

What phonology has learnt from Chinese

Moira Yip

University College London

Although Chinese phonologists made some of the earliest known observations about the internal structure of the syllable, there was a long period during which Western phonologists (with the exception of historical Sinologists) viewed Chinese as rather uninteresting. The focus of theorists was on alternations, and Chinese has few non-tonal alternations. Nonetheless, it turns out that once one digs deeper there is a wealth of interesting phenomena in Chinese that has helped illuminate a variety of areas of phonological theory. This is even more true once one looks beyond the well-known urban dialects like Mandarin, Cantonese, Shanghai, and Taiwanese, and makes use of the extensive body of descriptive fieldwork on the dialects. This is mostly published in Chinese, but native-speaker linguists have been branching out beyond the prestige dialects and showing a greater willingness to work on more rural dialects.

I divide my remarks into four main sections. In the first section, I look at areas in which Chinese data have been absolutely central to our understanding. This section includes, not surprisingly, tone, but also the syntax-phonology interface, and abstractness. Section two looks at areas where Chinese data has helped extend, clarify or refine details, but where the main structure of the theory does not depend on Chinese data. Section three identifies areas that look ripe for future research. Section four discusses how our understanding of Chinese has been enhanced by theoretical advances. There is an extensive bibliography, but only of papers published in English, and then mainly books or papers in general linguistics journals, because only work that is read by the larger community has been able to influence the thinking of the field as a whole.

1. Major contributions: where only Chinese has the necessary characteristics, or where it has them in a particularly noticeable and clear form.

1.1. Tone

1.1.1. Historical development

Chinese has several properties that make it a particularly fertile ground for studying tonogenesis and tonal evolution. First, it consists of a huge number of historically related dialects that are almost all currently tonal. Second, it has a long written tradition that makes the tracing of related forms, and thus reconstruction of earlier stages, much easier. Third, it has its own internal scholarly linguistic literature in which tones are explicitly described and discussed as far back as the seventh century in the famous Qieyun pronouncing dictionary. Fourth, there has been contact with other unrelated languages such as Japanese and Korean over many centuries, leaving borrowed forms in those languages and descriptions of Chinese tones by foreign scholars as early as the ninth century. No other language family has all of these properties. As a result, twentieth century researchers such as Haudricourt, Mei, Pulleyblank, Ting and many others largely agree on two things. The Chinese tonal system arose in the first place through the loss of syllable-final consonants, giving four tonal categories. Subsequently, a voiced/voiceless contrast in onset obstruents has been lost in many dialects, and converted into a tonal register contrast, with the reflexes of syllables with historically voiced onsets being lower in pitch. The four tonal categories can thus give rise to as many as eight tones. There is considerable disagreement over the timing of these changes, but little disagreement over the essentials. See Chen (2000:4) for a summary. These discoveries have illuminated our understanding of the relationship between tonal and laryngeal features in unrelated languages, including the influential work of Halle and Stevens (1971), Hombert, Ohala and Ewan (1979) and Kingston and Solnit (1988). It has also helped

provide a larger context in which to understand languages (fairly common in Africa) in which voiced obstruents either interfere with H-tone spreading or act to lower neighbouring high tones (the so-called depressor consonants). Foreshadowing the next section, the historical Chinese facts have led to proposals such as those of Halle and Stevens (1971), Bao (1990a) and Duanmu (1990) that identify one or more tonal features with the laryngeal features responsible for voicing in obstruents. For example, [+ slack vocal folds] has been proposed as the feature responsible both for low tone and obstruent voicing.

1.1.2. The features of level tones

The prevailing distinctive feature systems for tone are based heavily on data from Chinese. The reason is obvious: a distinctive feature system is supposed to be capable of characterizing all and only the tonal contrasts found in natural languages. It must thus be tested out on the richest possible systems. Languages with a simple H/L contrast and no underlying contour tones tell us little. Many of the tonal systems of Africa are of this type, although there are certainly some with three and occasionally four levels. In China, though, rich tonal systems are the norm, not the exception, and a large number of dialects have available published descriptions of their tonal inventories and often complex sandhi alternations (although the published data is largely in Chinese, and we have instrumental data on far too few of them).

Ever since Chao Yuan Ren in 1930 devised his notation system using digits 1 for low and 5 for high, it has been agreed that four and perhaps five levels must be handled by any adequate feature system. At a minimum, then, we need two binary features for tone, as proposed by Yip (1980), hierarchically arranged so that all possible combinations exist. One feature splits the pitch range into two registers, and the other feature subdivides each register into two tones. For more recent discussion see Fox (2000) and Yip (1995, 2002).

One phenomenon little reported in Chinese dialects, but which any feature system must be able to handle, is downstep, so here the Sino-linguist must defer to the Bantuist and the Meso-Americanist, and all must work together to design a truly universal feature system. See Hyman (1993, in press) for discussion.

1.1.3. Treatment of contour tones

The richness of Chinese tonal inventories is especially striking when it come to contour tones. Quite often, in fact, they have more contours than levels. Fuzhou, for example, has two levels and four contours on sonorant-final syllables. This has made them central to theoretical discussion on how to treat contour tones. Are they single units, with features such as [+ rising], or are they clusters of level tones occurring on a single syllable? Early writers, such as Pike (1948) and Wang (1967), considered Chinese contours to be indivisible units, quite different from contours in African languages. With the advent of non-linear representations (originally developed almost entirely on the basis of African data (Goldsmith 1976)) it became possible to represent syllables which hosted more than one tone, and suddenly the Chinese data provided confirmation of this power: this is what languages look like when they allow such multiple tonal specifications rather freely. Closer examination of Chinese tonal systems uncovered morphemes consisting solely of tone that could dock onto a neighbouring segmental host and create a tonal cluster - a contour. Conversely, a contour tone could break apart and attach itself to a sequence of two syllables, spreading out the rise or fall over a larger domain. Suddenly Chinese tone languages looked a lot like African tone languages.

