseous* than *Some passengers got nauseous or had trouble breathing* – both according to standard semantics assumptions, and as an apparent matter of empirical fact (viz. the fact noted above, that embedding the disjunctive variants under a DE operator (e.g. negation) leads to a *stronger* statement). The problem, then, is that Gricean/neo-Gricean acounts of (14) would seem to be at dire risk of predicting that the conjunctive effect should *not* be possible. Implicatures of ¬KXB/W – e.g. ¬K(John may have beer), ¬K(Some passengers got nauseous) – *directly* contradict it.[7] But paradoxically,

[7]In the case of plural existentials, the exact predictions of such a theory depend on what the plural means. If the truth conditional contribution of *Some passengers got sick/had trouble breathing* is that *2 or more* passengers did,

at the same time it seems that an account of the conjunctive effect in terms of scalar/Quantity implicatures seems to be on the right track – as witnessed by its disappearance in entailment reversing contexts, and by the fact that it can be canceled:

(19) Some students drank beer or wine (I can't remember which).

a. $\rightsquigarrow$The speaker isn't certain that some students drank beer, and the speaker isn't certain that some students drank wine.

(20) John may drink the beer or the wine (I can't remember which).

a. $\rightsquigarrow$The speaker isn't certain that John may drink the beer, and the speaker isn't certain that John may drink the wine.

(We note again that singular exisentials are systematically consistent with (14), and thus lead to no special problem.)

(21) Someone drank the beer or the wine.

a. $\rightsquigarrow$The speaker isn't certain that someone drank the beer, and the speaker isn't certain the someone drank the wine.

There are many logically possible approaches to resolving this apparent paradox. There have been many proposals about the conjunctive effect with possibility modals that are essentially *semantic* in nature (e.g. Higginbotham (1991); Zimmermann (2000); Geurts (2005)), yet do or can in principle still explain the facts noted here that seem to militate in favor of a pragmatic analysis (in terms of scalar implicatures). I won't discuss these interesting proposals in detail here, only noting that they do not extend to cover the conjunctive effect with plural existentials. Rather I'll focus on more recent pragmatic accounts which attempt to resolve the paradox by proposing essentially the following: disjunction is governed by a weaker kind of pragmatic reasoning than what is available to account for (14) under classical (neo-) Gricean assumptions (e.g. Sauerland (2004)).

The idea is to revise some assumptions about how the pragmatics generates (scalar) implicatures, such that both the Generalization (14) respecting 'reading'

implicatures of $\neg$K(Some passengers got sick/had trouble breathing) are consistent with the speaker knowing/it being the case that exactly one student had beer, and exactly one wine. This is the wrong result in general, and in particular does not capture what is expressed by the conjunctively strengthened 'reading' (on which there may be and naturally are many passengers of each type). The problem is even more severe once a more realistic semantics is adopted for plural existentials (Section 4.1).

and the conjunctive effect are available/consistent with the speaker being cooperative (and such that, in general, (14) is attested). An illustrative example is Alonso-Ovalle (2005), who essentially adopts Kratzer & Shimoyama (2002)'s proposal for free choice indefinites in German. The pragmatics of disjunction proposed by Alonso-Ovalle is such that there are two reasons why a cooperative speaker might assert $\Diamond(B \vee W)$: one is that he said exactly as much as he knew with respect to the the permissibility of B and W, i.e. he couldn't make either of the stronger claims $\Diamond B$, $\Diamond W$: $\neg$K$\Diamond$B, and $\neg$K$\Diamond$W. The other, in short, is that he couldn't "choose between" these stronger, simpler claims, in the sense that he they have parallel epistemic status: he knows each of B and W to be permissible, or each not to be.[8] If the speaker is in the latter epistemic position, then given that he believes what he literally asserted ($\Diamond(B \vee W)$), he must know that both A and B **are** permissible, i.e. the conjunctive effect follows.[9] Presumably, context will disambiguate which epistemic state the speaker is in (for which of the reasons the speaker opted out of asserting $\Diamond B/W$). A related approach is developed in Fox (2006), where the basic insight of the Kratzer and Shimoyama/Alonso-Ovalle approach is built into a syntactic/grammaticalized (i.e. non-Gricean) system for generating implicatures. A recursively available syntactic operator, akin to *only*, generates implicatures by associating with scalar terms; one instantiation yields the Generalization (14) respecting strengthened meaning, two stacked instantiations yield the conjunctive strengthening.

The problem faced by both of these accounts is that, having weakened the pragmatics to allow for conjunctive 'readings' of $\Diamond(A \vee B)$ they end up wrongly predicting that *singular* existentials (as well as plurals) should also give rise to them. This is for the simple reason that possibility modals are semantic analogues of singular existentials. For example, under Alonso-Ovalle's assumptions, one of the two cooperative reasons for asserting *Someone drank beer or wine* is – entirely parallel to the modal case – that the speaker couldn't choose between the (stronger, simpler) non-disjunctive statements *Someone drank beer* and *Someone drank wine* (i.e. knows each to be true, or each to be false). From this a conjunctive effect wrongly follows – if believes what he said, he must know them both to be true. In principle this severe problem might be addressed, at least within Fox's formalization of the basic idea, if a further distinction can be drawn between singular existentials on the one hand, and possibility modals

[8]The first reason is in fact not discussed by Alonso-Ovalle (or Kratzer and Shimoyama), but it follows from the assumptions he makes that it is a valid one – which is desirable given the cancelability of the conjunctive effect.

[9]Crucially, the speaker could cooperatively opt out of choosing $\Diamond(B \wedge W)$, since this is even stronger than what he ends up conveying.

and plural existentials on the other, in the nature of the alternatives used in pragmatic reasoning.[10] The proposal developed here starts from the opposite direction, and explores whether understanding the singular/plural distinction can pave the way to a general account of conjunctive effects.

3 Distributivity Implicatures

We propose that the key to understanding free choice permission lies in assimilating it to another puzzle, the fact that plural existentials lead to a parallel conjunctive effect. The key to understanding the latter is the observation that *overtly distributive* plural existential quantification also does:

(22) Some of the students each bought a car or a motorcycle.
[∃X: Students(X)] [∀x: x is-part-of X] C(x) ∨ M(x)

a. ⇝Some of the students (each) bought a car, and some of the students (each) bought a motorcycle.[11]

The conjunctive effect in (22) is every bit as puzzling as it is in the case of plural existentials without overtly realized distributivity; under the generalization (14) we expect implicatures that the speaker doesn't know that the non-disjunctive statements *Some of the students (each) bought a car/motorcycle* are true, but this is precisely the opposite of what is attested.[12] It seems plausible that the effect is somehow tied to distributivity itself, since distributive/universal quantification allows for a similar one, as noted in the previous section:

(23) Each of the students bought a car or a motorcycle.
[∀x: x is-one-of-the-students] C(x) ∨ M(x)

a. ⇝Some of the students (each) bought a car, and some of the stu-

[10]Fox for example suggest that a further (scalar) implicature of singular existentials blocks the conjunctive effect – namely a (scalar) implicature that an identical sentence but with a *plural* existential is false. For example, *Some passenger got sick* is claimed to implicate that it's not the case that two did. I'm not sure whether in general singular existentials give rise to such strong implicatures. Although they do seem to give rise to implicatures that the speaker *doesn't know* (or in some cases, care) whether the corresponding plural sentence is true, the systematic existence of the stronger implicature is crucial for Fox's suggestion to be able to save his approach, as far as I can tell.

[11]Parallel examples in the temporal domain cannot be constructed, since we don't find 'floating' *each* in the adverbial domain. Cf. however an inversely linked partitive such as *On each of several (special) occasions I drank beer or wine with dinner*, where (unsurprisingly) we find a conjunctive effect.

[12]Here again the conjunctive effect can be defeated/fail to arise, vis. the possibility of adding the rider '…but I don't know which.'

dents (each) bought a motorcycle.

Given this similarity, it is possible to understand the conjunctive effect of (22) as owing to exactly the kind of implicature found with (23), but calculated at an *embedded* level, within the scope of the plural existential. Looking at the scope of *some of the students* in (22), we have something which is essentially identical to (23): a universal quantifier with disjunctive scope. Adding the implicatures that such a configuration gives rise to when unembedded, but within the scope of the existential, gives exactly the conjunctive strengthening:

(24) There is a plurality P of students such that: each student in P bought a car or a motorcycle, and some some students in P bought a car, and some students in P bought a motorcycle
'Some of the students bought a car, and some of the students bought a motorcycle'

Accounting for the relevant implicatures of universals over disjunction is, as noted in the previous section, unproblematic on standard (neo-) Gricean accounts. What needs to be assumed is that, among the relevant ALTERNATIVES to $\forall(A \vee B)$ – its 'scalemates' – are $\forall A$ and $\forall B$. (And of course that the algorithm for computing implicatures allows the hearer to conclude that these are both false.) So, for example, for (23), *each of students bought a car* and *each of the students bought a motorcycle* can be concluded false. It follows from the assertion (23) in conjunction with these implicatures that at least one student bought a car (but not motorcycle), and at least one a motorcycle (but not a car). This isn't exactly the paraphrase we've been using for the conjunctive, which was stated in terms of a *plural* existential – *Some student**s** bought a car, and some student**s** bought a motorcycle*, but it is intuitively the correct one. (23) is fine in the (degenerate) case that there were just two students, one who bought a car, and the other a motorcycle, and is decidedly odd in case all the students bought both. It is shown below that the result is also what is wanted (at the embedded level) for (22).

To make the proposal for (22) clear, we adopt a notational system which captures the general spirit of the proposal in Chierchia (2001). Chierchia argued that many paradigm cases of non-truth conditional meaning classically treated as (neo-) Gricean quantity/scalar implicatures arise in *embedded* contexts, in ways that preclude the classical analyses. His conclusion and proposal was that such meanings are derived by essentially a *grammaticalization* of the (neo-)

Gricean system, which works in parallel with semantics, rather than posterior to it. For our purposes, it isn't crucial whether embedded implicatures are derived in Chierchia's 'semantic' way, or by positing *syntactic* operators that mimic the effect of (neo-) Gricean implicature computation Fox (2006) (a mixed theory is given Chierchia (2005)). For simplicity we adopt a highly simplified syntactic version. We crucially depart from Chierchia and classical neo-Gricean accounts in taking the competitors/scalar alternative to *X(A* $\vee$ *B)* to include not only *X(A* $\wedge$ *B)*, but also *XA* and *XB*, following Sauerland (2004):

(25) a. Let STRONG(*X*) stand for the neo-Gricean strengthening of *X*: i.e. X conjoined with the negation of its stronger neo-Gricean scale-mates.

b. STRONG($\forall$x: Ax $\vee$ Bx) is thus equivalent to

(i) ($\forall$x: Ax $\vee$ Bx) $\wedge$ $\neg$($\forall$x: Ax) $\wedge$ $\neg$($\forall$x: Bx) $\wedge$ $\neg$($\forall$x: Ax $\wedge$ Bx)

(ii) =($\forall$x: Ax $\vee$ Bx) $\wedge$ ($\exists$x: Ax $\wedge$ $\neg$Bx) $\wedge$ ($\exists$x: Bx $\wedge$ $\neg$Ax)

(22) (repeated) can now have the representation in (26a):

(26) Some of the students each bought a car or a motorcycle.

a. [$\exists$X: Students(X)] STRONG([$\forall$x: x is-part-of X] Cx $\vee$ Mx)

b. =[$\exists$X: Students(X)] ([$\forall$x: x is-part-of X] Cx $\vee$ Mx) $\wedge$ $\neg$([$\forall$x: x is-part-of X] Cx) $\wedge$ $\neg$([$\forall$x: x is-part-of X] Mx) $\wedge$ $\neg$([$\forall$x: x is-part-of X] Cx $\wedge$ Mx)

c. =[$\exists$X: Students(X)] ([$\forall$x: x is-part-of X] Cx $\vee$ Mx) $\wedge$
([$\exists$x: x is-part-of X] Cx $\wedge$ $\neg$Mx) $\wedge$ ([$\exists$x: x is-part-of X] Mx $\wedge$ $\neg$Cx)

d. 'At least one of the students bought a car (but not a motorcycle), and at least one of the students bought a motorcycle (but not a car).'

As noted above and indicated in the paraphrase (26d), the actual result is not quite equivalent to what we've been assuming as a paraphrase of the conjunctive strengthening. One difference is that the strengthened meaning derived doesn't require that there be multiple motorcycle buyers or multiple car buyers – just one or more of each, in contrast with the paraphrase we've been working with, which used a plural existential: 'Some of the students (each) bought a car, and some of the students (each) bought a motorcycle. This seems to be exactly what

is wanted: if there are three students who made a vehicle purchase, one of a motorcycle, two of a car, there's nothing strange about using (22) – so long as there was no need to be more precise, of course. In the further degenerate case in which there was just one car buyer, and just one motorcycle buyer (among the students), the account predicts (26) to have exactly the status of e.g. *Some students (each) bought a car*, in a context in which the speaker knows (and it is relevant that) exactly two students did, and this seems to be correct. There is a general tendency, it seems, for plural existentials to suggest vagueness in number, or to be odd where the number of witnesses hovers at barely plural (i.e. 2).[13] There is likely a pragmatic explanation for this fact, but all that matters for present purposes is that its existence means that the proposed analysis of the conjunctive effect has no special problem with the 'degenerate' case.

The second difference is that the derived strengthened meaning strictly requires that there be at least one student who bought a motorcycle but *not* a car, and at least one who bought a car but *not* a motorcycle. Again, this is intuitively correct, and the facts are entirely parallel to the pure universal case ((23)) – as predicted the by the account.

The account for (22) extends straightforwardly to plural existential cases like (3), given the assumption that the embedded implicature can be calculated just as well on a distributive operator if it doesn't happen to be pronounced. Evidence that non-overt operators can trigger implictatures like their overt counterparts can be found with existential readings for bare plurals, for example. *Alex saw girls from his section at the dance* implicates, just like *Alex saw some girls from his section at the dance*, that he didn't see every girl from his section.

(27) Some passengers got nauseous or had trouble breathing.

a. $[\exists X\text{: } P(X)]$ STRONG$([\forall x\text{: } x \text{ is-part-of } X]\ Nx \vee Tx)$

b. $=[\exists X\text{: } P(X)]\ ([\forall x\text{: } x \text{ is-part-of } X]\ Nx \vee Tx) \wedge \neg([\forall x\text{: } x \text{ is-part-of } X]\ Nx) \wedge$
$\neg([\forall x\text{: } x \text{ is-part-of } X]\ Tx) \wedge \neg([\forall x\text{: } x \text{ is-part-of } X]\ Nx \wedge Tx)$

c. $=[\exists X\text{: } P(X)]\ ([\forall x\text{: } x \text{ is-part-of } X]\ Nx \vee Tx) \wedge ([\exists x\text{: } x \text{ is-part-of } X]\ Nx \wedge \neg Tx) \wedge ([\exists x\text{: } x \text{ is-part-of } X]\ Tx \wedge \neg Nx)$

d. 'At least one passenger got nauseous (but had no trouble breathing), and at least one passenger had trouble breathing (but didn't get nauseous).'

