

*Measuring the processing load of Japanese words**

SO HIRANUMA

Abstract

A theory of processing difficulty (Hudson 1995, 1996b) grounded in Word Grammar (Hudson 1984, 1990) provides a means of computing the syntactic difficulty for a sentence in terms of the distance of dependencies. The purpose of this paper is to determine relative processing loads for Japanese words - content words versus function words - in the hope of advancing the method in such a way that syntactic difficulty will be measured with more accuracy. I shall illustrate through an experiment that in terms of memory cost, content words are essentially more expensive than function words.

1 Introduction

One of the main focuses of work on natural language processing has been the attempt to account for sentence processing breakdown caused by particular syntactic structures such as centre-embedding (processing overload effects), exemplified in (1), and by local syntactic ambiguity (garden-path effects) as shown in (2).¹

- (1) #The frightened child who the old woman who the rescue worker looked for had comforted survived the crash. (Gibson, Thomas and Babyonyshev 1995)
- (2) #The cotton clothing is made of grows in Mississippi. (Marcus 1980)

*I am grateful to Dick Hudson for his guidance and practical suggestions throughout my work.

¹The symbol “#” shows that the prefixed sentences are grammatical but either unprocessable or difficult to process to the extent that conscious reanalysis is required.

Within phrase-structure theories, a number of proposals concerning the causes of processing overload have been suggested with little accord as to the precise reasons why some syntactic constructions are harder to understand than others. Some well-known hypotheses attribute processing overload to the number of incomplete parsed phrase-structure rules (Yngve 1960; Chomsky & Miller 1963; Miller & Chomsky 1963; Miller & Isard 1964; Abney & Johnson 1991), or locally unsatisfied X-bar (Lewis 1993) or Case-assignment (Stabler 1994) or thematic (Gibson 1991) relationships.

Although the terms ‘processing complexity’ and ‘difficulty’ have been used interchangeably in the literature in terms of computing the processing load of a sentence, they should be distinguished from each other. One reason is that processing complexity must refer to objective facts about a sentence, e.g. how complex its syntactic or semantic structure is, whereas processing difficulty is measured in terms of the psychological difficulty which people have when they understand sentences. Furthermore, it is not always the case that the more syntactically complex sentences are harder to process. For example, (4) is easier to process than (3), although it is more complex, with one more word.

(3) #The woman saw the famous doctor had been drinking a lot.

(4) The woman saw that the famous doctor had been drinking a lot. (Sturt 1998)

This distinction between complexity and difficulty is fundamental, so I shall use the term ‘syntactic difficulty’ strictly to refer to the processing difficulty of a sentence or a phrase caused in particular by its syntax. In contrast, ‘syntactic complexity’ refers only to aspects of the syntactic structure itself.

The main question in this research is how syntactic complexity and difficulty are related. Earlier work assumed them to be identical, to the extent of using the two terms interchangeably. For instance, some kind of memory cost can be associated directly with nonterminal nodes (Frazier 1985; Frazier & Rayner 1988). However, difficulty and complexity have been distinguished in more recent work, especially by Gibson (1997), for whom difficulty is caused by integrating an input into the current structure and retaining unintegrated structures in working memory. My own work will also distinguish them.

In this paper, I shall report a preliminary study in preparation for a system which will enable us to estimate the syntactic difficulty of Japanese sentences and ultimately to illuminate the source for processing overload phenomena in Japanese. Head-final

languages, such as Japanese, are vital to investigate in the field of sentence processing, given that a natural language processor is efficient enough to cope with both head-initial and head-final languages. What makes Japanese more interesting besides its head direction is, as pointed by Mazuka and Nagai (1995), its other characteristics such as omission of shared information and scrambled word-order, all of which seem to increase processing difficulty.

The language theory I assume here is Word Grammar (Hudson 1984, 1990) whose roots lie in the tradition of dependency grammar (Tesnière 1959). So far as processing is concerned, the areas that have been treated by Word Grammar are comprehensive: a theory of parsing (Hudson 1996a), a theory of processing difficulty (Hudson 1996b), and NLP implementations applying Word Grammar (Fraser 1985, 1989, 1993; Hudson 1989; Shaumyan 1995). Hudson (1995) also conceived a means of measuring syntactic difficulty, ‘dependency distance’, based on the distance between a word and the word on which it depends. A motive for my assumption of Word Grammar in this research is that the measure seems to make plausible predictions about syntactic difficulty and readability in English (Hudson 1997). This method however, where any given word is weighted with equal processing load, leaves room for improvement, following the Syntactic Prediction Locality Theory by Gibson (1997) which postulates that integration cost (a component for the memory cost) monotonically increases according to the number of new discourse referents. The purpose of this paper is to show a more sophisticated way of measuring dependency distance by suggesting that different words should be associated with different memory cost, which is based on the results of the experiment I have conducted. I shall specifically present the approximate memory load of content and function words in Japanese.

2 Dependency distance

Word Grammar aims to express a syntactic relation between two words solely in terms of dependencies. This is an obvious advantage for a Word Grammar parser, since there is no need for the parser to build higher syntactic nodes or to consider the presence of empty categories. The parser’s only task is to join an incoming word with one of the processed words in working memory. The model of a Word Grammar parser therefore assumes word-by-word incremental processing, where the parser strives to optimise incremental comprehension of an input utterance as it is processed.

In this incremental parsing model, a greater processing load ought to be assigned to a word activated for longer in working memory. This idea is best measured as dependency distance, i.e. the number of words between a word and its parent. Thus the syntactic difficulty of a sentence is measured by the mean distance: the average value of the distances for all words in the sentence. Let us gauge, for example, the mean distances of sentences (5) and (6), where a lower-case number after each word indicates the dependency distance for that word.

(5)

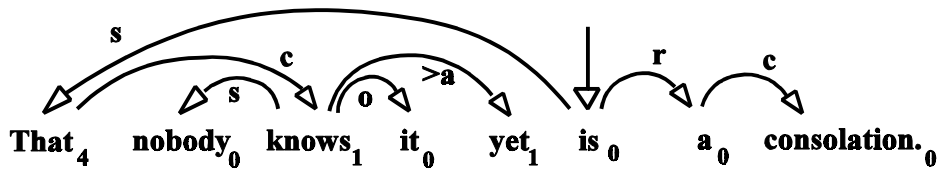


Figure 1

(6)

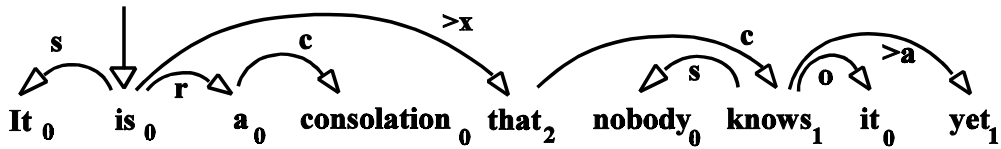


Figure 2

KEY: s=subject; r=sharer; c=complement; o=object;
>a=post-adjunct; >x=extraposee

	Mean distance	Total distance	Number of words
(5)	0.75	6	8
(6)	0.44	4	9

A longer subject phrase such as the clause that nobody knows it yet in (5) should be harder to process since the subject must be kept accessible, in order to be linked with its parent, for a longer period of time in working memory. Dependency distance correctly predicts the relative difficulty of (5) with its mean distance of 0.75 in comparison to the extraposed version (6) whose mean distance is 0.44. Extraposition, which is a source of complexity thanks to appending one extra dependency to a structure, is an alternative route to reduce difficulty. Hence, dependency distance constitutes a very simple and effective scale for calculating the syntactic difficulty of a sentence, although, as mentioned before, it can be improved by charging more precise weights for intervening words within a dependency.