The research of the last 20 years has continued to bridge the supposed gulf between systems in which contour tones were thought to be units, like Chinese, and ones in which they were thought to be derived from sequences of levels, like most Bantu languages, so that now the majority view would be that all contour tones can be decomposed into levels. Despite this, many contour tones display some residual unitary characteristics, as Bao (1990a, 1999) points out, so that the entire contour may spread as a unit in some cases (Chan 1991). For example, in Danyang, HL tones spread to a sequence of syllables, giving patterns like HL.HL.HL.LH, where the dots mark syllable boundaries. This sort of behaviour leads to a more nuanced view, in which contour tones in some languages may consist of two tones grouped under a higher node which can itself spread, so that they are at once both units and clusters, rather like segmental affricates (Yip 1989b). See Duanmu (1994a) for a different view.

Recently Zhang (2000b) looks closely at the phonetics of contour tones, and argues that many of the properties typically thought to be purely phonological are driven directly or indirectly by the phonetics. Discussion of his work is postponed till section 2.6.

1.1.4. Tone Bearing Units (TBU's) and tonal domains

Chinese contour tones have also helped illuminate the notion of tone-bearing unit (TBU). There are two areas in which this is so. First, in many Chinese languages the so-called Ru-sheng syllables which are closed by obstruents or glottal stops can only bear level tones. Open syllables, which always have long vowels, and sonorant-final syllables may have level or contour tones. This has led some researchers to suggest that stop-final syllables are mono-moraic, and all other syllables are bi-moraic, and that the mora is the TBU. Contour tones are two-tone sequences, and thus need two moras (Duanmu 1993). In its strongest form, Duanmu argues that the mora is cross-linguistically the only TBU, and that putative cases where the syllable is the TBU are open for re-analysis.

The second area in which Chinese offers evidence for the TBU is higher up the prosodic hierarchy. Duanmu (1993, 1999a) makes a distinction between syllable-tone and word-tone dialects of Chinese. In the former, each syllable has a tonal specification. In the latter, each word does. Duanmu argues that this distinction follows from facts about syllable-weight in the two types of language, in the following way. Only stressed syllables may keep their underlying tones. If syllables are bi-moraic, each syllable is a foot, and bears stress, and thus keeps its own tone. These are the syllable-tone languages, like Mandarin (although Mandarin does of course also have some mono-moraic syllables, called 'neutral toned', and which are unstressed and toneless as expected). If all syllables are light, only the one which bears word stress is stressed. Thus all other syllables lose their tones, and each word has only one tone. These are word-tone languages, like Shanghai. The claim that Mandarin syllables are heavy while Shanghai syllables are light is confirmed phonetically in Duanmu (1994). The general claim is thus that cross-linguistically there is no need for a parameter set at word-tone or syllable-tone: the distinction falls out from other properties of the languages involved.

1.2. Prosodic phrasing and the syntax-phonology interface

Data from Chinese has brought real insights into the syntax-phonology interface. Many Chinese languages have 'tone sandhi' rules in which the tones of one word change either (i) in some particular prosodic environment, such as non-finally or (ii) next to some other specific tone, provided both are in the same prosodic unit.

1.2.1. Syntactic conditioning of prosodic structure

Although the details are complex, the units within which these changes happen are at least partly defined by the syntax. Chen (1987) showed, for example, that in Xiamen the right edge of X” defined the end of a prosodic phrase, and that tones at the end of this phrase remained unchanged, while all others changed to a different allotone. Interestingly, he showed that only certain aspects of syntactic structure seemed to be accessible to the phrase-construction algorithm. The distinction between lexical categories and maximal projections mattered, but the lexical category itself was irrelevant: N“, V“, A” all behaved the same. The one additional piece of information that seemed to matter was the adjunct/argument distinction, but Lin (1994) builds on Chen’s work, showing that this distinction could be reformulated in terms of lexical government. Based on these results, any grammatical theory must allow for information from the syntax to be passed, at a minimum, to the prosodic portion of the phonology. For discussion see Inkelas and Zec (1990), especially Chen (1990), and Selkirk and Shen (1990) on Shanghai.

The interaction is quite intricate, since purely prosodic considerations can over-ride the syntactic influence. In Mandarin, for example, the domain of the famous third-tone sandhi rule does not usually cross the NP VP boundary, but domains are also minimally bi-syllabic, so if the NP is monosyllabic it joins the same domain as the VP, and the sandhi can apply. Focus also changes the domain structure (Shih 1986, 1997). It is therefore necessary either to allow for restructuring of some kind, or for the initial domain formation to be simultaneously sensitive to both syntactic and prosodic information. The former approach implies a derivation, something not allowed in current theories like OT. The latter approach implies that both syntactic and prosodic structures are visible at the same time.

1.2.2. The prosodic hierarchy

Chinese data has also posed some unresolved issues for the levels of the prosodic hierarchy. In most standard views (Nespor and Vogel 1986), phonological words (PhW) are grouped into Phonological Phrases (PhP), which in turn are grouped into Intonational Phrases. Each level is properly contained in the next higher level. Chen (1987) shows that Chinese data pose problems for this view. For example, the domains within which Min tone sandhi applies are larger than the word, but smaller than the utterance, so one would naturally expect them to be PhP’s, but it turns out that an intonational break - i.e. an Intonational Phrase boundary - does not block sandhi. Chen thus gives the sandhi domains a neutral name, Tone Group (TG), but it is not clear how to incorporate this into a cross-linguistic view of sentence-level prosody. I should note that in later work Chen (2000:413) concludes (for different reasons) that intonational phrasing is not dependent on PhP’s, and this leaves him free to conclude that after all the sandhi domains are indeed PhP’s.