[13]This explains why the the plural paraphrase for the conjunctive effect that we've been using up until now seemed intuitively correct – we continue to use it since it is appropriate in all but exceptional cases.

It should now be clear that under the present proposal, the reason that a conjunctive effect is unavailable for *singular* existentials is simply that no distributive operator is present.

The claim that distributivity is crucial to the conjunctive strengthening of (22) is independent of the particular account just given. Embedded implicature is a purely descriptive term, and as has been shown in other domains (Spector (2003), Sauerland (2004)), sometimes what the existence of embedded implicature shows us is that we didn't understand a particular (neo-) Gricean mechanism sufficiently, rather than that a (radically) non-Gricean approach is needed. Distributivity could be implicated in a number of ways. Do we find independent evidence that it *is* crucially involved?

An obvious question is whether we find the conjunctive effect with disjunctions of *collective* predicates. Naively one might expect that our proposal predicts that the answer should be no. Matters are complicated, however, since it is well known that there are 'intermediate' distributive readings – cases in which collective predicates are applied distributively to parts of a plurality, obviously non-atomic ones ('sub-pluralities') (Schwarzschild, 1996):

(28) The boys gathered in the hall
→can be true if the totality of boys did not (all) gather (together), but rather gathered into groups

It does seem to be the case the no conjunctive effect arises where the nuclear scope is a disjunction of collective predicates which themselves don't allow for any kind of distributive reading – compare (29) and (30). (The following context may be useful: a student, or group of students, asks what he/they should do for his art project...)

(29) (Some students wrote a poem or composed a song.
a. ⇝'Some students wrote a poem, and some students composed a song.'

(30) Some students wrote this poem or composed this song.
a. #'Some students wrote this poem, and some students composed this song.'

A predicate like *write this poem* or *compose this song* allows for no kind of distributive reading – if it applies to any plurality at all, it applies to exactly one and to none of its parts, whether atomic or plural (i.e. sub-pluralities). Still,

given the existence of intermediate distributive readings, it might seem puzzling from the perspective of our proposal that the conjunctive effect is absent in (30). A plurality of students composed of one group who wrote the poem, and one the song, would satisfy the proposed strong meaning if distributed down into just those groups. There are two possibilities. One could take these facts to show that the introduction of an (intermediate) distributive operator is not entirely free, but rather hinges on the properties of the predicate. If this is correct we have an argument that distributivity is crucial to the conjunctive effect. Although these considerations are inconclusive, there turn out to be stronger – but much more involved and indirect – ways of establishing the claim that distributivity is crucial; see Klinedinst (2006).

A further obvious question is whether there is independent evidence for an account in terms of embedded implicature of a genuinely non-Gricean variety, i.e. of a purely formal/grammatical one. The answer depends in large part on what properties are conceptually necessary to the latter type of account. It does seem that there is an intuition that the conjunctive effect enters as part of the content *asserted*, rather than as an inference (as reported in Simons 2005, who backs up the intuition with empirical evidence that is behaves differently that other types of Gricean inferences). If this is correct, it provides at least weak support for the type of account proposed here – since embedded implicatures per definition figure into the level of asserted content.[14]

4 Pluralities and Possibilities

Taking possibility modals to express plural existential quantification over possible worlds, an account of free choice permission can be given which is entirely parallel to that developed in the last section. The only further assumption required is one that is already completely standard: intensions are taken to be functions from *atomic* entities (possible worlds) to extensions. Distribution is therefore necessary in the scope of possibility modals, since these quantify over

[14]The matter is in actuality a bit more complicated. For reasons that will become clear when Chierchia's system is discussed, any theory of embedded implicature must have the ability to 'factor out', at the global level, implicatures added in embedded contexts. This means that the intuition that embedded implicatures are part of asserted content doesn't strictly come for free. In addition, in a theory like Chierchia's all scalar implicatures – even ones for which there is no such intuition – are generated by the same mechanism, and so it has to draw some further distinction. Thus it cannot even be stipulated that intuitions of asserted content track directly the output of this mechanism. I think that there is a principled way to draw the needed distinction, however, and in any case there remains point in favor against a global Gricean account: whatever its derivation of the conjunctive effect, it will be external to semantic content by definition.

pluralities, and so a simple modal statement like *John may have beer or wine* will translate as follows:

(31) $[\exists W: Acc_D(W)]\ [\forall w: w \text{ is-part-of } W]\ Bw \vee Ww$

'Acc_D' is to be understood distributively, i.e. as holding of pluralities of worlds each of which are deontically accessible.[15] (31) is exactly parallel in structure to the non-modal cases analyzed in the previous section, and the account applies in turn. Calculating an (embedded) distributivity implicature derives the crucial facts of free choice permission:

(32) a. $[\exists W: Acc_D(W)]$ STRONG$([\forall w: w \text{ is-part-of } W]\ Bw \vee Ww)$

b. $=[\exists W: Acc_D(W)]\ ([\forall w: w \text{ is-part-of } W]\ Bw \vee Ww) \wedge \neg([\forall w: w \text{ is-part-of } W]\ Bw) \wedge \neg([\forall w: w \text{ is-part-of } W]\ Ww) \wedge \neg([\forall w: w \text{ is-part-of } W]\ Bw \wedge Ww)$

c. $=[\exists W: Acc_D(W)]\ ([\forall w: w \text{ is-part-of } W]\ Bw \vee Ww) \wedge ([\exists w: w \text{ is-part-of } W]\ Bw \wedge \neg Ww) \wedge ([\exists w: w \text{ is-part-of } W]\ Bw \wedge \neg Ww)$

d. 'There is a plurality of worlds consistent with the rules which includes at least one world in which John drinks beer but not wine, and at least one world in which he drinks wine but not beer'

A more colloquial paraphrase: (1) John's options include the following: drinking beer (without drinking wine), and (2) drinking wine (without drinking beer). Notice that this means that John is not **required** to drink either – which seems to be empirically correct. There is however a further implicature that *John may have beer or wine* often carries, which is not yet derived: that John is not allowed to have both ($\neg\Diamond(B \wedge W)$).

$\Diamond(B \wedge W)$ fails to entail (32a); so the question arises how the negation of the former is to derived, given the working the assumptions adopted here, which effectively make the latter the actual semantic content of *John may have beer or wine* (on a given occasion of use). What needs to be assumed is that embedded implicatures are effectively ignored for the purposes of implicature calculation at higher levels; in this case the alternative $\Diamond(B \wedge W)$ needs to be compared to *John may have beer or wine* on its 'normal' meaning, $[\exists W: Acc_D(W)]\ [\forall w: w$

[15]Lurking behind this terminological point is an important question. Given the use of pluralities of possible worlds, and distribution over them, we can fairly ask whether there are natural language expressions/structures which express *collective* properties of pluralities of worlds (and what exactly this would mean). This interesting question is largely independent of the proposal, as long as the disjunctive complements of possibility modals that allow for conjunctive strengthening *can* have as meanings functions defined on atomic worlds. We have no reason to think that they can't.

is-part-of W] B(w) ∨ W(w), rather than (32a), where the local implicature has been added. The syntactic framework adopted here for expositional purposes would need to to be enriched to derive this result. Importantly, though, this is an independently crucial property of any theory of embedded implicature.

The (non-syntactic) system of implicature calculation in Chierchia (2001), for example, crucially works in this way. Chierchia is able to derive the correct implicatures for multiple disjunctions *(A ∨ (B ∨ C))*, namely that exactly one of A, B, C is the case. The way the system works is to calculate an implicaure for *B ∨ C* at the embedded level (in the scope of the higher disjunction), and add it to the normal meaning of *B ∨ C*, but effectively *remember* the latter, to be accessed in the computation of implicatures at higher levels. In effect, the rules encode that the overall implicature-strengthened meaning of *A ∨ (B ∨ C)* is obtained by adding the exclusivity implicature for *B ∨ C* locally *((B ∨C) ∧ ¬(B ∧ C)*, *=B* $\vee_{excl}$ *C)*, passing this through the computation to obtain *A ∨ (B* $\vee_{excl}$ *C)*, and then computing and adding to it a 'global' implicature yielding an exclusive interpretation for the higher disjunction, but based on the normal meaning of *B ∨ C*: *¬(A ∧ (B ∨ C))*. The resulting conjunction *(A ∨ (B* $\vee_{excl}$ *C)) ∧ ¬(A ∧ (B ∨ C))* is exactly the desired strong meaning – it is true iff exactly one of A, B, and C holds. What is crucial is that the wrong result would have been obtained by calculating implicatures for the higher disjunction taking into account the strengthened meaning of *B ∨ C* itself: *(A ∨ (B* $\vee_{excl}$ *C)) ∧ ¬(A ∧ (B* $\vee_{excl}$ *C))* is emphatically not strong enough, since it is consistent with each of A, B, and C obtaining. This 'memory' property of Chierchia's system is what distinguishes it from a theory that posits systematic ambiguity in scalar terms – which notoriously fails for multiple disjunctions – and is thus one that any theory of embedded implicature has to possess, one way or another

The desired total strengthened meaning for *John may have beer or wine*, under the plural analysis proposed here, can be derived in a modification of his system that imports the crucial feature of the analysis: that implicatures can be computed and added at the level of the embedded distributivity operator, based on comparison to the meanings of the distributivity operator applied to each of the disjuncts separately. Chierchia's algorithm, as shown by the example of multiple disjunction, provides for the possibility that this local distributivity implicature be ignored for the purposes of computing further implicatures. The derivation in our case is entirely parallel: the local distributivity implicature is calculated and added to the content (yielding free choice permission), but then ignored at the global level, where a further implicature to ¬◇(A ∧ B) is

calculated by *or-and* comparison, based on the normal meaning of *John may have beer or wine* (i.e. its meaning without the embedded implicature added).[16] See Klinedinst (2006) for a full formal implementation.

4.1 Negation and Plurality

A desirable feature of the standard semantics for possibility modals is that it captures the logical equivalence of $\neg\Diamond$P and $\Box\neg$P: you may not have beer if and only if you must not have beer. Interestingly, this equivalence is exactly what is expected on the assumption that possibility modals are in fact plural existentials, despite a possible impression to the contrary. (Doesn't 'there is not a plurality of worlds in which P' just mean that there are less than two – and not necessarily that there are none (that in all worlds not P)?) The crucial observation is that plurals behave systematically like singulars in downward entailing contexts (in the scope of decreasing functions):

(33) Alex doesn't have any friends (in Berlin).
   a. $\neq$Alex doesn't have two or more friends (in Berlin).
   b. $\approx$'There isn't anyone in Berlin who is a friend of Alex.'

(34) No students came to the party.
   a. $\neq$No group of two or more students came to the party.
   b. $\approx$'No student came to the party.'

In recent work by Spector (2005) and Anderson et al. (2005), it is argued on the basis of such facts that the plural has a weak semantics, such that plural variables range over entities which must simply contain at least one atomic part. The 'true' plurality conveyed in non-DE contexts – the fact that *Some passengers got sick* 'means that' at least two did – is argued to be derived as a pragmatic effect of competition with singular forms. Under both theories, the effect is predicted to disappear in DE contexts, so that (33), for example, conveys just its (desired) literal meaning – namely, that there is no group of one or more individuals which are friends of Alex.

Adopting this assumption, *John may not have beer* is represented as follows, where 'X' ranges over objects consisting of one or more atomic individuals, and we now understand the relation 'is-part-of' as extending to hold between

[16] Chierchia explicitly supposes that implicatures can be calculated based on a given scalar item only once (so not both globally and locally). Crucially, in our case two implicatures must be calculated based on the *or-and* scale: one locally, one globally. Mechanically, Chierchia's system *does* in fact allow for this possibility, as far as I can tell.

an atomic individual and the 'degenerate' plurality consisting of just that individual[17]

(35) $\neg$[$\exists$W: Acc$_D$(W)] ([$\forall$w: w is-part-of W] B(w)

a. 'There is no group of one or more accessible worlds, such that each world in it is a world in which John drinks beer (i.e. there neither one nor more than one accessible B world).'

Having adopted these more realistic assumptions about plurality, (35) is equivalent to $\Box\neg$B, as desired. I assume that, unlike the case of plurals in the individual domain, there is no inference to 'true' plurality based on competition/comparison: but the existence of a more than one accessible world, if there is any, will follow automatically from the vastness of the space of possibilities.[18]

A crucial observation we began with is that conjunctive strengthenings strongly tend to disappear in DE contexts: embed *John may have beer or wine* under negation, and the overwhelmingly natural reading is the negation of its standardly predicted meaning, not the negation of the free choice permission reading. This will remain a result of the present account, since any theory that generates embedded implicature has to derive in one way or another that they cannot be retained in DE contexts; cf.

(36) John didn't invite Bill or Mary or Susan

a. $=\neg(A \vee (B \vee C))$

b. $\neq\neg((A \vee (B \vee_{excl} C)) \wedge \neg(A \wedge (B \vee C))$
'It's not the case that exactly one of A, B, and C is true'

Since the distributivity implicature is calculated in an upward entailing context, the (immediate) scope of an existential, embedding this entire constellation in a DE context would then force the implicature to be calculated in a DE context, precisely the phenomenon which must independently be blocked. Taking the syntactic proposal entertained here seriously, this could possibly be derived as a distributional restriction on the implicature operator STRONG. In the system

[17]If the domain of plural objects is taken to be constructed by an operation of sum formation, such that plural objects are elements of join-semilattices, this amounts to saying that the individual John is part of the individual John. If it is taken to be sets, it just amounts to saying that John is a member of the set containing John. Neither claim is objectionable.

[18]Of course, in the case of free choice permission, the calculation of distributivity implicatures 'forces' there to be more than one world in the plurality quantified over, but this fact alone is uninteresting – the work done is to force there to be worlds of both both types (i.e. permissible beer drinking worlds and permissible wine drinking worlds, in the case of *John may have beer or wine*).

in Chierchia (2001), it is simply a built in property of the recursive semantic rules which calculate strengthened meanings. In sum, the proposal introduces no new complications to the theory of embedded implicature.

5 Conclusion

Given the considerations about plural meaning in the last section, the claim that possibility modals are plurals doesn't amount to anything radical – in fact it yields a semantics which is equivalent to the old one, once distributivity is taken into account.[19] But it is far from a vacuous refinement, as the present proposal shows: a unified account of two puzzles, an old one about modals, and a new one about plurals, becomes available within an existing general framework for calculating implicatures – something which has the core properties of the system in Chierchia (2001).

The core of the proposal is that **distributivity** is at base responsible for conjunctive strengthenings of existentials with disjunctive nuclear scopes – in both the modal and non-modal domains. Importantly, this proposal may remain intact and interesting even under a more sophisticated global approach to implicature. What would ensure that it will be is the existence of independent evidence that possibility modals are plural. There are some interesting considerations.