Dependency distance is derived from syntactic analyses based on Word Grammar syntax, which views syntactic structure as dependencies between pairs of words rather than phrasal constituencies. Because of the central role of words, problems arise where word-boundaries are unclear. This approach can therefore not always be applied straightforwardly to languages that have extensive agglutination, such as Japanese. Identifying words in Japanese is also problematic not only because word spaces are not institutionalised but because of abundant special clitics (Zwicky 1977; Pullman and Zwicky 1988) such as particles and auxiliary verbs. According to the status of word for these clitics, the following example can be analysed in two ways: clitics are separate words as in (7) or they are a part of their hosts shown in (8).

(7)

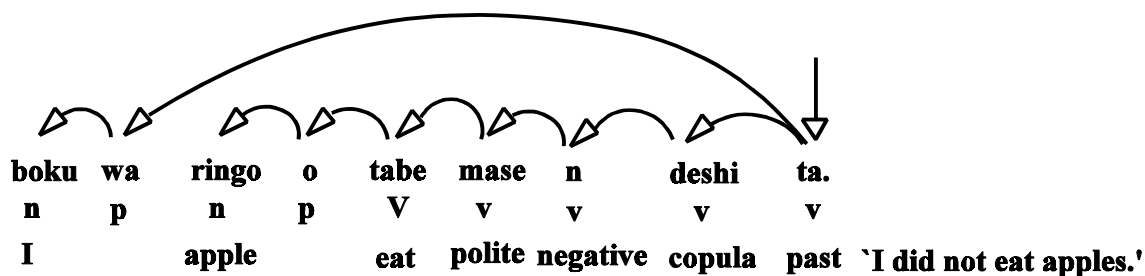
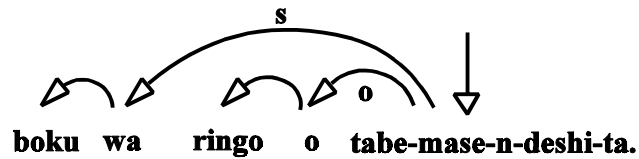


Figure 3

(8)

**Figure 4**

KEY: s=subject; c=complement; o=object; n=noun; p=particle;
 V=verb; v=auxiliary verb

The choice between these analyses is irrelevant to processing itself, though it is an important morpho-syntactic issue. What is needed in terms of syntactic difficulty is a system which allows us to provide the same amount of dependency distance for both (7) and (8), no matter which syntactic structure is assumed. One possible solution to this problem is to assign different processing loads to words or morphs according to their classification.

I now report an experiment that uses immediate recall to explore the relative memory constraints on different kinds of words. Immediate recall is a common technique to test short-term memory retention; Wingfield and Butterworth (1984) for instance, used a similar method in their experiments, where subjects were given control of a pause switch on the tape recorder and allowed to stop the tape-recorded passages whenever they wished to begin their reports.

In this experiment, first, I wished to determine the processing load for content words and function words including particles and auxiliary verbs in Japanese. Second, I wished to verify the processing load calculated from the experiment by applying it to the data in the experiment.

Table 1

order	sorts	μ	characteristics of 12 units
1.	Natural 1	48	grammatically structured natural sentence.
2.	Natural 2	48	grammatically structured natural sentence.
8.	Natural 3	48	grammatically structured natural sentence.
11.	S-Natural 2	48	12 units of 'Natural 2', arranged at random.
3.	Short w 1	48	4N, 4V, 4J, arranged at random.
12.	Short w 2	48	4N, 4V, 4J, arranged at random.
7.	Long w 1	60	4N, 4V, 4J, arranged at random.
10.	Long w 2	60	4N, 4V, 4J, arranged at random.
13.	Short w+p 1	60	4(N+p), 4(V+p), 4(J +p), arranged at random.
4.	Short w+p 2	60	4(N+p), 4(V+p), 4(J +p), arranged at random.
6.	Aux 1	48	4(N+v), 4(V+v), 4(jN+v), arranged at random.
9.	Aux 2	48	4(N+v), 4(V+v), 4(jN+v), arranged at random.
5.	Numeral qf	48	12(d+q), arranged at random.

KEY: N=noun; jN=adjectival noun; J=adjective; V=full verb; p=particle; d=digit; q=classifier; v=one or two auxiliary verbs.

3 Experiment

3.1 Method

Materials. Thirteen sequences (cf. Appendix) were utilised which consisted of three natural Japanese sentences and ten random strings of words. The length of the sequences was constant in terms of the number of morae; the strings were composed of 48 morae save four random sequences of 60 morae.² The consideration of morae was relevant to the experiment, because the number of morae were approximately corresponding to the number of the Japanese *hiragana* or *katakana* letters and subjects' task was to write the sequences down as precisely as possible. Every string contained twelve meaningful units; one unit consisted of four morae for 48 mora-strings, and of five morae for 60 mora-strings. In each unit, there was one content word which might be accompanied by one or more than one function words. I assume that content words are nouns, adjectival nouns, adjectives and full verbs, while function words are particles and auxiliary verbs. Table 1 above shows the characteristics of the sequences used in the experiment. The symbol ' μ ' stands for the number of morae.

The following are remarks on the thirteen sequences:

- 'Natural 1, 2 and 3' were the three syntactically structured sentences, while 'S-Natural 2' was the scrambled version of 'Natural 2', in which the twelve units of 'Natural 2' were arranged in a random order.
- Regarding the two pairs, 'Short w 1' and 'Short w+p 1', and 'Short w 2' and 'Short w+p 2', each pair contained exactly the same content words; of course most of their verbs and adjectives were inflected. The sole difference within the pair was that the 'w+p' versions accommodated one particle in every unit. Hence, the lengths of the strings were longer by twelve morae in comparison to their particle-less counterparts.
- 'Long w 1' and 'Long w 2' had entirely different content words from those of 'Short w 1' and 'Short w 2'. On the other hand, their lengths were equal to those of 'Short

²From a series of mini-scale pilot tests, conducted before this experiment, which I had conducted with several native Japanese undergraduate students at University College London, I had learned that a random sequence of eleven meaningful units with the length of about 45 morae was the limit for pure rote recall.

w+p 1' and 'Short w+p 2'. For the verbs and adjectives of the four 'w' strings, regardless of length, only the citation form was employed.