1.3. Abstractness:

1.3.1 The bi-uniqueness problem:

Chao (1934), writing soon after Bloomfield’s *Language* came out, wrote a highly influential paper reprinted in Joos (1957) which cast doubt on a fundamental aspect of phonemic theory. He showed that there was no unique phonemic analysis for certain types of Chinese data. In Mandarin, for example, the palatal series of consonants is in complementary distribution with three other series (velar, alveolar, and retroflex), and thus the choice of which phoneme they are allophones of is under-determined by the facts. Unusually for his time, he also emphasized the

importance of phonetic details in transcription and analysis. He was the first linguist to bring Chinese data to the attention of the rest of the theoretical linguistics community, and he was able to do this because he recognized the significance of that data to the issues of the day, and because he wrote in the language and terminology of the field at large.

1.3.2. Lexicon optimization

One property of Chinese that is superficially a handicap for any ambitions it might have to play a role in developing phonological theory is its well-known near-absence of segmental alternations. The relative paucity of affixation combined with little segmental interaction across compound or word boundaries means that most morphemes are invariant in pronunciation (although tonally of course alternations are common). Despite this, rule-based theories, which place a high premium on lexical economy, often analyzed these invariant morphemes as arising from very abstract underlying forms. For example, Mandarin has 15 surface vowels, each with a fairly limited distribution but Wu (1994) derives these from only 4 underlying vowels, with the surface forms resulting from the influence of the other segments in the syllable. With the advent of OT such approaches became much more controversial. Yip (1996) shows that abstract underlying forms may require complex rule systems in which, suspiciously, co-dependent features may require both rules in which feature A is derived in context of feature B, and also rules in which feature B is derived in the context for feature A. For example, in Cantonese, length, height and ATR are intimately connected. In open syllables, vowels must be long, and if necessary ATR is adjusted accordingly, as can be seen productively in the final vowel of the loanword ‘*number*’ [ɬmpa:]: [-high, +long] → [-ATR]. Before velar codas, non-low vowels lax, and length must then be adjusted, as can be seen productively in the contracted form of /mei tsʰʌŋ/ > [mɛ:ŋ]: [-high, -ATR] → [+long]. In contrast non-derivational output theories, in which underlying forms are (near) identical to surface forms, need only feature co-occurrence output requirements:

$$(1) \quad \begin{array}{cc} *_{\mu} & *_{\mu} \mu \\ | & \backslash / \\ [\alpha hi, \alpha ATR] & [\alpha hi, -\alpha ATR] \end{array}$$

For this and other reasons, she concludes that lexical economy coupled with elaborate rules should be abandoned in favor of more surface-like underlying representations as predicted by the Lexicon Optimization principle of OT (Prince and Smolensky 1993).

2. Areas where Chinese facts have enlarged or refined the theory:

2.1. Reduplication and affixation:

Chinese is rich in reduplication, which typically reduplicates a chunk that is co-extensive with the syllable and the morpheme (the two cannot always readily be distinguished). In most cases it looks more like self-compounding than affixation, and it is common for some portion of the input to be replaced by fixed segments. Affixation is much rarer, and often phonologically uninteresting, but there is a nominal suffix that in Beijing is a retroflex /-r/ and that has survived in quite different forms in other dialects. This is known as Er-suffixation, and is of real theoretical interest.

2.1.1. Templates

The size of the reduplicative template in Chinese is usually uncontroversially a syllable if viewed

as a prosodic entity, or a morpheme if viewed as a morphological entity. Some types of reduplication copy the entire word, so that bisyllables copy as units. In Mandarin an AB bisyllable becomes either ABAB in verbal reduplication or AABB in adjectival reduplication, depending on whether the word or the syllable is the base. Occasionally, mainly in onomatopoeia, we find triplication, which might require a bi-syllabic affixal template. By comparison with other reduplicative systems, this is all straightforward. Er-suffixation is more interesting, in that it imposes an output requirement that the suffix join the preceding syllable, rather than forming its own syllable: /*pan+r/* → [*par*], *[*pa.nər*]. Theoretically, this demands a template to which the root and affix must both map, competing for space in the single post-nuclear position. Lin (1993) points out that since the affix always wins at the expense of any root coda, the theory must include a principle that requires affixes to be realized. Such a principle has since become a standard tool in the OT constraint inventory. See Samek-Lodovici (1993), Akinlabi (1996) among others. Of course, some languages violate this principle (e.g. English '*I hit the ball every day/yesterday*'), but if the principle is a violable constraint, as in OT, this is as expected.

2.1.2. Fixed segments and The Emergence of The Unmarked (TETU)

A further area of interest in reduplication is the large number of cases where fixed segments are substituted for base segments. This is very common in language games and onomatopoeia, but is also found in the core language. In Fuzhou, verbs reduplicate to give a brief, casual, durative meaning, and /i/ substitutes for all vocalic material in the first syllable: *koŋ* → *kiŋ kwon*. There are two main competing analyses for these data. Either they are specified affixal segments that supplant base segments, much like the /-r/ affix does, or they are unmarked segments that emerge in reduplication in an attempt to reduce markedness, the TETU phenomenon. The first analysis is the more usual one (Bao 1990b, Chiang, 1992), and the second is a more recent development made possible by the insights of Optimality theory. Two facts make a TETU analysis appealing. First, overwhelmingly these segments are Coronal for consonants and the Coronal /i/ for vowels, and of course Coronal is known to be the least marked place of articulation. Second, in Shanghai *only* place contrasts are lost, while laryngeal contrasts of aspiration and voicing are retained: *ka* → *ka ta*, *ga* → *ga da*, *kʰa* → *kʰa tʰa*. Within OT, such phenomena are predicted to occur as result of constraint interaction. If a language has marked segments such as labials and velars in its inventory, then in OT input-output faithfulness, including IDENT-IO, must outrank markedness, such as *LAB, *DORS. If such segments fail to be copied in reduplication, however, it must be the case that the markedness constraints outrank base-reduplicant faithfulness, IDENT-BR. So IDENT-IO >> *LAB, *DORS >> IDENT-BR >> *COR, and we arrive at the right result. See Yip (2001) for details.