First we have some basic considerations of plausibility (to be clear: not an argument). *Any* (useful) statement involving existential quantification over possible worlds will by practical necessity have a (huge) plurality of witnesses to it, a simple consequence of the vastness of the space of possibilities. An absurd situation would obtain have to obtain for things to be otherwise. Suppose for example that there is **exactly one world** which is witness to (37) on a deontic or epistemic interpretation for the modal:

(37) Jenny may smoke

a. $\Diamond$S ([$\exists$w: Acc(w)] S(w)])

It would follow that:

(38) $\forall P(\Box(S \Rightarrow P) \vee \Box(S \Rightarrow \neg P))$

Which is to say that the law, or what is known, would completely determine the

[19] Again, in principle there could be different predictions if there are the equivalent of collective readings in the modal domain.

conditions under which S (under which Jenny smokes). In the epistemic case this means that discovering that Jenny smokes – suppose she actually does – is the one thing in the way of the speaker and total omniscience. In the deontic case it amounts to the law specifying in impossibly fine detail the conditions under which Jenny can smoke – do her toes have to be crossed or uncrossed? Can her mother have recently gotten an anchor tatoo? On which forearm?, etc. The reason is the elementary fact that for any single world, e.g. one compatible with what is known or required, a proposition is either true of/in it, or not. Adding a Kratzer style ordering source doesn't change this point: (38) would still follow from (37) if there were exactly one S world among the 'best' accessible worlds. Neither would using situations instead of worlds change anything. Any situation which can reasonably count as one in which Jenny smokes, no matter how 'minimal', will include an infinite number of details (the way her lips are pursed, for example) irrelevant to what the law says about Jenny smoking, and potentially underdetermined by our knowing that she does.

Stronger considerations come from the domain of cross-sentential anaphora, where we find that possibility modals pattern with true plural existentials, and against synonymous morphologically singular existentials, in their possibilities for anteceding pronouns. (We assume without argument here that 'would' (39) is anaphoric to the witness world(s) to the preceding existential modal statement (see e.g. Stone 1999)).

(39) (Don't smoke.) My sister might come in.
  a. She **would** (probably/definitely) kill us.
  b. . . . '(It is probable/certain) given that my sister comes in, she kills us.'

(40) At least one musician will come in. . .
  a. . . . He/#They will be female.
  b. . . . #Most/all of them will be female.

(41) One or more musicians will come in.
  a. . . . They/#he will be female.
  b. . . . Most/all of them will be female.

The generalization seems to be that plural indefinites require plural pronouns for cross sentential anaphora, while morphologically singular indefinites require singular pronouns. Given that 'at least one musician' and 'one or more musicians' are synonymous for all relevant purposes, and (crucially) that they don't

have different, relevant implicatures, the pattern they exhibit must be purely 'grammatical'. When anaphoric *would* restricts a modal adverb like *definitely* or *probably* as in (39a), it seems plausible that it must have plural reference (Stone (1999)) – at pain of collapsing the truth conditions in the two cases. Thus, it effectively follows from the generalization that emerges from (40)-(41) that possibility modals – e.g. *might* in (39)-(39a) – are plural. If *would* must have plural reference, and there is a purely grammatical requirement on antecedence, then *might* cannot be singular: either both *would* and *might* are unmarked, or both plural. But the two possibilities are essentially equivalent given the weak meaning we've assumed for plurals.

There is a final argument of a slightly more conceptual nature. Given that modals express generalized quantification over possible worlds, the question arises of why we only find modal (auxiliaries) with *existential* and *universal* force – in contrast, for example, with what is found in the individual domain. This contrast doesn't follow in general from a simple difference in the nature of the domains of modal and individual quantification, since there are ways of expressing non-universal and non-existential quantifications over worlds, periphrastically (e.g. *it is not the case that you may...*, compare *no* in the individual domain) and/or with expressions that aren't modal in the lexical-syntactic sense (e.g., *it is impossible that*, arguably *it is (50%) likely that*). Taking possibility modals as plurals raises the possibility to bridge the explanatory gap: for example, plurality can be taken to be an underlying feature of all modals, with a truly binary lexical distinction between indefinite (possibility) and definite (necessity) replacing the (more) stipulative existential/universal distinction.

It is my hope that these preliminary arguments can be fleshed out in further detail, to bolster the intuitive plausibility of the claim that possibility modals are plurals. I leave this for future research.

References

Alonso-Ovalle, L. (2005). Distributing the disjuncts over the modal space. In Bateman, L. & Ussery, C. (eds.), *Proceedings of nels 35*. GLSA, University of Massachusetts Amherst.

Anderson, J., Sauerland, U. & Yatsuhiro, K. (2005). The plural involves comparison. In Kesper, S. & Reis, M. (eds.), *Linguistic evidence*, Mouton de Gruyter.

Chierchia, G. (2001). Scalar implicatures, polarity phenomean, and the syntax/pragmatics interface. Ms., University of Milan-Bicocca.

Chierchia, G. (2005). Broaden your views: implicatures of domain widening and the "logicality" of language. Ms., University of Milan-Bicocca.

Fox, D. (2006). Free choice and the theory of scalar implicatures. Ms., MIT.

Geurts, B. (2005). Entertaining alternatives: disjunctions as modals. *Natural Language Semantics* **(to appear)**.

Higginbotham, J. (1991). Either/or. In *Proceedings of nels 21*. GLSA, University of Massachusetts Amherst.

Klinedinst, N. (2006). *Plurality and possibility*. Ph.D. dissertation, UCLA.

Kratzer, A. & Shimoyama, J. (2002). Indeterminate phrases: The view from japanese. In Otsu, Y. (ed.), *The proceedings of the third tokyo conference on psycholinguistics*. Tokyo: Hituzi Syobo. 1–15.

Ross, A. (1941). Imperatives and logic. *Theoria* **25**. 53–71.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy* **27**. 367–391.

Schwarzschild, R. (1996). *Pluralities*. Kluwer Academic Publishers.

Simons, M. (2005). Semantics and pragmatics in the interpretation of 'or'. In Georgala, E. & Howell, J. (eds.), *Proceedings of semantics and linguistic theory xv*. Cornell University Department of Linguistics, CLC Publications.

Spector, B. (2003). Scalar implicatures: exhaustivity and Gricean reasoning. In ten Cate, B. (ed.), *The proceedings of the eighth esslli student session*. Vienna. 277–288.

Spector, B. (2005). *(in progress)*. Ph.D. dissertation, Université Paris 7.

Stone, M. (1999). Reference to possible worlds (RuCCS TR 49). Technical report, Rutgers University.

Szabolcsi, A. (2004). Positive polarity-negative polarity. *Natural Language and Linguistic Theory* **22**.

von Fintel, K. (2006). Modality and language. In Borchert, D. M. (ed.), *Encyclopedia of philosophy – second edition*, Detroit: MacMillan Reference USA. http://mit.edu/fintel/www/modality.pdf.

Zimmermann, T. E. (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics* **8**. 255–290.

*Logic in Pragmatics**

Hiroyuki Uchida

Abstract

This paper argues against the complete elimination of logical introduction rules from the pragmatic inference system. To maintain the consistency of the inference system as a whole, which is meant to support one's truth-based judgment over propositions, the inference system should have access to both introduction and elimination rules. I show that the inclusion of introduction rules in the pragmatic inference system neither overgenerates propositions expressed nor cause non-terminating inference steps.

1 Free enrichment and alleged overgeneration

According to Relevance Theory (RT for short, Sperber & Wilson 1986/95), reference assignment and disambiguation are not the only pragmatic processes involved in the recovery of propositions expressed by (or the truth-conditional content of) utterances.

(1) a. Every presenter [*in the pragmatics session of CamLING07*] was impressive.
b. John took out a key and opened the door [*with the key*]. Cf. Hall (2006)

RT assumes that, given the linguistically provided information outside the square brackets in (1), the hearer can pragmatically add the contents given in the brackets when she recovers the proposition expressed (in appropriate contexts).[1]

* I am grateful to Robyn Carston, Nicholas Allott and Alison Hall for reading (part of) drafts and making suggestions. They do not necessarily agree with the final version and all the mistakes and misconceptions in this paper are the author's.

[1] More accurately, given the linguistic meaning outside the square brackets in (1) which the hearer can recover by decoding the semantic information encoded with the language expressions that the speaker used in the utterance, and given the context in which the utterance was made (which includes the speaker's intention), the hearer can pragmatically enrich to the truth-conditional content of the utterance which includes the pragmatically added contents inside the square brackets in (1). Crucially, the hearer may enrich the meaning of an expression before she recovers the encoded meanings of the other expressions. For more details about free enrichment, see Carston (2002).

Stanley (2002) claims that this free enrichment overgenerates. Suppose the sentence in (2) is uttered in a context in which the propositions in (3) are available as contextual premises.[2] Then, according to Stanley, RT wrongly predicts that the meaning of (2) can be enriched by conjoining it with the contextual proposition (3), deriving the proposition expressed in (4). Stanley himself does not literally identify this process as &Introduction, but for the purpose of this paper, let me identify this process as &I applied to (2) and (3a).[3] The recovery of (4) in this way would give the hearer enough cognitive effect, because an application of MPP between another contextual premise (3b) and (4) would lead to the relevant conclusion *John will not live long* (= *R*).

(2) John smokes. (= *P*)
(3) a. John drinks. (= *Q*)
b. If John smokes and drinks, he will not live long. (= (P&Q)→R)
(4) John smokes [*and drinks*]. (= *P&Q*)

Addressing this criticism, Hall (2006: 95-96) follows Sperber & Wilson (1986/1995) and postulates Conjunctive Modus (Ponendo) Ponens (CMPP) as in (5). With CMPP, one can derive the relevant conclusion, *John will not live long* (= *R*), without applying &I.

(5) Conjunctive Modus Ponens:

1. $(P\&Q)\rightarrow R$	Premise 1
2. P	Premise 2
3. $Q\rightarrow R$	1, 2, CMPP
4. Q	Premise 3
5. R	1, 2, 4, MPP

In her solution to the alleged overgeneration problem, Hall suggests a weaker claim that because of CMPP, the hearer does not have to use &I in order to derive the relevant conclusion *John will not live long*. Thus, the hearer *can* derive this

[2] My example sentences are overloaded with different kinds of information. They may represent sentences uttered, linguistic meanings or propositions, where propositions may be trivial, expressed or contextual (though in my views, trivial 'propositions' do not acquire propositional status until they are recognized as propositions expressed (but then they stop being trivial anymore). For example, (2) may represent the sentence uttered, but when it is identified with *P*, it represents a proposition.

[3] I make this assumption because the main argument of this paper is that inclusion of &I in the spontaneous inference system at the basic level neither leads to overgeneration with enrichment nor leads to infinite inferences. Reviewing what Stanley really meant in his overgeneration claim and arguing against it is not part of this paper's aim.

conclusion as in (5), without deriving the undesirable (4) as the proposition expressed by (2).

However, with this weaker claim, to prohibit the derivation of (4) as the truth condition of (2) in *any* instance of interpreting (2) in context, one would need some additional explanation why the hearer *always* uses the inference steps as in (5) rather than the application of &I followed by MPP, when free enrichment is involved.[4]

In this paper, I argue that introduction rules can be used in pragmatic inferences in general. Thus, after showing that it is problematic to eliminate the &Introduction rule from the pragmatic inference system in section 2, I provide an explanation about why &I is not used in enrichment in section 3, though the pragmatic inference system itself is equipped with this rule. I also argue that CMPP is only a convenient shorthand for a particular combination of inference steps, rather than an actual inference rule defined over logical connectives.

The stronger interpretation of RT's proposal[5] is that spontaneous inferences do not use (logical) introduction rules at all (and thus, (5) is the only way of deriving the conclusion *R*, given (2)~(3)). The reason for postulating this stronger hypothesis is not only the alleged overgeneration of propositions expressed by way of free enrichment. Sperber & Wilson, among others, argue that spontaneous inference should not have access to introduction rules because, otherwise, the system would generate infinite or non-terminating inferences. In section 4, I briefly explain this infinity problem and then show that the problem is not caused by the use of introduction rules in the system, and thus eliminating &I or other introduction rules is not the right way of coping with this problem. Section 5 shows some proofs to support my arguments. Section 6 deals with some loose ends and comments about use of logic in pragmatics from a general viewpoint. Section 7 provides concluding remarks.

This paper is based on certain theoretical assumptions. When we say that an inference system is incomplete with regard to the intended semantics, the 'intended semantics' does not mean the semantics of the inference data that the system aims to explain. It means the system-internal semantics that the person who proposes the system must define or provide for the language representations that are manipulated

---

[4] For some explanation that Hall suggests about why CMPP is preferred to &I followed by MPP, see Hall (2006: 96).

[5] Sperber & Wilson explicitly rule out introduction rules from the inference system. Thus, strictly speaking, what I call the stronger claim is the only possible interpretation of their proposal. On the other hand, they carefully avoid any modification of the logical system which may potentially underlie their inference system. Thus, it is theoretically possible to interpret their claim as the weaker one, in which we can use & with restrictions for purely application reasons, such as efficiency of inferential processes, even if that is not what Sperber & Wilson had in mind as a possibility.

in the system. When a deductive system is proposed, both the syntactic rules that define the well-formed syntactic objects (i.e. propositions in propositional logic) and the syntactic inference rules which operate over those syntactic objects (such as MPP, &I, &E etc.) must be presented, but that is not enough. It must also be specified how those syntactic objects and inference rules are intended to be interpreted. Moreover, such intended 'denotation' of syntactic objects and rules must be modelled in a well-defined semantic structure, such as the Boolean lattice for the classical propositional logic.[6] The 'completeness' of an inference system with regard to the intended semantic is a matter of checking (or 'proving') whether the syntax and the semantics match up completely in the proposed system, where both of these are system-internal concepts. Let me elaborate a little on this point with informal schemas.

(6) a. Syntax: $\{\ldots,\phi_1,\ldots,\phi_n,\ldots\} \vdash_{\mathsf{CLP}} \{\ldots,\varphi,\ldots\}$

b. Syntax simplified: $\phi_1,\ldots,\phi_n \vdash_{\mathsf{CLP}} \varphi$

c. Semantics: $\|\phi_1\|^M,\ldots,\|\phi_1\|^M \models \|\varphi\|^M$

In classical propositional logic, the syntax derives a set of propositions from a set of propositions, as shown in (6a) (CLP abbreviates Classical Propositional Logic). For the sake of simplicity, however, let me discuss the syntactic derivability as if we derived a proposition (rather than a set of propositions) from a set of propositions, by getting rid of 'irrelevant propositions' (indicated by … in (6a)) which do not play an essential role in the inference that we discuss at each stage.[7] That is, as shown in (6b), the proposition φ is syntactically derivable from propositions $\phi_1,\ldots,\phi_n$ (I also omit the set notation, $\{\cdot\}$, in the Antecedent to the left of $\vdash$ as well). This syntactic derivation is solely dependent on the set of syntactic

[6] The word 'denotation' may be misleading to linguists, because linguists tend to assume that denotations are some concrete objects in the world, but that is not necessarily the case. Functions from possible worlds to sets of individuals, for example, might be denotations of some logical expressions (such as predicates). Whatever one assigns as the (intended) interpretations of syntactic objects count as 'denotations.'