- There were two strings containing auxiliary verbs, namely 'Aux 1' and 'Aux 2'. Both sequences held thirteen auxiliaries; each unit contained one auxiliary verb except the third units of both sequences that contained two auxiliaries, namely *mashi* (polite) and *ta* (past).
- 'Numeral qf' was the random sequence of twelve numeral quantifiers. Each quantifier consisted of one or more than one digits plus one classifier.
- Some random sequences accidentally contained a string of two units which allowed an interpretation where the preceding unit modified the succeeding one. A common pattern was a string of two units consisting of an adjective followed by a noun. This type of string might be segmented into one chunk, which might encourage a better performance than had been expected on the units forming the chunk. My intention was to avoid chunks, but I failed to achieve it to a certain extent in designing the sequences. I discuss the effects of this chunking in §3.2.

Procedures. This experiment was conducted in the summer of 1996. My subjects were one hundred native Japanese students, aged on average about 20, at the Institute of International Education in London.³ In each test, I read thirteen sequences in the order shown in Table 1 to a group of between 10 to 20 subjects. My instructions to the subjects were that they might write each sequence down as soon as I had read it out, and stressed only accuracy of recall. The subjects were allowed to write in any Japanese letters: two alphabets *hiragana* and *katakana* as well as Chinese characters *kanji*. Arabic numbers were also permitted to be used for digits. The sequences were read in natural pitch pattern and normal reading speed, but each sequence was read only once. Thus the experiment examined the subjects' short-term retention.

The scoring system was based on the twelve units in a string; each unit was scored either as a success or as a failure. Since each sequence had twelve units, the sum of the successes and failures for each string should constantly be twelve. A recall of any unit

³I should like to acknowledge willing permission by Teruyuki Zushi for the conduct of the experiment at the institute.

was a success only if it was a verbatim copy of the original or discrepancies could be attributed to mishearing. Any other response including synonyms and abbreviations of words was counted as a failure; subjects were told to write exact words.⁴ Concerning displaced units, if disarrangement was owing to missing units, only these were counted as failures, and if displacement was caused by a reversal of units, scores were calculated in such a way that the reversed units could obtain as many as successes as possible.

3.2 Results and discussion

3.2.1 General Facts. The raw data comprising the number of failures and successes for each sequence was processed by various chisquare tests to monitor if there was association (non-independence) amongst the parameters on contingency tables. For all of the chisquare tests, the number of failures and that of successes for each sequence, and the sorts of sequences on contingency tables were used as the dependent variables. I shall begin the discussion by commenting on some general facts about the results of the experiment that I have found by counting the number of failures for each unit in all sequences, which are shown in Table 2. Figure 5 below shows the average values.

First of all, it can be noted that the overall accuracy of the performance was moderate; the subjects attained an average of 62.5% units correct. Figure 5 displaying an average value of failures for each unit is comparable to a serial-position curve (Bernstein, Clarke-Stewart, Roy and Wickens 1997: 240-241), where probability of recall is plotted as a function of its serial position in a list of items. The graph shows that comparatively better performance on units 1, 2 and 12, which conforms to primacy and recency effects, i.e. the first and the last few items in a list are most likely to be recalled. One peculiar aspect in this graph however, is a small dip at unit 9, which was confirmed by chisquare tests; there were statistically significant differences between units 8 and 9 and between 9 and 10 ($p=0.000$ in both cases).

⁴I also processed the data of the experiment by counting synonyms and abbreviated words as successes, then put the data to the same chisquare tests. The differences were not significant.

the processing load of Japanese words

Table 2

unit number	u1	u2	u3	u4	u5	u6	u7	u8	u9	u10	u11	u12	total
Natural 1	0	0	8	41	43	43	35	37	15	13	29	42	306
Natural 2	0	0	8	10	30	45	4	6	1	4	19	3	130
Natural 3	1	1	12	44	46	26	13	17	11	23	29	5	228
S-Natural 2	0	10	25	36	48	82	60	65	47	51	43	27	494
Short w 1	1	2	42	67	<u>68</u>	<u>47</u>	90	<u>55</u>	<u>46</u>	86	32	41	577
Short w 2	1	<u>7</u>	<u>11</u>	68	63	46	28	<u>57</u>	<u>36</u>	78	13	14	422
Long w 1	1	4	33	59	61	83	70	72	<u>46</u>	<u>56</u>	29	5	519
Long w 2	3	2	<u>52</u>	<u>46</u>	76	76	87	<u>29</u>	<u>30</u>	70	43	7	521
Short w+p 1	2	14	27	69	82	83	83	64	68	49	56	16	613
Short w+p 2	4	11	43	59	83	85	95	86	<u>9</u>	<u>15</u>	59	26	575
Aux 1	4	11	19	49	58	65	65	<u>40</u>	<u>55</u>	56	44	8	474
Aux 2	14	25	67	71	74	<u>56</u>	<u>61</u>	87	67	50	34	8	614
Numeral qf	2	7	27	47	67	49	54	38	32	24	24	5	376
Total	33	94	374	666	799	786	745	653	463	575	454	207	5849
Average	3	7	29	51	61	60	57	50	36	44	35	16	450

overall accuracy: $(12*13*100-5849)/12*13*100=0.625$

N.B. 'Average' displays an average value of all thirteen sequences for each unit.

Underlined numbers are adjective plus noun strings with possible chunking effects.