2.2. The Obligatory Contour Principle (OCP) and constraints as rule triggers

Although the OCP, which bans sequences of two identical elements, was first proposed to deal with tonal phenomena (Leben 1973), and although Chinese languages are tonal, their main contribution to our understanding of OCP effects has been in the segmental domain. Yip (1988) uses Cantonese OCP effects involving multiple labials as a central case in arguing that the OCP operates as a general well-formedness condition in the grammar, acting simultaneously as Morpheme Structure Condition (MSC), rule blocker and rule trigger in a single language. Subsequent work has confirmed that the OCP may be satisfied in several different ways in one language, as we would expect if it is a high-ranked output constraint rather than a felicitous by-

product of a rule conspiracy. See for example Myers (1997).

2.3. “Covert” prosody and binarity effects:

In stress languages, evidence for prosodic structure abounds. Chinese languages, however, are often reported not to have stress. See Duanmu (1995, 2000, 2001) and Chen (2000: 286) for useful discussion. It is therefore much harder to decide whether they construct prosodic constituents at any level at all. If it can be shown that they do, it suggests that prosodic structure is universally present, but is not necessarily realized as stress prominence. It turns out that there is considerable evidence for prosodic structure, even in Chinese.

The most convincing evidence for prosodic structure is surely binarity effects. In stress languages this shows up as such things as alternating stress patterns and iambic lengthening. In Chinese it surfaces instead as binary influences on tone sandhi domains and preferences for bimoraic syllables (as in Mandarin, see §1.1.4) and bisyllabic words.

Shih (1986, 1997) was the first to note that the domains in which Mandarin third-tone sandhi applies imply the avoidance of mono-syllabic left-overs, or, to put it another way, are minimally binary, as already discussed in section 1.2.1.

Duanmu (1995, 1997) on Shanghai goes further, showing not only that binary domains exist, but that they are left-headed feet. A tri-syllabic input composed of a bisyllable plus a monosyllable is footed ($(\acute{\sigma} \sigma)(\acute{\sigma})$), as shown by the fact that the first and third syllables keep their tones, but the second syllable loses its tone. A trisyllable with a monosyllable followed by a bisyllable is however not footed $*(\acute{\sigma})(\acute{\sigma} \sigma)$, but $(\acute{\sigma} \sigma \sigma)$, and Duanmu attributes this to an illicit stress clash between the first two syllables, repaired by merging the two feet into one.

The lexicon also shows binarity effects, with a huge preference for disyllabic words in Mandarin - 70% according to Duanmu (2000:146) - and for quadrisyllabic idioms containing exactly two binary feet. In addition, we see binarity effects in morphology. Yip (1994) shows that Cantonese familiar names and kinship terms must be bisyllabic, but this can be achieved in one of three ways: lexically, as in *ku tse* ‘father’s younger sister’; by reduplication, as in *p^ho p^ho* ‘grandmother (mother’s side)’; or by the use of a filler syllable *a yi* ‘mother’s younger sister’.

Even lexical insertion is influenced by prosody: some words have two synonymous forms, one monosyllabic and one bisyllabic. The choice of form for such words is determined by word position. Since non-heads have stress, in a V-O construction the non-head object cannot be ‘lighter’ than the head verb, so if the verb is bisyllabic the object cannot be monosyllabic. See Duanmu (2000:160), also Feng (2002).

As an aside, there is an interesting question as to what some of these bisyllabic units are. Although they are often referred to as feet, this is probably wrong. Consider Mandarin. Mandarin has reduced toneless syllables, the so-called neutral tone. They can only occur after full syllables, and the natural assumption is that a grouping of a full plus one or more neutral syllables constitutes a trochaic foot. But in that case a grouping of two full syllables must be something else. This suggests that the binarity requirement holds at a higher level than the foot level, and that binarity should be separated from foot-form. Of course, this does not mean that true feet may not also have to be binary. In Mandarin, as we saw in §1.1.4, the minimal foot is one bimoraic syllable (Duanmu 1999b). So in a single language we have heavy syllables as bi-moraic feet, and Minimal Words as bisyllabic units of some larger kind. Duanmu points out that this co-occurrence of both mora counting and syllable counting in one language goes contra the claims of McCawley

(1978) and that it is a sort of covert quantity sensitivity, supporting Kager (1992).

2.4. Cyclicity in OT:

The large literature on cyclic effects in phonology has been almost entirely based on word-internal cycles, and thus on data from languages with rich morphology. Recently, however, Chinese has also proved illuminating. Cyclicity remains an important issue because the standard analyses were couched in a derivational framework, with rules reapplying on multiple cycles, but in non-derivational one-step theories like OT no such mechanisms are available. With this in mind, Kenstowicz (1995) and Duanmu (1997) examine cyclic effects in Shanghai compounding. The data are especially interesting because compounding is so productive in Chinese. It is possible to look at both left-branching, right-branching and mixed structures, and at multiple levels of embedding, whereas affixation tends to be either prefixing or suffixing (and less often both), and often with an upper limit to the number of morphological cycles.

Duanmu uses data from compound stress in Shanghai (manifested by retention of underlying tones on heads), and shows that it can only be understood if there is an output-output constraint requiring that the stress locations in the immediate constituents of a compound are the same as when the constituents occur alone. For example, a six-syllabled compound consisting of a monosyllable (ó) plus a five-syllabled word (ó σ)(ó σ σ) is metrified (óσσ)(óσσ) rather than *(ó σ)(ó σ)(ó σ)(ó σ), thereby preserving the stress on the third syllable of the five-syllabled word. Interestingly, although Duanmu then shows that this constraint must be evaluated recursively, applying on every layer of morpho-syntactic bracketing, he also demonstrates that this can be done in parallel, rather than sequentially. It is thus compatible with the non-derivational claims of OT.