[7] As we see in section 4, a problem of the 'impure' definition of $\vee$I is that as well as it introduces the truth functional connective, it recovers one of these 'irrelevant propositions' from the background (i.e., $\{\ldots\}$) and put it in a noticeable place. This latter operation is structural weakening in the Succedent and because structural weakening exists independently $\vee$I, solving the (infinity) problem by controlling the application of (impure) $\vee$I is not only in the wrong track, it does not completely solve the problem, either, as we see in section 4. The same applies to &I in the Antecedent side with regard to the 'impure' left rule in Gentzen sequent presentation, as in the inference from $p \vdash p$ to $p\&q \vdash p$, which has incorporated structural weakening in the Antecedent. In contrast, the pure &I in (16a) abstracts away from the structural weakening. See section 5.

derivation rules which include &I, &Elimination, MPP (= →Elimination) etc. Now, in order for the system to be used in application, defining such syntactic rules is not enough. We have to define the semantic interpretation rule that interpret both the syntactic objects (such as $\phi_1,\ldots, \phi_n, \varphi$) and the syntactic derivability relation $\vdash$. Let $\models$ represent the interpretation of the syntactic derivability $\vdash$. This semantic relation $\models$ is harder to explain without introducing formal details, and I only make an informal presentation.[8] It informally means that for all the models M in which $\|\phi_1\|^M,\ldots, \|\phi_1\|^M$ are all 1 (or True), $\|\varphi\|^M$ is also 1. This semantic computation is based on the standard interpretation of the logical connectives, &, $\vee$ and $\rightarrow$ in terms of the truth tables. Or one may use the equivalent truth-condition presentations for the connectives, such as, for all models, M, $\|\phi_1\&\phi_2\|^M=1$ if and only if $\|\phi_1\|^M=1$ and $\|\phi_2\|^M=1$ as the semantics of &, etc. Now, because this semantic relation $\models$ applies generally, independent of the verdict of the syntax, if we assume that *p, q* and *p&q* are all well-formed formulas in the language that the inference system uses, then the semantics validates the argument, $\|p\|, \|q\| \models \|p\&q\|$ (the reader can check the validity by drawing truth tables, for example), even if we eliminates &I from the syntax and we can no longer 'syntactically' derive the sequent $p, q \vdash p\&q$. Thus, the truth-based semantics above would validate an argument that the syntactic system without &I (but which still makes use of the form $\phi_1\&\phi_2$ as a well-formed formula and which still uses &E and all the other elimination rules) can no longer support. This syntax would then be incomplete with regard to the suggested truth-based semantics.

The main part of this paper is just an elaboration of this incompleteness argument against the syntax without &I, relative to the 'truth-based' semantics as was sketched above, but let me concentrate on the formal status of the semantics I have just sketched. As I said above, this semantics is independent of the syntactic rules (that is why some arguments may be validated without the syntax being unable to support them). On the other hand, it is still part of the language system in two ways. First, any logical language without the provision of such a formal semantic structure as the one given above is not complete as a system. Without the intended semantics, the syntactic objects and rules may potentially be interpreted in different ways and thus the proposed syntactic system cannot be rigorously evaluated in terms of what it can do in application. Second, it is the intended semantics that is comparable to the data. The impression that one can compare the syntactic rules

[8] Wansing (1993), among others, interprets a proposition, *p*, as the set of information states in which *p* is true. Then, given additional rules to interpret connectives, such as &, $\vee$ and $\rightarrow$, the semantic relation $\models$ corresponds to the subset relation in set theories. This interpretation creates the Boolean lattice structure as the intended semantic structure.

directly to the inference data is illusory: one gets that impression because one has already assigned some arbitrary (or the 'most natural') interpretations to the syntactic rules. Because of these, any theory that makes use of some language in presentation should provide a precise definition of the intended semantics. Whether the intended semantics actually corresponds to the semantics as in the data is a separate issue. If it turns out that some system is incomplete with regard to the intended semantics, then it simply means that the system does not work in a complete way system-internally.

Having said that, because I am mostly concerned about 'truth-based inferences' of the spontaneous inference system, and because classical logic (which I claim to be essentially the same as one's spontaneous inference system at the basic level) is sound and complete with regard to the Boolean semantics as I sketched above (or, more accurately, the Boolean lattice structure as in footnote 8), I implicitly assume that the 'intended semantics' of the spontaneous inference system is simply the standard Boolean semantics, where the derivability in the syntactic derivation schema, $\phi_1,\ldots,\phi_n \vdash \varphi$, is interpreted as the semantic validity argument such that, for all the models in which $\phi_1,\ldots,\phi_n$ are true, φ is also true, as shown in (6).[9] This is convenient, because, as I said above, classical logic is generally complete with regard to this 'truth table' semantics on the one hand, and the truth-table semantics is in close correspondence to one's truth-based judgments over propositions in on-line inferences on the other hand. Thus, by using the Boolean semantics as the intended semantics, we can mostly ignore the difference between the intended semantics of the deductive system and the semantics that models the actual interpretation. Because of this, in this paper, I am often careless about the distinction between the intended semantics of the inference system and the semantics of the inference data. In this way, I aim to show that the stronger claim made by RT (as well as some other systems that eliminate introduction rules from the inference system) is problematic form both theoretical and applicational viewpoints.

Finally, I assume that the pragmatic inference system has properties of 'deductive systems' at the basic level. In other words, I assume that all the rules in the system, including the syntactic rules, the interpretation rules, and the relation(s) over semantic objects in the intended semantics, apply with their full generality. One might define additional rules in the syntax to explicitly control the application of some syntactic rules (such as &I), but then one would also have to provide the

[9] I mostly ignore the semantics of sub-sentential expressions in this paper, because the main topic of this paper is truth functional connectives. See footnote (31), though.

intended semantics for those additional rules, to show that the restrictions in terms of those additional rules really work in the intended way in the semantics.[10]

2 Problems of eliminating &I from the pragmatic inference system

In this section, I discuss some of the problems of eliminating &I from the spontaneous inference system. First, if truth-based judgment is at least part of one's spontaneous inference ability, then the inference system without &I become incomplete with regard to the intended semantics. Also the system uses a rule (that is, CMPP) which the syntactic rules of the connectives involved in that rule do not support. In other words, the system adds an additional theorem which is not supported by the rules of the logical connectives that are manipulated in the theorem. Thus, the inference system fails to be fully deductive.

(7) a. $p, q, (p\&q)\rightarrow r \vdash r$
b. $p\&q, (p\&q)\rightarrow r \vdash r$

Consider the two sequents (or the two 'arguments,' if we see them from a semantic viewpoint) in (7a) and (7b). The two atomic propositions *p* and *q* on the one hand and one complex proposition (*p*&*q*) on the other have the same truth-based interpretation in the antecedents of the sequents.[11] If the inference system cannot make use of &I, one cannot syntactically explain the same role that these formulas play in truth-based interpretations. Without &I, the syntactic system can still derive the entailment relation from (*p*&*q*) to *p*, and from (*p*&*q*) to *q* via &Elimination but that is not complete with regard to the truth-based semantics.[12] Thus the validity of (6a) cannot be explained without &I, and the stronger claim by RT requires CMPP as an essential rule, as has been discussed already.

---

[10] Note that there is asymmetry between the syntax and the semantics throughout the rule formation in the language system. That is, when we control some rule applications, we must first specify the control in the syntax, and then define the interpretation of such control in the semantics. The control in the syntax must be sound and complete with regard to the intended semantics so that the control can really work in the intended way.

[11] I do not show a minimal pair in the other direction. That is, in addition to (7), evaluation of the validities of a pair of sequents such as, i) $p, q, p\rightarrow(q\rightarrow r) \vdash r$ and ii) $p\&q, p\rightarrow(q\rightarrow r) \vdash r$, would be necessary to show the equivalence of the roles of 1) *p*, *q* and 2) *p*&*q*, in an antecedent of a sequent. I omit such a pair because one can prove them only with MPP and &E.

[12] See section 6.1 for further remarks about my recognition of the truth-based semantics as (part of) the intended interpretation of the inference system. I add some comments about Braine and O'Brien's 'procedural' semantics in section 6.3.

However, the reason why CMPP's successive application of $(p\&q)\rightarrow r$ to p and q separately in (5) does not cause a problem for the inference system as a whole is the logical equivalence relation in (8). The proof of this equivalence requires &I, as well as &Elimination (&E).[13]

(8) $(p\&q)\rightarrow r \dashv\vdash p\rightarrow(q\rightarrow r)$

Imagine a logical system with &E but without &I, and call it CPLe. (7a) is not provable in CPLe. Now, imagine that we add CMPP to CPLe and call the resultant system RT. Then, (7a) is provable in RT. Because RT is equipped with MPP (which is an 'elimination' rule of $\rightarrow$ and thus, would be preserved in RT), (7b) is also provable in RT. If we want to maintain congruence (cf. footnote 21) and transitivity of the system, which are both essential for a fully deductive system, then, there should be some path from p, q separately to $p\&q$ as one formula, whereas RT is lacking in this path, that is, &I. Thus, the system fails to be fully deductive.[14] To make my arguments clearer, let me review the inference in (7a), and consider how the stronger claim by RT would relate this inference to the inference in (7b) and how the inference system without &I would syntactically recognize the semantic equivalence between 1) p, q as separate propositions on the one hand, and 2) $(p\&q)$ as one complex proposition on the other, in the antecedent of a sequent in the truth based semantics.[15] Suppose that the inference system were lacking &I (as in the stronger interpretation of RT's proposal). Then, the inference system would not recognize (7a) as a derivable/provable sequent via &I. However, suppose that this hypothetical inference system were equipped with CMPP instead. Then, a person using this inference system could tell that (7a) is a derivable sequent. Next, the person equipped with this inference system could compare the inference in (7a) to another inference in (7b) which her inference system can recognize as derivable, too, but this time by using MPP. Now, she notices that the pair p, q and the complex formula $(p\&q)$ are replaceable with each other in the

[13] See the proofs in (23) in section 5.

[14] In a sense, RT is comparable to a hypothetical Combinatory Categorial Grammar system which insists that they can use function composition without abstraction rule (N.B. function composition as a higher order theorem is derivable from abstraction and association as axioms). I do not investigate whether we can preserve the deductive nature of the inference system without &I but with CMPP by decomposing CMPP into some axioms in a 'modular' system as is sketched in section 6.1 and 6.3. My guess is that there is not a lot of promise. Controlling structural associativity in a multi-modal deductive system is easy because structural rule neither introduces nor eliminates connectives such as &, $\vee$ and $\rightarrow$. In comparison, restricting the use of &I with presence of &E even in one mode would cause a problem to the deductive system. I leave further investigation about this point for another paper.

[15] Again, I show the recognition of the equivalence relation in one direction only.

antecedent of an otherwise equivalent sequent (that is, 7a and 7b are identical except for p, q and $p\&q$) and this replacement does not change the validity of the argument. To the degree that the person using this system finds that this is almost always the case, she can reflectively recognize the semantic fact that *p, q* on the one hand, and ($p\&q$) on the other, have the same (truth-based) interpretation in the premise of a sequent.[16] However, the hypothetical inference system that she is equipped with is still lacking a direct way of supporting this semantically valid inference, because it is lacking &I. Instead, the inference system would recognize it indirectly, as I have shown above.

I do not find a convincing reason to explain our intuition about the valid argument in (7a) and its relation to another valid argument in (7b) in this indirect way (or in this reflective way). In an informal (truth-based) semantic inference, the conditional $(p\&q)\rightarrow r$ requires both p and q to be true (as the standard truth table shows) in order for r to be true, but that is exactly how the propositions p and q are interpreted in the premise of an inference, and thus one can semantically conclude that r is true. The rule &I in classical logic is postulated just to support this semantic judgment, and the inference system as a whole should be equipped with it, in order to make the system complete with regard to this semantic inference.

As I explained above, one can see this incompleteness issue in terms of replacement possibilities between (sets of) propositions. Whenever p, q on the one hand, and ($p\&q$) on the other, appear inside the otherwise identical set of premises, RT can explain why the result of the inferences are the same only in an indirect way. Thus, for all the other cases in which our semantics tells us that the choice between 1) ϕ, φ and 2) $\phi\&\varphi$ in the premises of an argument does not influence its truth-based validity, [17] RT would require some rules analogous to CMPP. For example, consider the semantically valid argument, p, q, $\neg(p\wedge q)\vee r \vdash r$. RT without &I would require another rule analogous to CMPP to support its validity. The fact that $\neg(p\wedge q)\vee r$ and $(p\wedge q)\rightarrow r$ are inter-derivable in classical logic does not help, because, without any introduction rules, the RT inference system cannot recognize them as equivalent (again, without adding another formation rule such as $p\rightarrow q \dashv\vdash \neg p\vee q$ whose addition to the system would further spoil the deductive nature and completeness of the system without introduction rules for truth functional connectives).

---

[16] This is not always the case, because of the sequents, i) $p, q \nvdash p\&q$ and ii) $p\&q \vdash p\&q$ in RT's system. But this case is trivial in the current argument, because i) is exactly the sequent that RT claims that one needs to exclude. Again, I argue that i) should be maintained and i) does not do any harm in its application in spontaneous inferences.

[17] Carson (p.c.) claims that it is not clear why this is something that the spontaneous inference system should be expected to explain. See section 6.4 for this point.

In a similar way, RT might need an additional axiom to deal with the following case.[18] Consider the set of premises P1={p, q, (p&q) $\rightarrow r$, $\neg r$}. How would RT without &I but with CMPP deal with this premise set? As one possibility, RT can first apply CMPP between p and (p&q)$\rightarrow r$, concluding $q \rightarrow r$, to which RT can apply MPP with q, deriving r. Then, RT would get an inconsistent set of premises {r, $\neg r$} from which RT would conclude a contradiction that is, $\perp$. Thus, following this first route, the inference system would derive, p, q, (p&q)$\rightarrow r$, $\neg r \vdash_{\mathsf{RT}} \perp$.