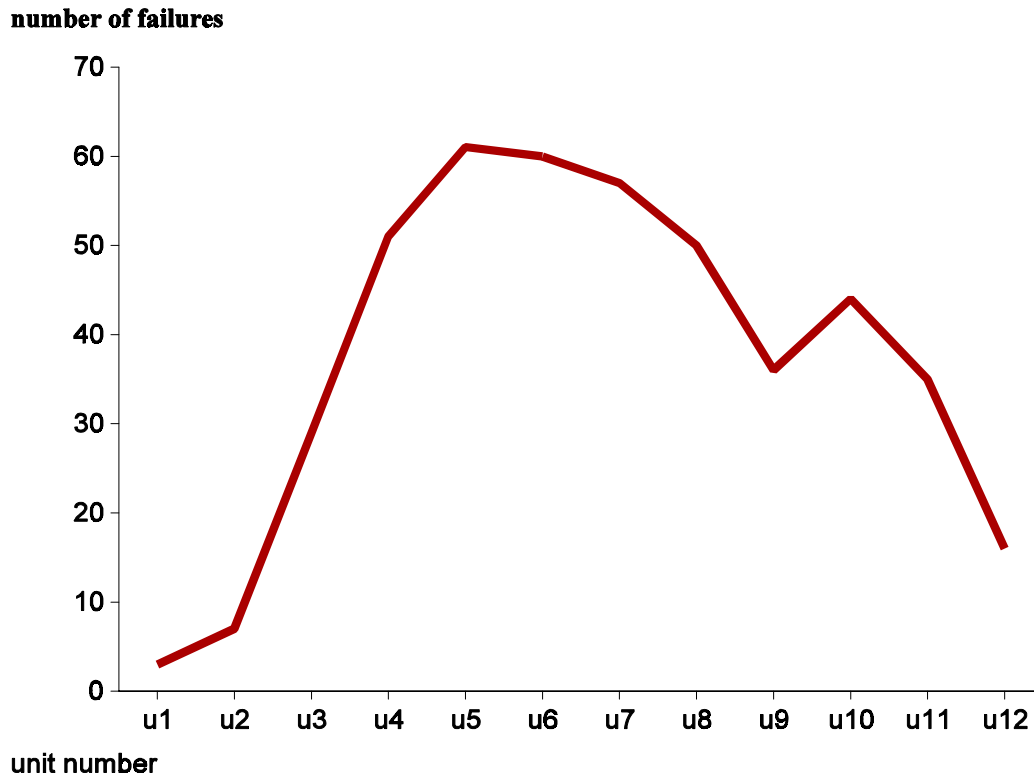


Figure 5

A possible account for the overall better performance on unit 9 is the existence of chunks by chance in this region. Wingfield and Nolan (1980) found that subjects' segmentation of ongoing speech occurred primarily at major syntactic boundaries, irrespective of the speech rate of the heard materials. Hence, if syntactic (and also semantic) chunks capable of being segmented perceptually are spotted around unit 9, they must be the cause of the relatively better result for the unit. Having checked the sequences, I have identified twelve adjective plus noun chunks, only two of which seem to be ineffective, that is, units 10 and 11 in 'Aux 1' and units 4 and 5 of 'Long w 1'. (I have excluded three natural sentences from consideration, since all items in these sentences are syntactically related to one another.) More saliently however, it has been

recognised that among the rest of ten chunks, six are clustered around the relevant region, as indicated in Table 2.

Referring to Table 2, we can observe that the number of failures drops at the responsible units for these six chunks above. Consequently, the fact that the six chunks in the region of unit 9 discovered out of ten random sequences have the comparatively better results seems to support our assumption that the drop at unit 9 of the 'Average' graph is due to chunking in this region.

In general, chunks seem to save space in short term memory, which resulted in the better performance in the experiment. Other chunks found in 'S-Natural 2' will be examined in §5.

3.2.2 Natural sentences. A chisquare test comparing three natural sentences showed, contrary to our expectations, no association among them ($p=0.000$). In fact, although I have tried every possible combination of chisquare tests on the data for these three sequences, every test confirmed this difference. This result could have been caused by ordering effects; the linear positions of the sequences may have effected subjects' performance. 'Natural 1' positioned in the first and 'Natural 3' placed in the eighth which was between the random sequences might be more unexpected and therefore harder to cope with than 'Natural 2' that immediately followed 'Natural 1'. This may explain why 'Natural 2' has the best score among the three (and also in the entire sequences).

We expect a difference between 'Natural 2' and its scrambled counterpart 'S-Natural 2', because the latter is essentially a random sequence, so its result should be worse than that of 'Natural 2'. This is borne out by a test, which showed independence of the data for the two sequences ($p=0.000$).

3.2.3 Words and particles. There were three pairs of sequences which shared the same characteristics: 'Short w 1' and 'Short w 2', 'Long w 1' and 'Long w 2', and 'Short w+p 1' and 'Short w+p 2'. On one hand, chisquare tests reinforced our anticipations for the latter two cases; correlations were significant between 'Short w+p 1' and 'Short w+p 2' and between 'Long w 1' and 'Long w 2' ($p>0.010$ in both cases). Especially, the outcome of the test for the two long content word-sequences was even more impressive ($p=0.934$). On the other hand, another test gave no evidence of a correlation between 'Short w 1' and 'Short w 2' ($p=0.000$). This unexpected result may again be explained by ordering effects. If we take into consideration that 'Short w 1' was the very first

random sequence in the list, its unexpected occurrence may therefore have induced subjects' performance to deteriorate (the successes of 'Short w 1' was rated at 51.9%, in contrast that of 'Short w 2' was 64.8%). Hence 'Short w 1' will not be used for the rest of chisquare tests due to its possible unreliability.

3.2.4 Word+particle versus word. Two tests assessed relationships among 'Short w 2', 'Short w+p 1+2' and 'Long w 1+2', the last two of which are shown as combined values for both the 'Short w+p' and both the 'Long w' strings, respectively. Our predictions are that 'Short w 2' and 'Short w+p 1+2' should be different, as should 'Short w 2' and 'Long w 1+2'. The reasons are as follows: 'Short w+p 1+2' must be syntactically (or morphologically) more difficult to process than 'Short w 2' thanks to the presence of a particle after each content word. Moreover 'Short w+p 1+2' is phonologically longer by twelve morae than 'Short w 2', which might cause difficulty of producing output simply because subjects had to write more letters down for 'Short w+p 1+2'. Recall that one Japanese alphabet sign roughly corresponds with one mora.⁵ In the case of 'Short w 2' and 'Long w 1+2', the second reason just mentioned will be applied in view of the fact that 'Long w 1+2' is twelve morae longer than 'Short w 2' with other factors shared. As to the results, the tests supported the prediction ($p=0.000$) in both cases.

The succeeding test probes into a possible association between 'Short w+p 1+2' and 'Long w 1+2'. Bearing in mind that both were identical in phonological length and contained the same number of content words, we can infer that the only difference between them is that 'Short w+p 1+2' contains a particle after each content word, but 'Long w 1+2' does not. An interesting prediction can therefore be made about this test that it would not show any correlation between the two, since 'Short w+p 1+2' should be more labourious to process than 'Long w 1+2' owing to the presence of particle. The result was as expected: the test supported the prediction ($p=0.000$). The conclusion can therefore be drawn that the processing weight of a string made of a content word plus a particle ought to be heavier than that of a single content word. In other words, particles should carry some kind of extra weight, even though this may not be as heavy as that of content words.

3.2.5 Auxiliary verbs. The following is a consideration given to the results of the two strings containing auxiliary verbs. Although 'Aux 1' and 'Aux 2' should have been

⁵Most subjects wrote sequences in *hiragana* on most occasions and in Chinese characters for some times. This may be merely because the former was faster to write than the latter.

the processing load of Japanese words

similar due to their equal conditions in terms of length and components, a test for these sequences disconfirmed our assumptions ($p=0.000$). Having examined the number of failures for each unit in the sequences in question, I have found interesting facts which may explain why subjects could recall 'Aux 1' with better accuracy than 'Aux 2'. Observe Figure 6 relevant to the discussion.