2.5. Loanword phonology as a probe:

Loanword phonology has proved a useful tool for phonologists, since it shows how the speakers of a language react to unfamiliar inputs, and how they strive to adjust them minimally to their own language. Chinese loanwords are fertile ground, for several reasons. First, there are a lot of them in some dialects, particularly Cantonese! Second, Chinese has a simple syllable structure, typically maximally CVC, with codas limited to nasals and in some dialects stops. If the source language is English, radical adjustments are needed to deal with the much larger and more complex English syllable. The form of the adjustments is illuminating.

2.5.1. The role of perception:

Silverman (1992), using data on English loans into Cantonese, argues that loanword adaptation takes place in two 'scans'. The output of the first scan, the Perceptual Scan, is the input to the second 'scan', the phonology proper. He argues that in the first scan Cantonese hearers may fail to perceive English distinctions that are missing in Cantonese, such as obstruent voicing. Thus the input to the Cantonese phonology for English 'bus' is [pas], not *[bas]. This scan also differentiates between unsyllabifiable and thus vulnerable salient and non-salient consonants, with [s] in clusters being salient but [l] and [r] in clusters being non-salient. Salient consonants are assigned syllable nodes, and then undergo epenthesis in the phonology: 'bus' [pa:si:], 'stamp' [si:ta:m]. Non-salient ones, on the other hand, will delete: 'freezer' [fi:sa:], unless the phonology assigns them a syllable, which it does in order to satisfy bi-syllabic Minimal Word requirements ('cream' [keyli:m], *[ki:m].) This account is in line with well-documented observations about the

inability of hearers to detect some contrasts (such as [l] vs. [r]) lacking in their own language, while also recognizing the need for outputs to conform to the L1 grammar.

Perception is presumably also responsible for the fact that main-stressed syllables are adapted with a high tone, but unstressed syllables are given a mid tone. Under normal intonation, English stressed syllables have a higher pitch than unstressed syllables, and the Cantonese hearer perceives this as tone.

2.5.2. L1 grammar effects:

Yip (1993) went one step further than Silverman, arguing that the phonological component should not be a set of rules for which there was no evidence in L1 itself, but simply the output-based OT grammar of L1 acting on new inputs. The difference can be seen very simply. Cantonese itself has no epenthesis rules. In a rule-based grammar, the loanword facts such as the added final vowel of [pa:si:] 'bus' require a special rule of epenthesis specific to loanwords. In an output-based grammar, on the other hand, Cantonese has a prohibition on fricatives in codas that is part of the L1 grammar, *CODA[CONT]. So long as preservation of input segments, MAX-IO, is high-ranked, the net result will be epenthesis. The reason we observe no epenthesis in L1 is simply that there are no inputs with fricative codas in the first place.

2.6. Phonetics-phonology interface:

In the last five years there has been a resurgence of interest in the phonetics-phonology interface, and Chinese data has played a role in the discussion.

When phonologists look for diagnostics as to whether some process is phonetic or phonological, we have traditionally taken gradient processes to be phonetic, and categorical processes to be phonological. In the case of tone, the available data is often too sketchy for us to be sure of our ground: is the process by which a H.L bi-syllable becomes H.HL phonological spreading, or just phonetic peak delay? One of the few tonal processes that has been studied sufficiently carefully for us to be confident that we can call it phonological is Mandarin third-tone sandhi. The third-tone sandhi rule is clearly categorical, going back to famous work by Wang and Li (1967) and most recently Peng (2000). A third tone, usually low-level non-finally and low-rising finally, changes to a high-rising second tone before another third-tone. The output is usually thought to be indistinguishable from an underlying second tone. Peng looks at the output of this rule from three angles: production, perception, and categorization. From the production perspective, he shows that the output tone is phonetically of *exactly* the same shape as an underlying second tone, but very marginally lower, about 2Hz throughout. From the perception point of view, he shows that hearers cannot reliably distinguish the output of the rule from an underlying tone two, performing at chance, and confirming earlier work such as Wang and Li. Lastly, he conducted an experiment to assess how hearers categorized the tones they hear. Hearers were asked to identify the tone of the first syllable in [yu2san3] 'umbrella', (from /yu3san3/, where /yu3/ means 'rain'; digits stand for the traditional names of the tones). Interestingly, although these output tones were produced almost identically to tone 2, and perceptually indistinguishable from tone 2, they were uniformly categorized as tone 3, showing again that hearers can retrieve the underlying forms of the tones for familiar lexical items even from radical phonological transformations.

The dividing line may be clear enough in the third-tone case, but other data are more intriguing. Zhang (2000a) argues that non-contrastive phonetic details can play a role in

phonological patterning. The data comes from nasalization in Er-suffixation in Beijing Mandarin and other dialects. Recall that /-r/ supplants a coda consonant, so that /pan + r/ becomes [par]. However, /paŋ + r/ becomes [pār], with a nasalized vowel. This has always been mysterious, but Zhang ties it to the observation that the vowel in unsuffixed [paŋ] has significantly longer nasal flow duration than the vowel in unsuffixed [pan]. In the suffixed forms, this nasality difference is preserved. He instantiates this as an OT grammar, in which output-output constraints pay attention to non-contrastive surface nasality. Although the final nasal itself must go, under pressure to realize the affix (see § 2.1.1), its nasality can survive, since this is on the nuclear vowel.

The same author also finds evidence for phonetic influences in phonology in the tonal domain. Zhang (2000b) shows that cross-linguistically, including in many Chinese dialects, contour tones are limited to certain contexts: long vowels, stressed syllables, final syllables, and syllables in shorter words. The common thread here is that all these are contexts in which the sonorous portion of the syllable is longer. He concludes that the phonological licensing requirements should pay attention to phonetic duration, whatever its source, and that in a given language the distribution of contour tones should correlate closely with the duration facts. Again, then, the division between phonetics and phonology is blurred or indeed denied.