Alternatively, starting from the premise set P1, RT can first apply MTT between (p&q)$\rightarrow r$ and $\neg r$, concluding $\neg$(p&q). Then the resultant premise set would become P2={p, q, $\neg$(p&q)}. Now, in terms of the truth based semantics, the three propositions p, q, $\neg$(p&q) cannot be all true in any model. Therefore, P2 should semantically lead to a contradiction. However, RT cannot syntactically derive a contradiction from P2, because RT is not equipped with &I. Thus, following the second route from P1, RT's verdict is p, q, (p&q)$\rightarrow r$, $\neg r \nvdash_{\mathsf{RT}} \perp$. Comparing the two inference-routes starting from the same premise-set P1, we might argue that the verdicts of the RT's spontaneous inference system is inconsistent, as well as pointing out again the incompleteness of the inference system with regard to the intended semantics (i.e. the verdict of the second route means that there should at least be one semantic model in which all the four formulas p, q, (p&q)$\rightarrow r$, $\neg r$ are true, whereas, as the reader can easily check by drawing a truth table, there does not exist such a model).

Allott suggests that maybe RT would get rid of MTT from the inference system, but as far as the treatment of the premise set P1 is concerned, I am not sure if that is the route they would take. RT would have to deal with the premise set P2 anyway. Thus, RT might define another axiom to derive the sequent, p, q, $\neg$(p&q) $\vdash_{\mathsf{RT}} \perp$. Again, addition of such an axiom that is not supported by the basic rules for the connectives involved in the axiom would spoil the deductive nature of the inference system.[19]

---

[18] Nicholas Allott (p.c.) suggested this case to me. Though I use the set of premises and the basic line of arguments that Allott provided, my analysis may differ from his.

[19] As I implied in the introduction, I do not have a strong view about the claim that the spontaneous inference system is not deductive at all. If that was the case, most of my arguments against a spontaneous inference system without introduction rules would become irrelevant. However, given the productivity of spontaneous inferences, and also given the more than superficial similarity between the logical systems as are investigated by proof theorists and the inference systems that are investigated by psychologists or more empirically minded linguists/philosophers, I do not think that we have to accept the split between the two types of logical systems at the foundational level, rather than at the level of application.

I have shown some reasons not to eliminate &I from the inference system. Now, should we still preserve CMPP in our spontaneous inference system? If the system is equipped with &I, we do not need CMPP as an inference rule. However, because of the equivalence of the two formulas in (8), one may still use CMPP as a shorthand for a set of inference steps in application. That is, given (8), applying $(p\&q)\rightarrow r$ successively to p and q does not cause any problem to the inference system as a whole, based on the basic property of the logic in which replacement of a sub-formula in a sequent with a logically equivalent formula does not influence the provability of the sequent.[20] Analyzing interpretation data is beyond the scope of this paper, but to the degree that data suggest that one may use the inference step as in CMPP, we can still treat CMPP as a shortcut in application.

In this section, I have shown that the inference system cannot recognize the truth-conditional equivalence between certain propositions without &I. Though truth-based arguments are not the only kind of arguments that the spontaneous inference system is intended to support, as long as such semantic arguments are at least important in spontaneous inferences, failing to support them at the basic level of the system compromises the explanatory power of the system as a theoretical tool. In terms of congruence,[21] RT's stronger claim can cause situations in which

---

[20] Sperber & Wilson (1986/1995: 99-100) argue that the rule of CMPP is psychologically plausible in terms of the maximization of the usefulness of the new information that one gets in spontaneous inferences (see also Hall 2006: 96). Roughly speaking, when one has the proposition in the form of $(P\&Q)\rightarrow R$ as one premise, the possibility of finding P and Q separately as other premises is greater than finding $P\&Q$ together. I reserve my view to this point in terms of probability. In terms of efficiency, however, processing one premise after another makes some sense. To support that point, the proof of the sequent $P, Q, P\rightarrow(Q\rightarrow R) \vdash R$ is algorithmically less complex than the proof of the sequent $P, Q, (P\&Q)\rightarrow R \vdash R$, in terms of the complexity measure based on the number of connectives involved in the proofs (i.e. the latter proof includes the introduction of & to conjoin P and Q which increases the complexity of the proof by one). Thus, if one can automatically interpret $(P\&Q)\rightarrow R$ as $P\rightarrow(Q\rightarrow R)$ during a spontaneous inference given the availability of P at that stage of the inference, then the deductive steps that use CMPP will be less complex than the steps using &I, followed by MPP (N.B. the former inference steps would not really derive $P\rightarrow(Q\rightarrow R)$ from $(P\&Q)\rightarrow R$, rather, given the availability of P, one can apply $(P\&Q)\rightarrow R$ directly to P, whereas the logical equivalence between $(P\&Q)\rightarrow R$ and $P\rightarrow(Q\rightarrow R)$ as shown in (23) justifies this successive application of $(P\&Q)\rightarrow R$ to P and Q. So the inference steps will be exactly like (5), though in our system, with a formal underpinning from the basic logical system). The point is that my analysis is totally compatible with the use of CMPP as an application shortcut (CMPP is a higher order theorem that is provable from the basic axioms), and it can also show the efficiency of the spontaneous inference steps using CMPP.

[21] Informally, for all X, Y. X is congruent with Y (in the antecedent of a sequent) iff, for all Z_1, Z_2, W. $[(Z_1, X, Z_2 \vdash W) \Leftrightarrow (Z_1, Y, Z_2 \vdash W)]$ (where X, Y, Z_1, Z_2 are meta-variables for sets of propositions, and commas between such sets represent the set-union $\cup$). That is, X and Y are

one may replace a pair of propositions with a (complex) proposition without influencing the provability of the sequent, but in which one cannot derive the latter from the former directly. This corresponds to the incompleteness of the syntax with regard to the Boolean semantics, but seen from a different viewpoint, we could also say that the syntactic system uses a rule (i.e. CMPP) which is not supported by the basic rules of the connectives involved (that is, the rules for & and → but without the introduction rules). This spoils the deductive nature of the inference system.

3 Alleged over-generation by way of enrichment

In this section, I show that RT does not have to eliminate &I to prevent the alleged overgeneration via free enrichment.

In propositional logic, &I requires as premises two formulas that can be assigned truth values, as is informally shown in (9).

(9) a. Syntax: $p, q \vdash_{\&I} p\&q$
b. Semantics: If $\|p\|$ = True and $\|q\|$ = True, then it follows that $\|p\&q\|$ = True.

Because of this, if one also assumes that the proposition expressed is the first meaning representation that the hearer derives out of a language expression to which the hearer can assign a truth value (in context),[22] it follows that &I (or introduction rules for any truth functional connectives) cannot be used in the derivation of a proposition expressed.[23] Consider (2)~(4) again. Propositions in (3a, b) as contextual assumptions are fully propositional on their own. On the other hand, the semantic content of the sentence in (2), which is uttered by the speaker, acquires a fully propositional status only after it is recognized as the proposition expressed by that utterance. Thus, one can conjoin (2) with (3a) via &I only after recognizing the semantic content of (2) on its own[24] as the proposition expressed.[25]

---

congruent in the antecedent iff for no sequent, replacing one with the other in the antecedent influences the provability of the sequent.

[22] I stipulate that the hearer does not assign a truth value to a minimal proposition. This does not prohibit the hearer from evaluating a minimal proposition or a trivial proposition in the recovery of the proposition expressed. See the end of this section.

[23] Sperber & Wilson explicitly make their mental logic operate over some non-propositional representations, so we have to do some more work to apply this criteria to their system. See section 6.1.

[24] This is inaccurate because one may enrich the semantic content of (2) before applying &I or any other rules for truth functional connectives. Either the literal meaning of (2) or the result of enrichment based on it can enter into &I with (3a). But such enrichment cannot include rules of truth functional connectives because of the semantic requirements of those connectives, plus the assumption that the semantic content of (2) does not become truth evaluable just because the

Assuming that the hearer may derive only one proposition expressed per utterance, it follows that (4) cannot be the proposition expressed by the same utterance of the sentence in (2). Note that in this explanation, one does not need CMPP as an actual logical inference rule. CMPP might still be used to describe an on-line inference step that arises as a result of routinization of certain logical inference steps in application. But use of CMPP does not spoil the fully deductive nature of the inferential system as a whole, either. With &I, the system can recognize the equivalence between the role of the two premises *p* and *q* separately, on the one hand, and the role of (*p*&*q*) as one complex premise, on the other.

I stipulated that one can apply introduction rules for truth conditional connectives only after one enriches the meaning of the overtly used expression to the *proposition expressed*. As well as its role in keeping the inference system sound and complete with regard to the intended semantics, the assumption is supported by the semantic claim that nothing that the hearer recovers from the language expression during the process of deriving the proposition expressed needs to enter into truth based inferences (other than the proposition expressed itself). For example, if the hearer does not see the literal meaning of *John drinks* as relevant enough in the context of an utterance, she does not need to assign a truth value to the proposition that corresponds to that literal meaning (and she does not have that proposition enter into truth based inferences). She only has to assign a truth value to the proposition that she takes as being expressed, say, "John drinks alcohol," for example. Thus, from some sort of economy consideration, I assumed that the speaker does not assign a truth value to a proposition unless she either sees it as the proposition expressed or it is one of the contextually available propositions (which, because of the roles that contextual premises play in inferences, should be assigned a truth value by definition). However, the proposal would have a problem if one had to apply a truth based logical inference rule to a propositional representation that has not yet been accepted as the proposition expressed in other well-attested cases. Some might argue that 'trivial propositions' as in (10) are such cases.

---

semantic content of it is type *t* expression (or, informally, just because what is uttered is a 'sentence.'

[25] Hall (2006) postulates two kinds of pragmatic inferences, local inferences that are applicable to sub-propositional expressions, and global inferences that apply only to fully propositional expressions. Allott (p.c.) suggests that this division might inherently be present in Sperber & Wilson (1986/1995). One can regard my proposal in section 2 as one interpretation of this division between two kinds of inferences.

(10) a. John has a brain.
(uttered to express the proposition, *John is smart.*)
b. Meg is human.
(uttered to express the proposition, *Meg may make mistakes, etc.*)

An argument against my proposal above would be that, in order to derive the propositions expressed (e.g., the ones provided in the parentheses in (10a) and (10b)), the hearer has to evaluate the literal meanings of (10a) and (10b) as trivially true propositions. Some might argue that such evaluation involves truth based inferences.

However, (10a) and (10b) do not pose a problem for my proposal. To recognize the literal meanings of (10a) and (10b) as trivially true, one does not have to apply proper logical inference rules. In other words, to recognize them as trivially true, one does not have to have the trivially true propositions interact with other contextual assumptions in terms of logical inference rules.[26] To explain this point, I first evaluate an application of &I from semantic viewpoints and then come back to inferences involved in trivial proposition cases. In the case of &I, what the semantics of & conjoins is a pair of truth values. Thus, one must know the truth values of *P* and *Q* in order to compute the truth value of *P&Q*. But before *P* and *Q* are accepted as propositions expressed or unless *P* and *Q* are contextual propositions (which are by definition fully propositional), one cannot decide on the truth values of these propositions.

Because semantic validity arguments (e.g., for the provable sequent *P*, *Q* ⊢ *P&Q*, the semantic argument will be, *for all the models in which* P *and* Q *are true,* P&Q *is also true*) abstract away from the choice of a model, some might argue that one does not really need to assign truth values to all the premise propositions of an inference rule for a truth functional connective. They might argue that the choice of a model (which will correspond to the context of the utterance in question) is irrelevant even in inferences for truth functional connectives and thus, one does not need to assign truth values to the premise propositions in such inferences. However, quantification over models presuppose the truth evaluability of each proposition in any of those models, and if one cannot see some bit of the decoded meaning as truth evaluable in spontaneous inferences, one may not use that bit as an input to a truth functional inference rule, such as &I. Thus, though there are still some speculative elements in my proposal, I assume that use of proper classical logic inference rules in spontaneous inferences requires the full truth evaluability of

[26] Such propositional interactions prior to the recovery of the propositions expressed are not necessary even when the literal meanings of (10a, b) are recognized as being informative enough and are accepted as propositions expressed. Contextual premises, plus the linguistic meanings of the relevant expressions, will provide enough clues without such interactions.

all the premise propositions. In use of &I in a spontaneous inference, one has to see "John has a brain" (for example) as the proposition expressed, before one uses it as a premise of &I.

In comparison, let me examine a case of a trivial proposition in more detail. Nicholas Allott (p.c.) claims that, to see that (a) "John has a brain" is trivially true, one would need to retrieve or construct (b) "John is human" and (c) "Humans have brains" and let the three propositions interact somehow. However, there is a crucial difference between this sort of inference on the one hand, and inferences for truth functional connectives (or any inference rules which have properly corresponding rules in classical logic) on the other. For presentation reasons, let me change the proposition (c) into a different form, that is, into (c′) "For all individual *x,* if *x* is human, *x* has a brain." In Allott's example above, (a) is the semantic content of the language expression uttered[27] (i.e., *John has a brain*), whereas (b) and (c′) are contextually available propositions which the hearer may use as premises of an inference rule. Because (b) and (c′) are full propositions (i.e. fully truth evaluable) by definition, the hearer may apply MPP between them and conclude (d) "John has a brain." Because the literal meaning of the utterance, that is, (a), is the same as (d) (in the role that it can play in truth based inferences), the hearer decides not to take (a) as the proposition expressed, and thus does not assign a truth value to (a) (or more accurately, the hearer does not take the proposition "John has a brain" as the proposition expressed by this particular utterance of the sentence *John has a brain*). Note that this process does not require an assignment of a truth value to the semantic content, *John has a brain* which one presumably derives as the result of the linguistic decoding. What is required instead is the recognition that this literal semantic meaning and the contextually derived proposition (d) are the same,[28] and

[27] One could call "John has a brain" a proposition even though it is judged to be trivial for whatever reasons, rather than the 'semantic content' of the sentence, because this proposition exists independently of the English sentence, *John has a brain.* However, because trivial propositions are the cases in which the hearer does not judge such a proposition as the proposition expressed by the utterance (and thus does not accept it as the proposition for the utterance, in my analysis), I do not use the word *proposition* for (a) here, to avoid unnecessary confusion. In order to be complete, I should consider all the different reasons why such propositions are judged to be trivial, relative to the language expressions used in the context. For lack of space, however, I leave such a complete exposition for another paper.

[28] To be complete in my arguments, I would need to show case by case that this identification process actually does not involve any use of logical inference rules as in classical logic. Things will be easy if there are only two cases involved in trivial propositions, that is, either (a) is a tautology or (a) as a candidate for the proposition expressed and (d) as a conclusion derivable from contextual assumptions only are exactly the same proposition, because the use of classical logic inference rules is not necessary in the triviality judgment in either of these two cases. But I am not certain whether this is always the case. I will leave it for further research.

thus, identification of (a) as the proposition expressed by this utterance does not allow the hearer to do anything more than she could do otherwise.

Though my informal exposition here still contains some arbitrary elements and should be formulated more accurately in some other occasions, it relates the trivial proposition case as in Allott to cases that involve tautologies, such as *Boys will be boys* or *People who study math study math*. If tautologous propositions are real tautologies, by definition, they are always true and thus the addition of them into the premise set of propositions will not allow the hearer to derive any other conclusions that she could derive without them. Thus, given an utterance of the sentence *People who study math study math*,[29] the hearer may start enriching its meaning before she lets the literal meaning of the utterance interact with other positions in the context at all.