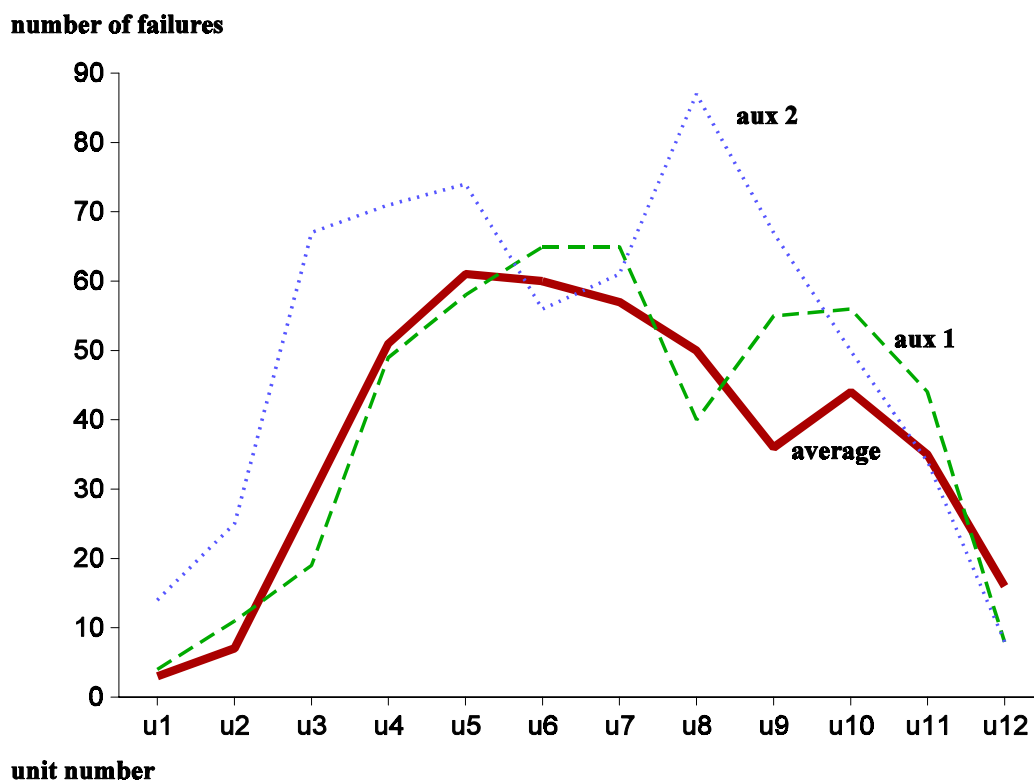


Figure 6

It can be noticed from Figure 6 that the performance in units 6 and 7 of 'Aux 2' was comparatively better. This suggests that chunking effects should have taken place in these regions. As for 'Aux 1', the figure also indicates possible chunking occurring between units 8 and 9.

What is notable is the worse performance in units 1 to 5 of 'Aux 2', compared with the corresponding parts of 'Aux 1' and 'Average', where they roughly overlap each other. We may be able to interpret this pattern as a result of possible ordering effects on 'Aux 2', considering that the sequence was positioned immediately after the syntactically structured sentence 'Natural 3'. After hearing 'Natural 3', the subjects might have expected another ordinary sentence, but what was to be processed was a random sequence. This conflict between the subjects' expectations for a natural sentence and coping with a random sequence in reality could have yielded the unpredictably poorer performance on 'Aux 2' (especially in the first five units), which consequently led to the dissimilar results between the two sequences in question. It should be remembered in relation with this that the result of the random sequence 'Short w 1' which was preceded by the natural sentence 'Natural 2' was faultier than that of 'Short w 2'. This may also explain why the results of 'Aux 1' and 'Aux 2' were different from each other.

3.2.6 Word+auxiliary verb versus word. Continuing the discussion of auxiliary verbs, another chisquare test showed a difference between 'Short w 2' and 'Aux 1+2' (the combined data of the two 'Aux' strings). Considering that both had the same number of content words in a string and the length of the sequence was equivalent, we can interpret the result as evidence that auxiliary verbs are probably separate words. This is because if the auxiliary verbs in 'Aux 1+2' had been parts of their preceding content words, there would have been the same number of words in both 'Short w 2' and 'Aux 1+2', which should have led the test to show a similarity between them. It is therefore assumed that auxiliary verbs must convey some amount of processing cost which would weigh comparatively smaller than that of content words.

3.2.7 Numeral quantifiers. Finally, we examine the data regarding the sequence consisting of twelve numeral quantifiers. It is readily noticed that subjects' performance on this string was surprisingly good with the success score of 824 out of 1200, which was ever better than that of 'Short w 2' (778). Since Arabic numerals were allowed to be used for writing the digits, the number of characters to complete 'Numeral qf' were considerably fewer compared to those of other sequences, which must have enhanced the subjects' performance of 'Numeral qf'. The figures are especially interesting because each of the twelve units to be recalled was morphologically complex, so each part of the complex (the digit and the classifier) counts for far less than a whole word in terms of processing load.

4 Weighing Japanese words

In this section, I shall attempt to weigh the processing cost of content and function words. It should be pointed out here that no estimation can be made from the experiment about relative weights for each word-class for content words. It will therefore be the case that ‘noun’, ‘adjectival noun’, ‘full verb’ and ‘adjective’ are assigned the same amount of weight.

Processing load is calculated on the basis of the number of failures. The sequences considered are as follows: ‘Long w 1+2’, ‘Short w 2’ ‘Short w+p 1+2’, and ‘Aux 1+2’. It should be noted again that ‘Short w 1’ is excluded from deliberation because of its likely unreliability due to ordering effects. We shall take the figure for ‘Long w 1+2’ as the standard against which all the other units are measured: since the failures for ‘Long w 1+2’ are 520, this number counts as a processing weight of 1 unit. The rates for the other units are given in (9) below.