The contributions discussed above by no means exhaust the role Chinese data have played in our understanding of human language, but for reasons of space I have had to be somewhat selective. In historical linguistics, I have focussed on tonal development, but comparative dialectal work has also enhanced our understanding of coda decay (Chen and Wang 1975), among other things. Work on speech errors (Wan and Jaeger 1998) remains the only work I am aware of on speech errors in a tonal language. There is a huge phonetics literature on Chinese tone, investigating issues in perception and production such as the relative contributions of pitch, duration and amplitude, and the amount of co-articulation in connected speech.

For a language family often written off as ‘uninteresting’ by phonologists, Chinese, as I have tried to make clear, has made substantial contributions to phonological theory as far back as the seventh century, throughout the twentieth century, and now on into the twenty-first.

3. Areas in which Chinese might make contributions in the future

3.1. Syllable-internal structure: the issue of pre-nuclear glides

The Chinese linguistic tradition noticed as early as the seventh century that the syllable seemed to have internal structure. Undoubtedly this insight was facilitated by the almost complete coincidence of syllable, morpheme, and written character, thus bypassing the need to isolate the syllable from some larger morphological entity. They used evidence from poetic rhyme, and also from the most natural point at which to break a syllable into two parts, and proposed that the syllable was composed of an Initial and a Final, and that the Final contained a medial glide, and a Rime. The famous Qieyun dictionary was constructed in which the pronunciation of each character was denoted by two different characters, one of which began with the same Initial, and the other of which ended with the same Final. For example, /man/ might be entered with the characters for /mei/ and /dan/. See Norman (1988: 24) for a good and accessible summary. As far as I know, their work is the earliest contribution to a theory of the syllable. It remains a very strong influence on Chinese phonologists to this day, but more importantly it can be seen as the intellectual pre-cursor of current Onset-Rime theories of the syllable.

In modern times, Chinese has not yet figured largely in the theoretical debate on syllable-

internal structure, but I include this section because they have one promising property. Chinese syllables are by large rather simple, usually maximally CVC. The one area in which they are of real interest is that a glide may intervene between C and V, creating a sequence of rising sonority, as in Mandarin [tyən]. Cross-linguistically these are not common (Rosenthal 1994, 1997). There is a large literature on the treatment of this glide, and it is variously analyzed as part of the Onset, part of the Rime, not in either, a secondary articulation on the preceding consonant, moraic or non-moraic! For example, Harris (1983) for Spanish and Bao (2000) for Fuzhou locate them in the Rime; Pike and Pike (1947) for Mazateco and Bao (1990) for Mandarin locate them in the Onset; and Duanmu (1990) for Mandarin considers them secondary articulations on the onset consonant. It remains a test case for any theory of the syllable, and is unresolved at the time of writing. See Yip (to appear) for arguments that their inconsistent behaviour within a single language, Mandarin, poses problems for any Onset/Rime approach, but is straightforward under a moraic analysis.

3.2. Acquisition of tone

Acquisition research requires two things: a fairly well-understood adult grammar, and meticulous recording of large quantities of infant and child utterances over the first few years of life. It tends to be carried out in the first instance by native speakers who are professional linguists collecting data from their own children, although the data may then be analyzed by other researchers in subsequent years. Among tonal languages, Chinese has perhaps the largest cohort of native speaker trained linguists, and thus there is more work on acquisition of Chinese than there is of other tonal languages. The focus has been on the early acquisition of tonal contrasts. Here Chinese provides fertile ground for study, because of the richness of its tonal inventories, but it must be said that at present it holds more promise than results. In particular almost nothing is known about the acquisition of productive tonal process such as third-tone sandhi in Mandarin, but see Li and Thompson (1977) and Yue-Hashimoto (1980).

So what *do* we know? From research starting back in the 1970's by Li and Thompson, Clumek and many others, and continuing to the present day with work by Lee, Tsay and others, it appears that children by the age of about two are quite accurate in their tonal production, even in languages with elaborate tonal systems, like Cantonese. At this same stage their segmental production may still be quite distant from the adult norms. Less clear is whether the order in which tones are acquired is cross-linguistically the same. Despite suggestions that level tones are acquired before contours, and high level tone first of all, the evidence is quite confused. One thing is clear: these questions will not be answered in full by looking at languages with inventories limited to two level tones: The flamboyance of Chinese tonal systems is again an essential asset. See Yip (2002) for a summary.

4. How theoretical advances have enhanced our understanding of Chinese phonology

When language data are used successfully to argue for a theory, the corollary is that the new analysis offers a more explanatory account of the data than previous approaches. As a result we simultaneously gain new insights into the language in question. In this section I briefly gather together some of the most important ways in which we have learnt to see Chinese differently as a result of many years of theoretical advances.

I start with tone. Until the 1970's, Chinese contour tones tended to be viewed as indivisible units. Autosegmental phonology allowed us to recognize three things: (i) that they are in fact

composed of levels, just like contours in tonal languages from other areas. (ii) that tones in Chinese too could be mobile, migrating away from their point of origin, and also (iii) that morphemes consisting solely of floating tones explained otherwise mysterious tonal changes. A simple example shows how a floating tone can combine with a level one to produce a contour. In Cantonese familiar forms of names all have high level or high rising tones on the last syllable. If the basic name was high, it stays high. If the basic name was non-high, we get a rise. If it was already high rising, there is no noticeable change. If the names are formed by adding a high tone suffix, all this follows. But without autosegmental phonology, these facts were just called ‘changed tone’, and simply listed.

A second advance came from examining the details of tonal changes. The traditional approach was to list the basic tone, and the sandhi tone, and to view it as wholesale replacement of one tone by another, usually referred to by the name of the tone, and not by its phonetic properties. For example, one would read that the first of a sequence of two ‘third tones’ became a ‘second tone’ in Mandarin. From research on non-Chinese tonal languages, however, we know that tonal changes are usually assimilations or dissimilations, and this knowledge allowed us to realize that the rule can be more perspicuously viewed as the insertion of a H between two L tones, such that the first one becomes rising, and further that this rule is triggered by the OCP. A large proportion of tonal changes (though not all) can now be viewed as phonologically natural, rather than idiosyncratic replacements that are a historical residue.