Dealing with Allott's example above, I let the semantic content of the sentence uttered interact with another contextually derived proposition. This is not problematic because the analysis only prohibits 'truth based' interaction between the linguistic meanings of the expressions used and contextually available information. Thus, after accepting my assumptions, one can still enrich the meanings of component expressions by using the information provided by contextual assumptions. In fact, one may even 'mimic' some of the seemingly truth based inferences.[30] For example, with model theoretic relations such as sub-set relations, one can mimic logical entailment relations without deriving a fully propositional representation. In this way, one can enrich the meanings of predicate expressions *smart* or *human* via set-containment relations, for example, without deriving a full proposition. On the other hand, I argue that proper logical introduction rules do require fully truth evaluable elements as premises (just as is the case in standard logical systems) and thus, they cannot be mimicked in terms of relations between sets.[31]

---

[29] What proposition is expressed by the use of this sentence will depend on each context and is not relevant to the discussion here.

[30] Of course, the claim is that such 'seemingly' truth based inferences are not really truth based. In fact, the claim is stronger than that. It would state that pragmatic inferences that modify the meanings of the expressions uttered prior to the recovery of the proposition expressed are not 'propositional' inferences, whether propositions are interpreted in terms of their truth values or not (see section 6.1). But I do not have space to explain this point, and thus I use 'truth based inferences' as the guiding criteria.

[31] I regard generalized conjunction as in Partee and Rooth (1983) only as a rule of PF-LF mapping. For example, at an intermediate stage of the syntactic derivation for the sentence, *Jack and Eva smoked*, the syntactic system might interpret the natural language expression *and* as a lambda term, $\lambda x.\lambda y.\lambda P.P(x)\&P(y)$, so that this derived functor expression can be applied to the individual tems jack′ and eva′ successively, deriving another lambda expression, $\lambda P.P(\text{jack}')\&P(\text{eva}')$. However, note that the logical connective & itself stays as a truth functional connective in any of these lambda terms (i.e. it conjoins two propositions, such as $P(x)$ and $P(y)$). Thus, the use

Finally, I briefly discuss an example of a minimal proposition that Carston suggested (p.c.). Interpreting the utterance of the sentence *John poisoned Bill and Bill died* in a relevant context, the hearer may recover the minimal proposition "John poisoned Bill and Bill died" (= A) in the process of recovering the proposition expressed, "John poisoned Bill and Bill died because of that poisoning" (=B), adding the causal relation to the truth-conditional content via enrichment. However, nothing in my analysis prohibits this enrichment process. All I am claiming is that this enrichment process is not 'truth functional' inferences, as &I, &E and MPP etc. are. Causal relations are not truth functional relation, as one can see from the meaning of *because*, which is not a truth functional connective, as one can see in a sentence such as *John came because Eva came*.[32] Evaluation of sub-propositional elements or even propositional elements recovered from the sentence uttered is still possible, and from empirical considerations, is necessary, as Carston suggests. But my argument is that though such inferences are still pragmatic inferences (and thus, are constrained by the principles of Relevance), there is a formal difference between such inferences and 'truth functional' inferences which are deriving (sets of) propositions from (sets of) propositions solely based on their truth value assignments (in all models). Given that difference, I stipulate a certain feeding relation between these two kinds of pragmatic operation at the theoretical level. That is, 'non truth-functional' processes which includes enrichment theoretically feed into the truth functional ones that include the rules for logical connectives.

I leave for further research exactly which inferences can be mimicked in this way and which cannot be.

This section explained why introduction rules are not applicable in enrichment. The next section deals with the alleged 'infinite inference' problem.

4 Alleged Infinity problem caused by introduction rules

This section briefly addresses the claim that if one's pragmatic inference system were equipped with logical introduction rules, one would run infinite or non-terminating inferences. Because such non-terminating inferences are not attested in

---

of generalized conjunction in terms of the lambda abstracted terms as above does not influence the fully truth functional status of &, $\vee$, $\rightarrow$, etc at the level of logical forms.

[32] The fact that the proposition B entails the proposition A does not mean that B must be deduced from A by using 'truth functional' inferences as is used in classical propositional logic. Entailment relations may come from the preservation of the lexical meanings, and at least in this example of minimal proposition, none of the classical logic rules is used in the enrichment process from A to B (because, again, the causal relation is not truth functional, and neither is the temporal precedence relation, such as "After P, Q").

interpretation data, it must be the case that the pragmatic inference system does not have access to introduction rules. Arguing against such claims, I show that this problem is caused independently of the use of introduction rules and should be solved independently.

Johnson-Laird (1997: 391) claims that introduction rules, if they are used in spontaneous inferences, may lead to infinite inference steps, as schematized in (11).[33]

(11) a. $P, Q \vdash_{\&I} P\&Q \vdash_{\&I} P\&Q\&P \vdash_{\&I} \ldots$
b. $P \vdash_{\vee I} P\vee Q \vdash_{\vee I} P\vee Q\vee R \vdash_{\vee I} \ldots$

However, the alleged infinity in (10a) is because of the expansion of '*P*' to '*P*, *P*,' and '*Q*' to '*Q*, *Q*.'[34] It is not because of &I per se. Thus, eliminating &I from the system does not solve the problem completely, to the degree that the problem exists. Also, with regard to this structural expansion rule, note that one occurrence and more than one occurrence of the same formula have the same interpretation in truth-based inferences. Thus, the alleged infinity might be just a matter of the imperfect representation system, rather than some imperfection of the inference system. In fact, even at the level of represented deductions, logicians have tried to eliminate un-decidability induced by structural rule applications. Without going into details, one may apply a structural rule only when the consequence of that rule application is required by the next step of the inference.[35]

In (11b), $\vee$I presupposes weakening of the succedent set. Because the standard introduction rule for $\vee$ implicitly includes the structural weakening in the Succedent side, one has to separate the concept of $\vee$I and the concept of weakening, first, and then find out which of these has created the alleged infinity problem.[36] Because $\vee$I persists across different logical systems with different

[33] Braine and O'Brien also describe this version of the problem. Cf. O'Brien (2004).

[34] This notation is slightly sloppy, because *P, Q* must stay as premises of inference in order to be interpreted as 'And.' Gentzen sequent presentation captures the semantic equivalence of *P*, *Q* and *P&Q* in the antecedent of a sequent in a better way (see 16a), though comparison is not straightforward. Because of some technical details, (16a) formally corresponds to &E, rather than &I. Such technical details, however, do not matter. With $\wedge$L and $\wedge$R in (16a), the Gentzen system is complete with regard to the intended interpretation, whereas the system without &I (such as RT's stronger claim) is not.

[35] Braine and O'Brien proposed a similar, but a different proposal in spirit. That is, they modify the underlying algorithm of their system. Because they divide rules of inference into groups which are not supported by the underlying logical system, it causes several problems, incompleteness as one. See section 5.3.

[36] Došen (1988) and Belnap (1996), among others, recommend rule presentations which separate the two concepts, a) rules of connectives and b) structural property of the system (where

structural management properties (e.g. presence or lack of structural weakening), and because it is weakening in the succedent that increases the number of propositions such as Q and R in (11b),[37] the infinity problem, to the degree that it gives problems to the inference system, is a matter of structural weakening rule, rather than $\vee$I. Thus, one cannot fully control this problem merely by eliminating $\vee$I. Just like expansion of the formulas, it is beyond the scope of this paper to discuss whether weakening does cause problems to the spontaneous inference system, and if it does, how to control it. One may adopt Intuitionistic logic which is lacking in weakening in the succulent, for example. But using a sub-structural logic makes the system incomplete with regard to the truth based semantics. Thus, it cannot be used to explain one's truth based inferences in a complete way.[38]

Instead of modifying the underlying algorithm of the inference system, I would rather control structural rules at the level of application, as was suggested above. That is, one might set up the forward looking inferences in such a way such that one may apply structural weakening rule only if it's output is required by a further inference step.[39] Informally, this means that one weakens P to P, Q, only if, say, one has $(P\vee Q)\rightarrow R$, as another premise.[40] Alternatively, one may try a proof representation system which does not incorporate the structural weakening into the rule of $\vee$I, but which leaves the weakening rule implicit, so that the spurious ambiguity that is caused by application or non-application of the weakening rule simply does not arise. This analysis requires some technical explanation, and I leave the details for another paper.[41]

As another variety of the alleged infinity associated with introduction rules, some might argue that recursive applications of &Introduction followed by &Elimination would produce infinite inference steps, but this infinity does not arise in standard proof representations without a Cut, such as Gentzen sequent presentation without Cut. Some proofs are listed in section 5 (see (17)~(20)).

---

b) is explicitly represented as structural rules separately from a). In that conception, $\vee$I is independent of structural weakening in the Succedent. For example, "The rules for the logical operations are never changed: all changes are made in the structural rules." (Došen, 1988: 352).

[37] One can rewrite (10b) as $P \vdash P, Q \vdash P, Q, R$, …etc., without introducing $\vee$.

[38] Whether Intuitionistic logic is still useful in a 'modular' inference system in one of its modules is a separate issue. See section 6.1.

[39] One can prove that controlling structural rule application in this way does not influence provability of sequents. Thus, the system will stay complete. See section 6.3.

[40] In this case, structural weakening feeds into $\vee$I, which feeds into $\rightarrow$. Thus, in this particular case, it leads to the same result as Braine and O'Brien. But the way that we achieve it is better, for the reason that we explained already. In this proposal, $\vee$ itself is freely applicable, as long as there are P and Q in the Succedent side.

[41] Roughly, notions such as 'monotonicity' and 'purity' may be assigned to the system itself. See Avron (1993) for the explanation of these ideas. See Wansing (1998:92) as well.

Finally, I discuss a more sophisticated infinity argument. Consider (12).

(12) (Non-) Frame problem.
a. Antecedent Set $\vdash_{\Delta}$ Succedent Set
b. $P \vdash_{\{Q\}\&I} P\&Q$
c. cf. $P, Q \vdash_{\&I} P\&Q$

(12a) represents a spontaneous on-line inference step. Though the logical inference rules are the same as in classical logic, (12a) distinguishes between two kinds of databases that are used as premises. The antecedent embodies the set of premise propositions that are active in the context, including the proposition expressed by the utterance. To draw a conclusion in the succedent set, one can also use premise propositions in the 'dormant database' set Δ, which contains the whole of the (propositional) knowledge that one has.

With these assumptions, some might argue that the inference system would wrongly predict the existence of an infinite inference as in (13).

(13) $P \vdash_{\{Q, R, S...\}\&I} P\&Q \vdash_{\{R, S,,...\}\&I} P\&Q\&R \vdash_{\{S,...\}\&I} P\&Q\&R\&S \vdash_{\{...\}} \ldots$

In (13), one may extract one proposition after another from the dormant database set and conjoin them with the proposition *P* in the active premise set. If one assumes that the amount of one's knowledge is almost infinite, this model wrongly predicts that one may actually run an almost infinite inference.[42]

However, note that this alleged infinity is not a matter of &I per se. As I have already pointed out, in the antecedent set, *P* and *Q* as separate propositions on the one hand, and *P&Q* as a single complex proposition on the other, play the same role in the classical logic. Thus, the above infinity problem will arise independently of the use of &I. What is problematic then is the introduction of *Q, R. S* into the active data-base, not the conjunction of those newly introduced propositions with a proposition that is already in the active data base. Thus, what one needs is a systematic way of constraining the introduction of propositions from the dormant database to the active database.

This section has shown that use of introduction rules in the inference system is not the cause of the alleged non-terminating inferences, and that the elimination of introduction rules does not solve the problem.

[42] Also, in a spontaneous inference, one does not typically access all the pieces of knowledge that one has, even if the pieces of knowledge are relevant to the argument one is making.

5 Some Gentzen sequent proofs

In this section, I show that successive applications of &I (or &R in this section) and &E (or &L) do not lead to undecidability. I also show that $(p\&q)\rightarrow r$ and $p\rightarrow(q\rightarrow r)$ are inter-derivable. The proofs here are elementary, and are a simple application of the Gentzen sequent presentation of classical logic as in Girard (1987) or Takeuti (1987).

(14) Sequent to prove (e.g.) $p, q, (p\&q)\rightarrow r \vdash r$

Gentzen sequent proof representation places the sequent to prove at the bottom of the derivation. Then, one logical connective after another is eliminated upwards along the chain, as is shown in below examples. If the proof is successful, the sequents at the top of the proof are all identify axioms in the form of (15).

(15) Axiom: $A \vdash A$

By convention, *p, q, r…*represent atomic propositional letters, *A, B, C* represent any (propositional) formulas, and *X, Y, Z* represent sets of such formulas. I omit the set notations both in the antecedent (i.e. the left-hand) side of each turnstile and the succedent (i.e. the right-hand) side. (16) shows the axioms for the connectives, & and $\rightarrow$ (I use $\wedge$ for & in Gentzen sequent presentation for some technical reasons). I omit the rules for other connectives. *Cut* is an admissible rule[43] which is not necessary for the proof system, but is useful for improving the efficiency of the proof.

(16) Logical rules:

a. $\dfrac{A, B \vdash X}{A\wedge B \vdash X}\ \wedge L$ $\qquad$ $\dfrac{X \vdash A \quad Y \vdash B}{X,Y \vdash A\wedge B}\ \wedge R$

b. $\dfrac{X \vdash A \quad Y, B \vdash Z}{X, Y, A\rightarrow B \vdash Z}\ \rightarrow L$ $\qquad$ $\dfrac{X, A, Y \vdash B}{X, Y \vdash A\rightarrow B}\ \rightarrow R$

c. $\dfrac{X \vdash A \quad A \vdash Z}{X \vdash Y}\ \text{Cut}$

I have omitted some of the 'contextual' structural variables (i.e. *X, Y, …*) for readability. Note that $\wedge$L in (16a) is 'pure' in the sense that it does not introduce a new propositional variable in the inference from the top to the bottom, as opposed

[43] That is, if any sequent that we can prove with Cut is provable without Cut.

to the 'impure' inference rule from $A \vdash X$ to $A \wedge B \vdash X$, which is valid in classical logic, but has incorporated structural weakening in the Antecedent. The 'pure' presentation is preferable for the reason that we have discussed in section 4. In (16), except for the Cut rule,[44] the number of the connectives decreases by one along each consecutive step upwards. Because there are only a finite number of connectives in each sequent to be proved, any proof is decidable in a finite step, unless Cut is used.

Remember the successive use of &I (= $\wedge$R here) and &E(=$\wedge$L), which may allegedly lead to an infinite inference. With Cut, this claim is substantiated, as in (17).