(9)

	failures	processing load
Long w 1+2	520 points=(519+521)/2	1
Short w 2	422 points	0.81=422/520
Short w+p 1+2	594 points=(613+575)/2	1.14=594/520
Aux 1+2	544 points=(474+614)/2	1.04=544/520

The main difference between ‘Long w 1+2’ and ‘Short w 2’ was their lengths; each unit (i.e. content word) of ‘Long w 1+2’ consisted of five morae whereas that of ‘Short w 2’ was composed of four morae. I shall therefore assume that the processing cost for a content word comprising more than four morae is 1 unit, while a content word made up of less than five morae charges approximately the load of 0.8 units. In order to determine the weights for a particle and an auxiliary verb, we need to subtract the load of ‘Short w 2’ from ‘Short w+p 1+2’ and from ‘Aux 1+2’ respectively, bearing in mind that the lengths of the content words used in ‘Short w+p 1+2’ and ‘Aux 1+2’ were less than five morae. According to the calculations, the processing weights for a particle and an auxiliary verb will be about 0.3 and 0.2 units, respectively. In consequence, the approximate processing loads for content words, particles and auxiliary verbs will be assumed as follows:

the processing load of Japanese words

(10)

	processing load
content word $\geq 5\mu$	1.0 unit
content word $\leq 4\mu$	0.8 unit
particle	0.3 unit
auxiliary verb	0.2 unit

5 Verifying the weights

We will end this report with the verification of (10) on the observed data from the experiment. There was one random sequence of a scrambled grammatical sentence used in the experiment, which was ‘S-Natural 2’. We will first of all calculate its predicted processing load based on (10) then work out the expected numbers of failures and successes for ‘S-Natural 2’. Finally, we will compare the assumed figures and the raw data for ‘S-Natural 2’ earned from the experiment.

There were eleven content words whose lengths were less than five morae, one content word longer than five morae, ten particles and one auxiliary verb. Hence the total anticipated processing weight for ‘S-Natural 2’ will be 13.0 units. According to (9), the processing load of 1 unit corresponds to 520 points of failures, so we can derive from (9) that the expected failures for the sequence in question should be 563 ($=520 \times 13.0/12.0$) points.

A chisquare test examining an association between the actual result of ‘S-Natural 2’ and its predicted figures did not confirm a similarity between the two sequences ($p < 0.050$). In fact, we can recognise that the subjects’ performance on ‘S-Natural 2’ was surprisingly better than had been expected. One possible explanation for this unexpected difference would be that the arrangement of ‘S-Natural 2’ might have made it possible for the subjects to segment and parse the whole twelve units into fewer chunks, which would have given rise to an unpredictably better result. Potential syntactic relations among the first five units in ‘S-Natural 2’, where chunking should have been likely to happen are shown in (11).

(11)

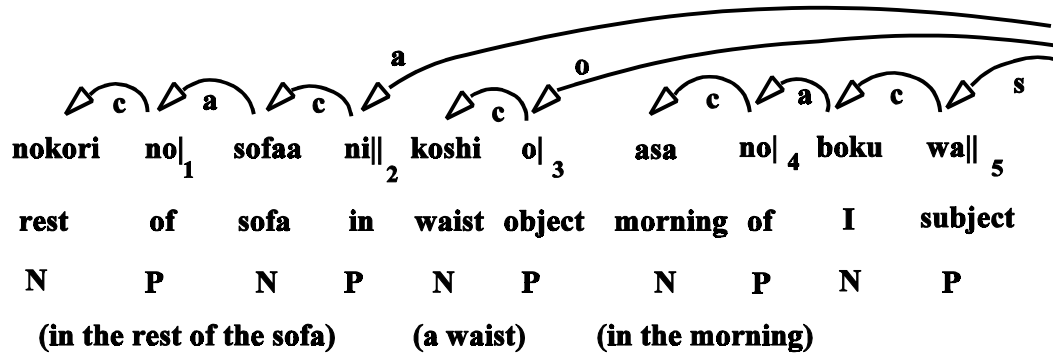


Figure 7

the processing load of Japanese words

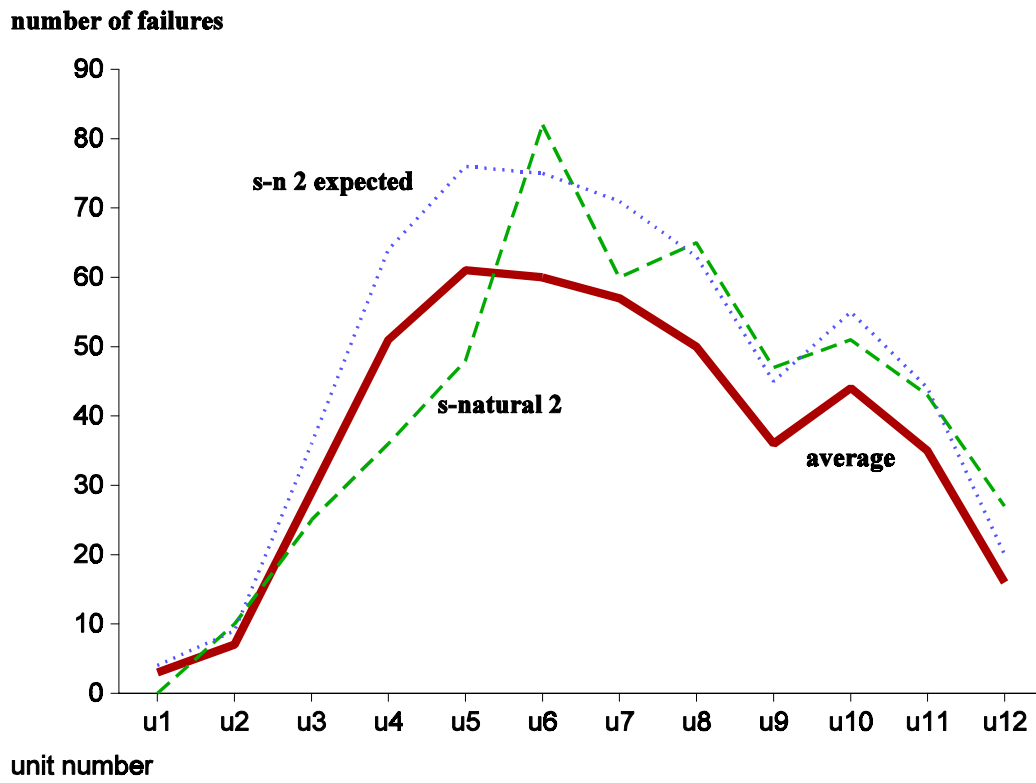


Figure 8

To reinforce the above assumption, let us monitor the number of failures for each unit of ‘S-Natural 2’ in Figure 8 above. Regarding the graph of ‘S-N 2 expected’, the number of failures expected for each unit was calculated by multiplying each figure of ‘Average’ by 1.25 on the basis that the ratio of the total failures between ‘Average’ and ‘S-N 2 expected’ is roughly 1 to 1.25 (563/450).

We can observe from Figure 8 considerable overlap between ‘S-Natural 2’ and ‘S-N 2 expected’ in units 8 through 12. There is also a similarity between the two graphs in units 1 and 2 which indicates serial position effects on these units. Furthermore, apparent differences between the two sequences in units 6 and 7 were disconfirmed by chisquare tests comparing the graphs of ‘S-Natural 2’ and ‘S-N 2 expected’ at unit 6 and unit 7 individually ($p > 0.010$ for both tests). We shall therefore focus our attention on the remaining of the graphs.