The second major area where Chinese is now viewed quite differently is its prosody. The importance of prosodic structure in conditioning phonological rules has proved enormously useful in Chinese. Retention of tone on certain syllables versus loss on others can now be seen as prosodically driven, and linked to other facts about the language such as syllable weight, as well as to cross-linguistic facts about vowel reduction. Grouping of syllables into pairs in the application of tonal rules can now be seen as an instance of binary foot construction. Preferences for bi-syllabic words can be seen as requiring a binary minimal word. All of this is new, and required a realization that the absence of an obvious stress system did not mean that strings had no prosodic structure.

Finally, we have learnt to think beyond the syllable. For years the morphology, the writing system and phonological notations in which every syllable was recorded with its own tone obscured the fact that in Chinese, as in other languages, phonological processes may operate over a larger domain. A string recorded as having three syllables with [54 43 21] tones may phonologically have just a H tone at the start, a L at the end, and phonetic interpolation from one to the other. Two advances have made this possible. One is our understanding of the role of prosody, but the other is the role of the phonetic component, such that not every tonal detail requires full tonal specification in the phonology.

The lesson to take from this summary is that the wise phonologist never studies solely his or her own language. Insights gained from theoretical advances instigated by work on quite unrelated languages bring real dividends when applied to one’s own language, and we ignore them at our peril.

References

- Akinlabi, A. 1996. Featural affixation. *Journal of Linguistics* 32: 239-289
Bao, Zhiming. 1990a. *On the nature of tone*. PhD dissertation, MIT.

- Bao, Zhiming 1990b. Fanqie language and reduplication. *Linguistic Inquiry* 21, 317-350.
- Bao, Zhiming. 1999. The structure of tone. New York, Oxford University Press.
- Bao, Zhiming. 2000. Syllabic constituency and sub-syllabic processes. *Journal of East Asian Linguistics*. 9.3:287-313.
- Chan, M. 1991. Contour-tone spreading and tone sandhi in Danyang Chinese. *Phonology* 8.2: 237-260.
- Chao, Y. R. 1930. A system of tone letters. *Le Maitre Phonétique*. 45:24-7.
- Chao Y. R., 1934. The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*. Vol IV, Part 4: 363-97. Reprinted in Joos, M. Ed. 1957. *Readings in Linguistics I*. Chicago: University of Chicago Press.
- Chen, M. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4. 109-150
- Chen, M. 1990. What must Phonology Know about Syntax. *The Phonology-Syntax Connection*. Inkelas, S. and D. Zec, eds. University of Chicago Press. pp 19-46
- Chen, M. 2000. *Tone Sandhi*. Cambridge University Press
- Chen, Matthew, and William S.-Y. Wang (1975). Sound change: Activation and implementation. *Language* 51: 228_281.
- Chiang, Wenyu. 1992. The prosodic morphology and phonology of affixation in Taiwanese and other Chinese languages. Ph.D. Dissertation, U. of Delaware.
- Duanmu, S. 1990. A formal study of syllable, tone, stress and domain in Chinese languages. Doctoral dissertation, MIT, Cambridge, Mass. Distributed by MIT Working Papers in Linguistics.
- Duanmu, S., 1993, "Rime length, stress, and association domains," *Journal of East Asian Linguistics* 2.1, pp. 1-44.
- Duanmu, S., 1994a.. Against contour tone units. *Linguistic Inquiry* 25.4, pp. 555-608.
- Duanmu, S., 1994b. Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology* 11.1, pp. 1-24.
- Duanmu, S., 1995, "Metrical and tonal phonology of compounds in two Chinese dialects," *Language* 71.2, pp. 225-259.
- Duanmu, S., 1997, "Recursive constraint evaluation in Optimality Theory: evidence from cyclic compounds in Shanghai," *Natural Language and Linguistic Theory* 15, pp. 465-507.
- Duanmu, S. 1999. Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics* 8.1: 1-38.
- Duanmu, S., 2000a, *The Phonology of Standard Chinese*. Oxford University Press.
- Duanmu, S., 2001 "Stress in Chinese" In Xu, (ed.) pp. 117-138.
- Feng, Shengli. 2002. Minimal Word in Mandarin Chinese. ROA # 507-0302
- Fox, A. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. Oxford University Press.
- Goldsmith, J. 1976. *Autosegmental Phonology*. Ph.D. dissertation, MIT.
- Halle, M. and K. Stevens 1971. A note on laryngeal features. *Quarterly Progress report* 101.MIT.
- Harris, James 1983. *Syllable structure and stress in Spanish*. MIT Press, Cambridge, Mass.
- Haudricourt, A-G. 1954 De l'origine des tons en vietnamien. *Journal Asiatique*. 242: 68-82.
- Haudricourt, A-G 1972 Two-way and three-way splitting of tonal systems in some Far-Eastern