(17) *Proof 1*

$$\dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash p \wedge q}\wedge R \qquad \dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash p \wedge q}\wedge R}{p \wedge q \vdash p \wedge q}\wedge L \qquad r \vdash r}{(p \wedge q), (p \wedge q) \rightarrow r \vdash r}\rightarrow L}{p, q, (p \wedge q) \rightarrow r \vdash r}\,Cut$$

(18), in which Γ and Δ represent the two sub-proofs of (17), represents the proof in (17) in brief. If the Cut rule is used, then this proof might not terminate in a finite step, given the sequent to prove, p, q, $(p \wedge q) \rightarrow r \vdash r$.

(18) *Proof 1 (with abbreviation)*

$$\dfrac{\Gamma \qquad \Delta}{p, q, (p \wedge q) \rightarrow r \vdash r}\,Cut$$

In the position of the sub-proof Γ in proof 1, one could insert a larger sub-proof, e.g., the whole of the proof 2 in (19).

[44] Cut is not a logical rule (which is a rule for a truth functional connective/operator). Its inclusion in (16) is for presentational convenience only.

(19) *(Sub)-proof 2*

$$\dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash p \wedge q}\wedge R \qquad \dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash p \wedge q}\wedge R}{p \wedge q \vdash p \wedge q}\wedge L}{p, q \vdash p \wedge q} Cut$$

Note that the left premise and the conclusion of Cut are both *p, q* ⊢ *p*∧*q*. Thus, we can use the conclusion sequent as a left premise of another Cut, by repeating the whole of the right premise of the original Cut as the right premise of this additional Cut. Thus, there is no maximal limit to the size of the sub-proof in (19), leading to the infinity (or undecidability) problem.

However, as Girard (1987) and others showed, Cut is an admissible rule in Gentzen sequent presentation. Without Cut, Proof 1 is represented as Proof 3.

(20) *Proof 3* (*Without Cut*)

$$\dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash p \wedge q}\wedge R \qquad r \vdash r}{p, q, (p \wedge q) \rightarrow r \vdash r} \rightarrow L$$

Other than Cut, all the rules in (15)~(16) reduce the number of connectives by one along each consecutive step upwards, and thus, all the proofs are decidable in finite steps. Consequently, successive use of ∧L and ∧R does not lead to an infinite inference.

Finally, (22) show that the equivalence in (8), repeated here as (22), is provable only with &I (or ∧R here) as a rule of the logic. The proof in (23a) requires ∧R in the top left sub-proof.

(22) (*p*∧*q*)→*r* ⊣ ⊢ *p*→(*q*→*r*)

(23) a. ⊢

$$\dfrac{\dfrac{\dfrac{\dfrac{p \vdash p \quad q \vdash q}{p, q \vdash (p \wedge q)}\wedge R \qquad r \vdash r}{p, q, (p \wedge q) \rightarrow r \vdash r} \rightarrow L}{p, (p \wedge q) \rightarrow r \vdash q \rightarrow r} \rightarrow R}{(p \wedge q) \rightarrow r \vdash p \rightarrow (q \rightarrow r)} \rightarrow R$$

b. ⊣

$$\dfrac{\dfrac{\dfrac{p \vdash p \quad \dfrac{q \vdash q \quad r \vdash r}{q, q \to r \vdash r} \to L}{p, q, p \to (q \to r) \vdash r} \to L}{p \wedge q, p \to (q \to r) \vdash r} \wedge L}{p \to (q \to r) \vdash (p \wedge q) \to r} \to R$$

In this section, I showed that successive use of &I(=∧R) and &E(=∧L) does not lead to an infinite inference in Gentzen sequent presentation without Cut. I also showed that we need &I as an inference rule to support CMPP as an application rule in spontaneous inference. I did not show how we can prevent infinity which could be induced by the use of structural rules (such as expansion and weakening) in the proof presentations, but for some rough ideas (in the context of Modal logic), see Hudelmaier (1996).

6 Loose ends and speculations

This section deals with some loose ends. The discussion will be mostly speculative and incomplete.

6.1 Mental logic over non-propositional representations.

As I wrote in section 3 (cf. footnote 23), Sperber & Wilson make their mental logic operate over 'non-propositional representations,' and I add some comments to this claim.

One can interpret this claim in two different ways. One interpretation is that, in Sperber & Wilson's inference system, propositional letters are not always interpreted in terms of their truth values. With this interpretation, it is misleading to state that their mental logic operates over 'non-propositional representations,' because the underlying system may still be propositional logic, only with different resource management properties (or, less technically, with different 'interpretations' of propositional formulas). Thus, we can still stay inside propositional logic, only with varying ways of interpreting the propositional language.

Though truth based inferences may not be the only kind of inferences for Relevance Theory, they are still an important target of their pragmatic analysis. Thus, we can safely assume that part of the tasks of their mental logic is to explain one's spontaneous truth-based inferences. But then the incompleteness problem, among others, is as serious a problem in their system as in an inference system with

the truth-based semantics as the 'only' intended interpretation. Because of this, the claim that their mental logic operates over 'propositional' letters that may be interpreted in a different way from their truth values does not make their system without &I any less problematic. It will still underachieve its intended tasks (or they may need to add stipulative side conditions to make it work in a complete way).

On the other hand, it is true that such a claim makes the application of my proposal less straightforward. Instantiation of the proposal in S&W's inference system requires further research. In this section, I only sketch a speculative way of applying the proposal to such a multi-purpose inference system.

As long as S&W can use the same inference language, only interpreted in varying ways depending on which kinds of inferences they are dealing with, we can keep the logical language more or less the same as the one in classical propositional logic. To simplify things, let me stick to the propositional logic language that we have used in this paper, made out of expressions such as P, $P\&Q$, $P\rightarrow Q$ etc. To instantiate a multi-purpose inference system as above, we may see the expressions in the logical language in multi-modal interpretations.[45] That is, in one mode of interpretation, one will interpret the formulas in terms of their truth values (then in this mode of interpretation, the inference system is basically the classical propositional logic). In another mode, one may interpret *P, Q* etc. as 'resources,' as in linear logic (then in this mode, one occurrence and two occurrences of the same formula, say, P, make a difference, just as one bottle of beer and two bottles of beer as resources are interpreted differently). With such multi-level interpretations, as long as the inference system is equipped with the whole set of introduction and elimination rules for all the connectives in the mode of truth based interpretation, one can at least confirm that the system does its job in a complete way as far as the truth based inference is concerned. But we can also propose a multi-modal system in which the system is sound and complete in every mode, with regard to the intended interpretation in each mode (that is, the system will do its job in a complete way in each mode of interpretation). As I sketched in section 4, the alleged infinite inference problem is caused by simplistic application of structural rules. Thus, the multi-modal system may solve this problem without stipualtively banning introduction rules for truth functional connectives. Instead, we can choose the right set of modes of interpretations with appropriate structure management properties as its sub-systems.

---

[45] To have the interpretations in different modes interact with one another, we would have to modify the (logical) language expressions in the system, such as introducing some ways of signifying the different modes of interpretations or defining interaction rules between different modes, where such interaction rules might in turn require an addition of modal operators into the logical language. I leave a more accurate exposition of such a multi-modal system for another occasion.

As I said above, I only provide a sketch of how to instantiate this paper's proposal in such a multi-modal system. First, with regard to the 'fully propositional status' of the linguistic meaning of some language expression, we may simply base such status on the truth-evaluability of the expression in its interpretation in the truth-based mode of interpretation.[46] That is, the rules for propositional connectives/operators cannot apply in any mode until all the premises of the rules are judged to be fully truth evaluable in the truth based mode of interpretation. This will solve the alleged overgeneration in enrichment case. What is more difficult is how to control the alleged infinity in terms of structural rule application (again, I assume that the problem is not caused by introduction rules per se). This will require more careful work, because, by definition, structural properties of logical expressions vary across various modes of interpretations. However, the different structure management properties in different modes mean that we can naturally get rid of certain structural rules in certain modes of interpretations without stipluatively banning those rules (e.g. in Intuitionistic logic, weakening in the succedent is impossible, and thus, the alleged infinity in (11b) simply does not arise).

In this paper, I do not investigate whether such mode internal variation of structural properties is enough to cover all the kinds of inferences that Sperber &Wilson aim to explain. For example, an interesting question with regard to the multi-level interpretations of the logical expressions is whether we need to include a mode of interpretation in which the expressions are interpreted as tokens. But note that even in this mode, it does not make sense to see '&' as a token as well, if we still see it as a 'deductive' system.[47] Remember that formulas such as p, q, r on the one hand, and the connectives, operators such as &, $\vee$ have different statuses in logic. For example, only the former can stand on their own as well-formed expressions in the language, whereas the latter could not do this, as shown by the

[46] I abstract away from the mapping between natural language expressions used in utterances and the corresponding expressions in logical languages because it requires a more complex exposition. Roughly, the statement in the main text would change into the following: "we may simply base the truth evaluability of a language expression uttered (i.e. a 'sentence' in the case of a trivial proposition) on whether the corresponding logical expression (i.e. a formula in the case of a trivial proposition) can play a non-trivial role in the truth based inference in the context of the utterance." To remind the reader of my argument in section 3, one can decide whether the logical expression in question can play a non-trivial role or not in such inferences without letting it interact with other contextually available propositional formulas in terms of truth based inference rules. See section 3 for details.

[47] We could stop seeing this level or mode as part of the deductive system, but then it would become impossible to relate this mode to the other modes which are defined to be deductive. As an example, note that in Lambek calculus, which pairs LF as a relational structure with PF as a relational structure, phonological strings include (interpretations of) 'logical connectives,' such as the binary connective · which connects, *the* and *boy* producing, (*the*· *boy*).

ill-formed expressions, &, $\vee p$ and $q\rightarrow$. These connectives are not independent elements in the language; they play a role of mapping (simpler) formulas to (more complex) formulas in terms of functional derivability, where the derivability is inherently related to the basic property of the deductive system. Thus, insisting on the difference between p, q on the one hand, and ($p\&q$) (in the Antecedent) just because only the latter has & as a token misses the point. If we interpreted & etc. as a token (and if we did not add proper logical connectives instead of them, which would support the 'token' based deductive system), then the system would simply stop being deductive. The interpretations of the connectives are inherently related to the properties of the deductive system that makes use of them. That is why 'commas' in the Antecedent of a sequent have the same interpretation as & and 'commas' in the succedent of a sequent have the same interpretation as $\vee$. Just like 'commas' in proof representations cannot be treated as tokens, the connectives whose interpretations are inherently related to those commas cannot be treated as tokens. From a different viewpoint, propositions or any well-formed formulas can be interpreted as tokens because they are part of the set of well formed expressions in the language. On the other hand, the connectives such as $\wedge$, $\vee$ and $\rightarrow$ are not well-formed expressions on their own. They are defined to express the derivability relations supported in the chosen deductive system. Thus, if they were treated as tokens and were not assigned the semantics that are expected from the basic properties of the deductive system, then the system would stop becoming deductive. This means that even in the interpretation of logical expressions as tokens, we cannot treat & as a token. We would have to define some functional interpretation which maps the token interpretations of p and q to the token interpretation of $p\&q$.[48] I leave the structural property in that mode of interpretation for future research.

The second way of interpreting the claim that Mental Logic operates over non-propositional expressions is that their inference system deals with sub-propositional expressions (such as individual terms and predicate expressions), as well as propositional expressions. But we can accommodate this requirement just by using a system such as predicate logic (or a variant of higher order logic) at sub-propositional levels. In this paper, I did not explicitly look into substructures of the propositional expressions such as P, Q, R, but we presupposed the necessity of sub-propositional expressions in the inference language in many parts of the paper. For example, in order to model entailment relations at sub-propositional level in terms of set containment relations, I would need to use sub-propositional expressions in the logical language. Incompleteness at the sub-propositional level is a separate

[48] As in linear logic, we would need different kinds of &, such as a multiplicative one as opposed to an additive one.

issue that I abstract away from.[49] At the moment, as long as the inference system is complete at the level of the truth-based propositional calculus, it serves our purpose (note that in the standard predicate calculus, the connectives &, ∨, → etc. are still interpreted as propositional/truth functional connectives. See footnote 31).

[49] This does not mean that incompleteness does not matter at the sub-propositional level. Rather, it is not yet clear to me how we can use deductive inferences to explain spontaneous inferences at that level. A Lambek Calculus NL, which regards syntactic categories as formulas, such as NP→S, S, NP•NP (cf. $p \rightarrow q$, q, $p \& q$, respectively), has been shown to be complete with regard to a well-defined semantic model (i.e. free groupoid), and this shows that it is possible to achieve completeness with regard to some well-defined semantics by providing formulas-as-types to both propositional and sub-propositional logical expressions. However, in our case, the inference is not about compositional derivations of LF representations from the lexical level (which the use of logic in the syntax can take care of). Rather, it is about spontaneous inferences manipulating the out-put of the syntactic derivation at LF after its near isomorphic translation to Language of Thought (LoT) representations. I am not sure how the use of deductive systems can constrain spontaneous inferences at the sub-propositional level in LoT. Lexical enrichment is an example of such sub-propositional inferences. However, whereas 'narrowing' in lexical enrichment might be captured in terms of set-containment relation between sub-propositional concepts, it is not clear how using predicate logic (or higher order logic) can provide insight in the process of loosening. If we use first-order predicate logic or a variant of simply typed lambda expressions to represent LoT, then to the degree that the logical language expressions are constrained by some formal properties of the language, spontaneous inferences using such language expressions will be constrained as well. However, as one can informally understand by considering how the use of English may constrain one's general thinking in English, the way that one's inference is constrained by the language that one uses in inference (because of the limited expressive power of the language) may be quite different from the way that the classical propositional logic may constrain one's spontaneous propositional inferences as has been discussed in this paper. One possibility is that 'sub-propositional inferential processes' involved in enrichment etc. are not deductive in a direct way. That is, to the degree that both the starting point and the end result of (lexical) enrichment are part of propositional representations which have some formal structures as are expressible in simply-typed lambda expressions, for example, enrichment may still be constrained by the syntax of such logical expressions, but again, that is a different kind of constraints from the constraints that (the deductive rules of) the classical propositional logic may provide for propositional inferences. In this interpretation, the process of enrichment is not explainable in terms of a logical rule such as &I, MPP etc. It is only that enrichment manipulates some logical language representations. We can explain possible use of set-containment relations in enrichment (i.e. narrowing) in this indirect way. That is, enrichment may manipulate certain (logical) properties of the language representations that it operates over, and also, because the final product of sub-propositional inferences is the proposition expressed which does enter into properly deductive propositional inferences, enrichment may be geared towards the preservation of 'logical entailment relation' (again, mimicked by set-containment relations) at the sub-propositional level. However, the process of enrichment itself might still not be definable as a deductive rule. I leave this issue for future research.

6.2 Denotational or procedural views of semantics and soundness and completeness of the system

In this paper, I adopted a 'denotational view' of semantics. That is, our semantics modelled the denotations of the logical expressions, rather than modelling the syntactic proofs/derivations themselves, as Heyting did (see the next paragraph). Braine and O'Brien (e.g. O'Brien 2004), on the other hand, explicitly advocate a kind of 'procedural semantics.' The claim is that the semantics of the connectives are based on what they allow one to do with them.