The interesting region in Figure 8 is between units 2 and 5 because on one hand, the graph for ‘S-N 2 expected’ rises abruptly at unit 2, while on the other, the one for ‘S-Natural 2’ rises in proportion to the unit number up to unit 5 and then registers acute increase. This suggests that some kind of chunking may have occurred between units 2 and 5 in ‘S-Natural 2’. Referring to (11) in which possible chunking into three parts, i.e. units 1 plus 2 and unit 3 and units 4 plus 5 is assumed, we can recognise that our assumption is justified by the implication of the graphs in Figure 8.

Moreover, the abrupt increase at unit 5 in ‘S-Natural 2’ may signify that three chunks assumed in (11), all of which were expecting a full verb as their parent, failed to be linked with the verb *nomi* (to drink) in unit 5 owing to its semantic mismatch especially with the object item *koshi o* (waist) of unit 3. It follows that the unexpectedly good performance in units 3 to 5 of ‘S-Natural 2’ may be due to chunking.

6 Conclusion

In conclusion, we have developed a method for estimating processing load for Japanese words. We have also determined the approximate processing load for content words belonging to ‘noun’ ‘verb’ or ‘adjective’ and for function words such as particles and auxiliary verbs. By applying relative processing weights to different words, we shall be able to measure the dependency distance of a Japanese sentence or phrase with more precision.

We could however, further refine the approximation in (10) by considering the memory cost for content words with fewer morae than four and the word-classes

unclassified in (10), specially ‘digit’ and ‘classifier’. This gap is because the experiment does not offer adequate data required to calculate the approximate processing load for these. Despite the sequence consisting of twelve numeral quantifiers in the experiment, the data may be unreliable since the subjects were allowed to use Arabic numbers. For the questions I have laid out here, further investigation will be required.

References

- Abeny, S. P. & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research* 20, 233-250.
- Bernstein, D. A., Clarke-Stewart, A., Roy, E. J. & Wickens, C. D. (1977). *Psychology*. (4th ed.) Boston, MA: Houghton Mifflin Company.
- Chomsky, N. & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology* (Vol. 2, pp. 269-321). New York, NY: John Wiley.
- Fraser, N. (1985). A Word Grammar parser. London.
- Fraser, N. (1989). Parsing and dependency grammar. *UCL Working Papers in Linguistics* 1.
- Fraser, N. (1985). Dependency parsing. London.
- Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. Cambridge, UK: Cambridge University Press.
- Frazier, L. & Rayner, K. (1988). Parameterizing the language processing system: Left- vs. right-branching within and across languages. In J. Hawkins (ed.), *Explaining Language Universals* (pp. 247-279). Oxford: Blackwell.
- Gibson, E. (1991). *A Computational Theory of Human Linguistics Processing: Memory Limitations and Processing Breakdown*. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Gibson, E., Thomas, J. & Babyonyshev, M. (1995). Processing center-embedded and self-embedded structures in English and Japanese. Paper presented at the NELS.
- Gibson, E. (1997). Linguistic complexity: Locality of syntactic dependencies.
- Hudson, R. A. (1984). *Word Grammar*. Oxford: Blackwell.
- Hudson, R. A. (1989). Towards a computer-testable Word Grammar of English. *UCL Working Papers in Linguistics* 1, 321-339.
- Hudson, R. A. (1990). *English Word Grammar*. Oxford: Blackwell.
- Hudson, R. A. (1995). Measuring Syntactic Difficulty. London.
- Hudson, R. A. (1996a). Discontinuity. London.
- Hudson, R. A. (1996b). The difficulty of (so-called) self-embedded structures. *UCL Working Papers in Linguistics* 8, 283-314.
- Hudson, R. A. (1997). Weighing other people's words: Syntax and readability. London.

- Lewis, R. (1993). *An architecturally-based Theory of Human Sentence Comprehension*. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Marcus, M. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.
- Mazuka, R. & Nagai, N. (1995). Japanese sentence processing: An interdisciplinary approach. In R. Mazuka & N. Nagai (eds.) *Japanese Sentence Processing* (pp. 1-7). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Miller, G. A. & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology* (Vol. 2, pp. 419-491). New York, NY: John Wiley.
- Miller, G. & Isard, S. (1964). Free recall of self-embedded English sentences. *Information and Control* 7, 292-303.
- Pullman, G. & Zwicky, A. (1988). The syntax-phonology interface. In F. Newmeyer (ed.), *Linguistics: The Cambridge Survey I: Linguistic Theory Foundations*. Cambridge: Cambridge University Press.
- Syaumyan, O. (1995). *Parsing English with Word Grammar*. London.
- Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton Jr., L. Frazier & K. Rayner (eds.), *Perspectives on Sentence Processing* (pp. 303-336). Hillsdale, NJ: Lawrence Erlbaum.
- Sturt, P. (1998). *Monotonic parsing and reanalysis*. Paper presented at UCL Linguistics seminar.
- Tesnière, L. (1959). *Eléments de Syntaxe Structurale*. Paris.
- Wingfield, A. & Nolan, K. A. (1980). Spontaneous segmentation in normal and in time-compressed speech. *Perception and psychophysics* 28, 97-102.
- Wingfield, A. & Butterworth, B. (1984). Running memory for sentences and parts of sentences: Syntactic parsing as a control function in working memory. In H. Bouma & D. Bouwhuis (eds.) *Attention and Performance X* (pp. 351-363). Hillsdale, NJ: Erlbaum.
- Yngve, V. H. (1960). *A Model and an Hypothesis for Language Structure*. Paper presented at the Proceedings of the American Philosophical Society.
- Zwicky, A. (1977). *On Clitics*: Bloomington.

Appendix: sequences used for experiment

N.B. | indicates a boundary between meaningful units in a sequence.

KEY: N=noun; jN=adjectival noun; J=adjective; V=full verb; A=adverb; v=auxiliary verb; p=particle; dN=digit; qN=classifier

1. Natural 1

Akiya	de	wa ₁	at	te	mo ₂	tanin	no ₃	ie	no ₄
unoccupied house		topic	to be		although	stranger	of	house	of
N	P	P	V	P	P	N	P	N	P
shikichi	ni ₅	katte	ni ₆	ashi	o ₇	humiireru ₈	koto	wa ₉	houritsu
premises	on	own way		foot	object	to step into	thing	subject	law
N	P	jN	P	N	P	V	N	P	N
ni ₁₀	hanshi	ta ₁₁	koui	da. ₁₂					
to	to be against	past	act	copula					
P	V	v	N	v					

(Entering a stranger's house without permission, even if it is unoccupied, is an illegal act.)