- languages. C. Court (trans.). In J. Harris and R. Noss (eds.) *Tai phonetics and phonology*. 58-86
- Hombert, J.-M., J.J. Ohala and W.G. Ewan. 1979. 'Phonetic Explanations for the Development of Tones,' *Language* 55, 37-58.
- Hyman, L. 1993 Register tones and tonal geometry. In H. van der Hulst and K. Snider (eds.), *The Phonology of Tone: The Representation of Tonal Register*. (pp. 75_108). Berlin: Mouton de Gruyter
- Hyman, L. In press. Tone systems. In M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, (eds.) *Language typology and language universals: an international handbook*. 2 vols. Berlin and New York: Walter de Gruyter.
- Inkelas, S. and D. Zec (eds.), 1990. *The Phonology-Syntax Connection*. Chicago: University of Chicago Press.
- Kager, R. 1992. Shapes of the generalized trochee. *WCCFL* 11: 298-312.
- Kenstowicz, M. 1995. Cyclic vs. Non-Cyclic Constraint Evaluation. *Phonology* 12.3: 397-436.
- Kingston, J. and D. Solnit. 1988. The tones of consonants. Amherst and Ann Arbor: U. of Massachusetts and U. of Michigan.
- Leben, Will 1973. *Suprasegmental phonology*. Doctoral dissertation, MIT.
- Li, C. and S. Thompson. 1977. The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*. 4, 185-199.
- Lin, J. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11:237-276.
- Lin, Yen-hwei 1993. Degenerate Affixes and Templatic Constraints: Rime Change in Chinese" *Language* 69.4
- McCawley, J. 1978. What is a tone language? In V. Fromkin, ed. *Tone: A Linguistic Survey*. Academic Press, New York, pp.113-131.
- Myers, S. 1997 OCP Effects in Optimality Theory. *NLLT*: 15.4. 847-892
- N. Smith and Wang Jia-Ling, eds. 1990. *Studies in Chinese Phonology*. Foris, Dordrecht.
- Nespor, M. and I. Vogel. 1986. *Prosodic phonology*. Dordrecht, Foris.
- Norman, J. 1988. *Chinese*. Cambridge University Press.
- Packard, J. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.
- Peng, S-H. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. In M.B. Broe and J.B. Pierrehumbert, (eds.), *Acquisition and the lexicon: papers in Laboratory Phonology V*. Cambridge University Press, Cambridge. pp 152-167.
- Pike, K. 1948. *Tone Languages: a technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. University of Michigan Publications in Linguistics, no. 4. Ann Arbor: University of Michigan Press.
- Pike, K. and E. Pike 1947. Immediate constituents of Mazateco Syllables. *I.J.A.L.* 13: 78-91.
- Prince, A. and P. Smolensky 1993 *Optimality Theory: Constraint Interaction in Generative Grammar* Ms., Rutgers University and U. of Colorado, Boulder
- Rosenthal, S. 1994. *Vowel/glide alternation in a theory of constraint interaction*. PhD Dissertation, U. Mass, Amherst.
- Rosenthal, S. 1997. The distribution of pre-vocalic vowels. *Natural Language and Linguistic Theory*. 15.1: 139-180.

- Samek-Lodovici, V. 1993. Morphological gemination. Paper presented at Rutgers Optimality Workshop I, Rutgers University, October 1993.
- Selkirk, Elisabeth O. and Tong Shen. 1990. Prosodic domains in Shanghai Chinese. *The Phonology-Syntax Connection*. ed. by Sharon Inkelas and Draga Zec. CSLI. 313-338.
- Shih, Chilin. 1986. *The prosodic domain of tone sandhi in Chinese*. PhD. Dissertation, UCSD.
- Shih, Chilin 1997 Mandarin third tone sandhi and prosodic structure. In Wang, Jialing and N. Smith, eds., *Studies in Chinese Phonology*. Mouton de Gruyter, New York: 81-124.
- Silverman, Daniel. 1992. Multiple Scansions in Loanword Phonology: Evidence from Cantonese. *Phonology* 9.2.
- Wan, I-P, and J. Jaeger 1998. Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15.3: 417-461.
- Wang, W. 1967. The phonological features of tone. *IJAL* 33.2: 93-105
- Wang, W. and K. P. Li. 1967. Tone 3 in Pekinese. *Journal of Speech and Hearing Research*. 10:629-636.
- Woo, N, 1969. *Prosody and phonology*. PhD dissertation, MIT.
- Wu, Yuwen. 1994. *Mandarin segmental phonology*. Ph D Thesis, U. of Toronto.
- Xu, De-Bao, ed. 2001. *Chinese phonology in generative grammar*. Academic Press, London.
- Yip, M. 1980. *The tonal phonology of Chinese*. Ph D Dissertation, MIT. Published 1990, Garland Publishing, New York..
- Yip, M. 1988. "The Obligatory Contour Principle and Phonological Rules: A Loss of Identity." *Linguistic Inquiry* 19.1:65-100.
- Yip, M. 1989b. Contour tones. *Phonology* 6.1 :149-174.
- Yip, M.1993. Cantonese Loanword Phonology and Optimality Theory. *Journal of East Asian Linguistics* 2.3: 261-293.
- Yip, M. 1994. Isolated Uses of Prosodic Categories. In J. Cole and C. Kisseberth, eds, *Perspectives in Phonology*. , Center for the Study of Language and Information. Stanford, California. pp. 293-311.
- Yip, M. 1995. Tone in East Asian Languages. In J. Goldsmith, ed., *Handbook of Phonological Theory*. Basil Blackwell, Oxford. pp 476-494
- Yip, M.1996. Lexical Optimization in Languages without Alternations. In J. Durand and B. Laks, eds. *Current Trends in Phonology: Models and methods*. CNRS, Paris-X and University of Salford: University of Salford Publications. pp 757-788
- Yip, M. 2001. Segmental unmarkedness versus input preservation in reduplication. In L. Lombardi, ed. *Segmental phonology in Optimality Theory*. Cambridge University Press. pp 206-230.
- Yip, M. 2002. *Tone*. Cambridge University Press.
- Yip, M.. (to appear). The non-existence of the Onset-Rime boundary. *Lingua*.
- Yue-Hashimoto, A.O. 1980. Word play in language acquisition: a Mandarin case. *Journal of Chinese Linguistics*. 8.2: 181-204.
- Zhang, Jie. 2000a. Non-contrastive features and categorial patterning in Chinese diminutive affixation: MAX-[F] or IDENT[F]. *Phonology*: 17.3: 427-478
- Zhang, Jie 2000b. Phonetic duration effects on contour tone distribution. M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim (eds.) *Proceedings of NELS 30, Vol 2*. 775-785.

C:\moira\GLOT\outline2.doc