There are two ways of interpreting this claim. In one interpretation, their semantics is based on what mental logic rules allow one to do with them *in the semantics*. In this case, however, there is no inherent difference between their conceptions of the semantics and the above mentioned 'denotational' view of the semantics. In the denotational view, the semantics of logical language models what the syntactic system can do by way of interpreting the syntactic objects in the intended semantic structure. In that sense, the semantics of the deductive system does correspond to what the syntactic system allows one to do in the semantics. Whether or not they use truth tables in the intended semantics is a separate issue. We could interpret the inference language in a semantic structure that is different from the one represented by truth tables (i.e. the Boolean lattice). Braine and O'Brien could define whatever semantic structure is suitable for their purpose as long as the intended semantics is formally well-defined. But whatever denotations they may assign to the inference language, they must check whether what the system allows one to do at the level of syntactic derivations matches up with what the system allows one to do at the level of the (system internal) denotations in a sound and complete way, so that one can confirm that the system actually does the job that they intend it to do.

The alternative interpretation of their claim is more interesting. They might be assuming that their semantics directly model their 'proofs' (or their deductive steps). This reminds me of Heyting's semantics. In interpreting propositional languages, Heyting did not try to find out when each propositional formula is true. Instead, he tried to find out what the proof of each formula is (cf. Girard 1989: 5). Thus, Heyting first stipulated that the interpretation of each atomic formula (say, P) is its proof.[50] After that, he stipulated that a proof of $P \wedge Q$ is a pair (p, q) consisting of a proof p of P and a proof q of Q (cf. Girard, 1989:5).

---

[50] What counts as a proof of an atomic formula is not clear, but consider what one would do to prove each P, such as 2+3=5. Probably one could place two objects on the table, add three more objects to them, and then count the total as five, which may count as a proof of 2+3=5. In any case, the point of the direct interpretation of proofs is that, once we agree upon the proof of each atomic formula as an interpretation of that formula, then we can compute the proofs of more complex formulas out of the proofs of atomic formulas at the level of model structure, in the way

However, I do not think that an attempt to interpret the syntax of their mental logic in this direct way would be successful for Braine and O'Brien for various reasons. First, in the case of Heyting, he was interested in the direct interpretations of the proofs themselves. Thus, for Heyting, it does not matter what the resultant semantics turns out to be, as long as it directly represents the syntactic proofs/derivations (and as long as the semantic structure turns out to be well-defined). This is not the case for Braine and O'Brien, they have some empirical phenomena to explain, that is, spontaneous inferences as psychological phenomena. Thus, their semantics should have an appropriate structure as a model of one's system for spontaneous inferences.[51] Secondly, the incompleteness of the system without &I[52] comes from the 'incompleteness' of the syntactic system at the basic level.[53] Given this inherent incompleteness of their syntactic system, direct interpretation of their system is not likely to help Braine & O'Brien's system (which does not have the property of symmetry) with regard to soundness and completeness relative to the direct interpretation. This is because the semantic structure is evaluated with full generality. Even though the initial interpretation rules map only the objects that their syntactic rules allow one to generate onto some

---

suggested in the main text. Note that even in this direct interpretation method, Heyting had to provide some abstract 'denotations' to the language expressions, whether they are atomic or complex. The difference between the 'denotational view' and the direct/syntactic view of semantics then is simply that for the denotational view, one provides a matching semantic structure to a syntactic system (such as a Boolean lattice to the classical propositional logic) in the first place and then tries to prove soundness and completeness of the syntactic system with regard to that semantics, or for the syntactic view, one tries to directly represent each syntactic object and all the proof steps in one's semantic model, hoping that the resultant semantics constitutes a well-defined structure. The benefit of the first strategy is that the semantic structure is already well-defined at the start, because one picks up a well-defined structure in the first place. But soundness and completeness might not hold (and then one might look for another well-defined semantic structure as a candidate). With the latter view, one tries to set up the semantics in such a way that it follows each syntactic proof step. Ideally, the syntax should become sound and complete with regard to the resultant semantics created in this way. However, a problem of this second strategy is that there is no confirmation that one can create a well-defined semantic structure at the end of the day. Also, for technical reasons, maintaining soundness and completeness turns out not to be so easy to sustain even in this way of setting up the semantics tailor-made for the syntax. See chapter 1 of Girard (1989) in this regard.

[51] In other words, though the (system-internal) semantics of a logical language is independent of the (system-external) semantics as is represented in the inference data, these two kinds of semantics should match up quite closely (ideally, they should be isomorphic to one another) so that one can use the system in an empirically meaningful way.

[52] Or in Braine and O'Brien's theory, the incompleteness of the system which puts &I into a different group from the core group of rules. See the next subsection, 6.3.

[53] Informally, a syntactic system itself would become 'incomplete' if it is equipped with only one of the pair of rules for a connective used in the language, unless this elimination falls out from the basic structural property of that language.

semantic objects (and thus, their syntax will be sound and complete with regard to the semantics at this initial stage), evaluation of the resultant semantic structure with its fully general representational capacity will justify addition of some more semantic objects whose syntactic correspondents the syntax cannot generate with the given set of rules.

I wait for another occasion to provide a full review of Braine and O'Brien's analysis. In this section, I have added some speculative comments about alternative ways of interpreting deductive systems.

6.3 Multi-modal inference system.

As I mentioned in section 4 (footnote 35), Braine and O'Brien divide mental logic schemas into different categories. The basic ones (such as MPP) apply automatically. Some others (including &I) are only applied if their output will feed one of the basic ones. They claim that this solves the alleged infinity problem. For example, one cannot apply &I as in $P \vdash P\&P$ unless this application feeds one of the main schemas, such as MPP, avoiding the alleged infinity in (11a) in section 4.[54] I do not review this proposal in detail in this paper, but there are several problems. First, as I show in the main text, the alleged infinity, even if it exists, is not caused by introduction rules themselves. Thus controlling the use of introduction rules does not solve the problem in a complete way (without further stipulations that make the system even more complex). Also, dividing the rules for logical connectives into different groups is dangerous, because there are certain derivability relations among these rules and separating them into groups with restricted feeding relations risks making the system incomplete not only in each group but at the level of the whole system.[55] Compare their proposal with the informal suggestion in section 4 in which one controls structural rule application. To require that one may apply structural rule only if the output feeds into a logical rule is less harmful in several ways. First, the division between logical rules (i.e. rules for logical operators connectives, such as &, $\vee$, and $\rightarrow$) and structural rules (i.e. weakening, contraction, etc.) are already there in logic. There are several

---

[54] Without further restrictions, Braine and O'Brien's system does not solve the alleged overgeneration with enrichment as in (2)~(4) because in that case, the output of &I does feed MPP. But use of &I in relevance theory has not been one of their concerns.

[55] For each connective, having the elimination rule without the introduction rule is problematic, as I have shown in the paper. This is the same (in a slightly different way) if the introduction rule and the elimination rule are put into different groups of rules. Also, consider the equivalence between $(P \rightarrow Q)$ and $(\neg P \vee Q)$. Given equivalence relations like this one, putting the rules for $\rightarrow$ in one group, and the rules for $\vee$ in another, restricting the use of the latter, then also risks making the system incomplete with regard to the intended interpretation.

diagnostics that one may use for telling the differences between them. For example, simply count the number of connectives before and after each rule application. Logical rule application necessarily influences the number of connectives (or, equivalently, it influences the complexity of the formulas or the structured configurations of formulas). On the other hand, application of a structural rule in itself does not influence the complexity of the formulas.[56] It is not only that logical rules and structural rules are different in nature. See the footnote 36 for an observation to the effect that logical rules are independent of the structural properties of the system. This independence of logical rules from structural properties allows us to limit the application of structural rules in the way that we explained in section 4. In fact, there are several established proofs that show that applying a structural rule only if the output feeds into some logical rule does not influence the set of derivable (or provable) sequents (see Hudelmaier 1996, for example, in this regard). Thus, we have some confirmation that restricting the application of structural rules in this way to avoid the alleged overgeneration does not influence the derivability of sequents (or, semantically, the validation of arguments) in the inference system. If the original system without such control of structural rule application is complete with regard to the intended interpretation, then the same system with such control is also complete.

I leave further evaluation of Braine and O'Brien to further research.

6.4 Truth-based judgment in spontaneous inference

In (7) (repeated here as (24) below) in section 3, I argued that the spontaneous system at the basic level should be equipped with the rule which can directly support the truth functionally equivalent role that p, q on the one hand, and $(p\&q)$ on the other, play in the Antecedent of a sequent.

(24) a. $p, q, (p\&q)\rightarrow r \vdash r$
b. $p\&q, (p\&q)\rightarrow r \vdash r$

Carson (p.c.) claims that it is not clear why this equivalence is something that the spontaneous inference system should be expected to explain. I agree that it is

[56] Contraction of $P\&Q$, $P\&Q \vdash X$ to $P\&Q \vdash X$ may seem to influence the number of connectives, but this reduction of the complexity of the form is not really because of the structural rule application itself. Note that from $P\&Q$, $P\&Q \vdash X$, one may first eliminate two occurrences of & via &E, producing, $P, P, Q, Q \vdash X$. After that, one can apply contraction, producing, $P, Q \vdash X$. Based on the assumption that different routes to reach the proof of the same sequent actually represents the same proof, we can claim that structural rules do not influence the complexity of the structured configurations of formulas

debatable whether we can directly recognize the same semantic roles that *p*, *q* separately and *p*&*q* together play as premises in our spontaneous truth-based inference at the data level. However, I am not discussing the thing with regard to the semantics of the inference data only. I am also evaluating the spontaneous inference system with regard to whether it does its job in a complete way as a well-defined deductive system. A system that lacks &I but with &E cannot do &I in the syntax by stipulation, but the intended semantics (if the Boolean semantics as I suggested is the intended semantics) predicts that the system can validate that syntactically impossible sequent in the semantics. Thus, the system is inconsistent between the verdict in the syntax and the (contradicting) verdict in the intended semantics. They could provide an alternative semantics as the intended semantics so that this inconsistency can be resolved, but for the moment, I find it hard to come up with such an alternative which matches with their syntax in a complete way. Also, even if they could successfully provide the semantics with regard to which their suppression of &I from the inference system is complete, that alternative would mean that we could only recognize the above mentioned equivalence in the roles played by *p*, *q* and *p*&*q* only in an indirect (or reflective) way as I explained in section 3. I am not sure if I feel as if my own recognition is only indirect in my spontaneous inference. Given that it makes the definition of the intended semantics far more difficult, I argue that it is better to equip the inference system with &I at the base level, and explain why &I is not used in our spontaneous inference in certain cases for independent reasons. For example, as I have sketched in 6.1 and 6.3, with multi-modal interpretations of the inference language, we may recognize the difference between *p*, *q* and *p*&*q* as premises in a mode of interpretation that is different from the truth based one. This may explain why we feel as if there are differences between these two at the level of intuitions.

7 Conclusion

If a pragmatic inference system is to explain one's truth-based inference (possibly among other kinds of inference), it is not desirable to eliminate logical introduction rules completely from the inference system, with view to preserving the consistency of the system as a whole. Use of introduction rules in the inference system as a whole does not lead to overgeneration via enrichment. Introduction rules can apply only with fully propositional elements as premises, and thus, such rules cannot be applied before the recovery of the proposition expressed. The alleged infinite inference steps are not caused by introduction rules per se, and the problem must be solved independently.

References

Avron, Amon. (1993). Gentzen-Type Systems, Resolution and Tableaux. *Journal of Automated Reasoning,* 10, 265-281.

Belnap, Nuel. (1996). The Display Problem. In Heinrich Wansing (ed.), *Proof Theory of Modal Logic*. Dordrecht: Kluwer, 79-92.

Carston, Robyn. (2002). *Thoughts and Utterances: the pragmatics of explicit communication*. Oxford: Blackwell.

Došen, Kosta (1988). Sequent systems and groupoid models I. *Studia Logica,* 47, 353-389.

Girard, Jean-Yves. (1987). *Proof theory and logical complexity*. Amsterdam: Elsevier.

Girard, Jean-Yves. (1989). *Proofs and Types,* Translated and with appendices by *Paul Taylor* and *Yves Lafont.* Cambridge: Cambridge University Press.

Hall, Alison. (2006). Free enrichment or hidden indexicals. *UCLWorking Papers in Linguistics* 18:71-102.

Hudelmaier, Jörg. (1996). A contraction-free sequent calculus for S4. In Heinrich Wansing (ed.), *Proof Theory of Modal Logic*. Dordrecht: Kluwer, 3-15.

Johnson-Laird, Philip N. (1997). 'Rules and illusions: a critical study of Rips' "the psychology of proof."' *Minds and Machines* 7.3:387-407.

O'Brien, D.P. (2004). 'Mental-logic theory: What it proposes, and reasons to take this proposal seriously.' In J. P. Leighton & R.J. Sternberg (Eds.), The nature of reasoning, Cambrdige: Cambridge University Press, 205-233.

Partee, Barbara and Mats Rooth. (1983). 'Generalized conjunction and type ambiguity.' In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow (eds.), *Meaning, Use, and Interpretation of Language*. Berline: de Gruyter, 361-393.

Rips, Lance. J. (1997). 'Goals for a theory of deduction: reply to Johnson-laird.' *Minds and Machines* 7.3:409-424.

Sperber, Dan. and Deirdre Wilson. (1986/1995) *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stanley, Jason. (2000). Context and logical form. *Linguistics and Philosophy* 23:391-434.

Stanley, Jason. (2002). Making it articulated. *Mind and Language* 17 1& 2:149-168.

Takeuti, Gaisi. (1987). *Proof Theory: Studies in logic and the foundations of mathematics, Vol 81*. North-Holland: Elsevier Science Ltd; 2nd edition.

Wansing, Heinrich. (1993). *The Logic of Information Structures*. Lecture Notes in AI 681, Berlin: Springer-Verlag.

Wansing, Heinrich. (1998). *Displaying Modal Logic,* Dordrecht: Kluwer.

Appendices

Contents of UCLWPL *volumes 1-18*

Volume numbers in parentheses.

Ordering UCLWPL

Ordering whole volumes
You can order individual copies of *UCLWPL* up to and including volume 18 (2006). Price per copy: £12.50 (students), £17.50 (others).

Ordering off-prints
You can order off-prints of individual papers. Price per copy: £2.50.

Subscribing
You can take out an annual subscription.

Exchange
We can arrange exchanges with other publications.

Payment
All prices are in Sterling and include postage and handling. Please pay by cheque (UK only), Eurocheque or International Money Order, made payable to *University College London*.

Contact

Molly Bennett
Department of Phonetics and Linguistics
University College London
Gower Street
London WC1E 6BT
UK

e-mail: molly@phon.ucl.ac.uk