2. Natural 2

Boku	wa ₁	sofaa	ni ₂	koshi	o ₃	oroshi	te ₄	asa	no ₅	nokori	no ₆
I	subject	sofa	in	waist	object	to lower		morning	of	rest	of
N	P	N	P	N	P	V	P	N	P	N	P
koohii	o ₇	nomi	nagara ₈	atama	no ₉	naka	o ₁₀	sukoshi ₁₁			
coffee	object	to drink	while	head	no	inside	object	a little			
N	P	V	P	N	P	N	P	A			
seirishi-te-mi	ta. ₁₂										
to try clearing	past										
V	v										

(Sitting on the sofa and drinking the rest of the morning coffee, I tried clearing my head a little.)

3. Short d 1

huusen ₁	hurui ₂	muzukashii ₃	taberu ₄	akarui ₅	depaato ₆	tazuneru ₇
balloon	old	difficult	to eat	bright	department store	to ask
N	J	J	V	J	N	V
oishii ₈	nezumi ₉	totonoe ₁₀	okurimono ₁₁	umareru ₁₂		
tasty	mouse	to arrange	gift	to be born		
J	N	V	N	V		

4. Short w+p 2

atama	ga ₁	komakaku	te ₂	kurai	to ₃	kangae	te ₄	hatarake	ba ₅
head	subject	fine		dark		to think		to work	if
N	P	J	P	J	P	V	P	V	P
nuigurumi	o ₆	hohoemu	to ₇	tsutsumashiku	te ₈	itaria	no ₉		
stuffed toy	object	to smile		modest		Italy	of		
N	P	V	P	J		P	N	P	
gyuniku	wa ₁₀	ureshikere	ba ₁₁	mayot	te ₁₂				
meat	topic	glad	if	to wonder					
N	P	J	P	V	P				

5. Numeral qf

jup	pon ₁	hachi	mai ₂		juu-ichi	ji ₃	ni-jut	tou ₄	
10	bottles	8	thin flat objects		11	o'clock	20	animals	
dN	qN	dN	qN		dN dN	qN	dN dN	qN	
ichi	i ₅	ni	wa ₆	san	ko ₇	go-juu-san	pun ₈	kyuu	dai ₉
1	place	2	birds	3	pieces	53	minutes	9	machines
dN	qN	dN	qN	dN	qN	dN dN dN	qN	dN	qN
roku	satsu ₁₀	yon	kai ₁₁	shi	gatsu ₁₂				
6	books	4	times	4	month (April)				
dN	qN	dN	qN	dN	qN				

6. Aux 1

kirei	da ₁	hon	desu ₂	mi	mashi	ta ₃	uso	da ₄	shizuka
beautiful	copula	book	copula	to see	polite	past	lie	copula	quiet
jN	v	N	v	V	v	v	N	v	jN
da ₅	kaeri	masu ₆	ason	da ₇	yukai	na ₈	namida	da ₉	
copula	to return	polite	to play	past	pleasant	copula	tear	copula	
v	V	v	V	v	jN	v	N	v	
genki	na ₁₀	hana	desu ₁₁	kikoe	ta ₁₂				
healthy	copula	flower	copula	to sound	past				
jN	v	N	v	V	v				

7. Long d 1

reizouko ₁	konpyuutaa ₂	chirakasu ₃	kagayakashii ₄	tamago ₅	itametsukeru ₆
refrigerator	computer	to scatter	glorious	egg	to torment
N	N	V	J	N	V
atsukamashii ₇	moushikomu ₈	suzushii ₉	yakusoku ₁₀	kashikoi ₁₁	wakimaeru ₁₂
presumptuous	to apply	cool	promise	wise	to distinguish
J	V	J	N	J	V

the processing load of Japanese words

8. Natural 3

Tonari	no 1	musume	wa 2	boku	no 3	tekubi	ni 4	kusuriyubi	o 5	oi
next	of	girl	subject	I	of	wrist	on	little finger	object	to put
N	P	N	P	N	P	N	P	N	P	V
te 6	katachi	no 7	sadamara	nai 8	kimyou	na 9	zukei	o 10		
	shape	of	to be fixed	negative	strange	copula	figure	object		
P	N	P	V	v	jN	v	N	P		
soko	ni 11	kai	ta. 12							
there	on	to draw	past							
n	P	V	v							

(The girl next to me put her little finger on my wrist and drew a strange and shapeless figure there.)

9. Aux 2

nigiyaka	da 1	hune	desu 2	yomi	mashi	ta 3	kome	da 4	kai
bustling	copula	ship	copula	to read	polite	past	rice	copula	to buy
jN	v	N	v	V	v	v	N	v	V
masu 5	hima	na 6	usagi	desu 7	benri	na 8	ton	da 9	
polite	leisure	copula	rabbit	copula	convenient	copula	to fly	past	
v	jN	v	N	v	jN	v	V	v	
tomat	ta 10	kirai	da 11	anata	da 12				
to stop	past	to dislike	copula	you	copula				
V	v	jN	v	N	v				

10. Long d 2

nagusameru 1	mayoneezu 2	isogashii 3	dairiseki 4	urayamashii 5	otoroeru 6
to comfort	mayonnaise	busy	marble	envious	to decline
V	N	J	N	J	V
shiraberu 7	kanashii 8	keisatsukan 9	omoidasu 10	niwatori 11	misuborashii 12
to examine	sad	police officer	to remember	fowl	shabby
V	J	N	V	N	J

11. S-Natural 2

nokori	no 1	sofaa	ni 2	koshi	o 3	asa	no 4	Boku	wa 5	nomi
rest	of	sofa	in	waist	object	morning	of	I	subject	to drink
N	P	N	P	N	P	N	P	N	P	V
nagara 6	seirishi-te-mi	ta 7	sukoshi 8	koohii	o 9	atama	no 10			
while	to try clearing	past	a little	coffee	object	head	of			
P	V	P	A	N	P	N	P			
oroshi	te 11	naka	o 12							
to lower		inside	object							
V	P	N	P							

12. Short d 2

hataraku ₁	kurai ₂	atama ₃	hohoemu ₄	tsutsumashii ₅	kangaeru ₆	itaria ₇
to work	dark	head	to smile	modest	to think	Italy
V	J	N	V	J	V	N
ureshii ₈	gyuuniku ₉	komakai ₁₀	mayou ₁₁	nuigurumi ₁₂		
glad	meat	fine	to wonder	stuffed toy		
J	N	J	V	N		

13. Short d+p 1

hurukere	ba ₁	nezumi	ga ₂	totonoe	te ₃	akaruku	te ₄	huusen	wa ₅
old	if	mouse	subject	to arrange	bright			balloon	topic
J	P	N	P	V	P	J	P	N	P
tabere	ba ₆	umare	te ₇	depaato	ni ₈	muzukashiku	te ₉		
to eat	if	to be born		department store	to	difficult			
V	P	V	P	N	P	J			P
okurimono	o ₁₀	tazuneru	to ₁₁	oishikere	ba ₁₂				
gift	object	to ask		tasty	if				
N	P	V	P	J	P				