Sentence prosody and syntax in speech perception

David House

ABSTRACT

In order to test the use of sentence prosody related to syntax in sentence processing, a listening test was constructed consisting of marked and unmarked prosodic versions of sentences with various syntactic structures. Results of a pilot test showed considerable listener differences in sentence interpretations. While listeners were able to use prosody as an aid in sentence processing, they also relied heavily upon other cues such as simplicity strategies and morphological restrictions supplied by the lexicon. When the cues provided conflicting information, listeners, on an individual basis, tended to select the most salient cue to determine syntax.

1. INTRODUCTION

One of the functions of sentence prosody is to provide the listener with acoustic cues which represent aspects of syntactic structure. This is accomplished primarily by means of fundamental frequency excursions (intonation) and manipulation of the time dimension such as final lengthening and pauses. These aspects of the speech signal can be used by the speaker to emphasize certain structural relationships within a sentence and by the listener to confirm underlying and surface structures or to resolve structural ambiguities.

In a study of the production and perception of ambiguous sentences, Lehiste (1973) found a correlation between surface constituent structure and success in disambiguation. Speakers were able to convey structural differences and listeners were able to perceive them. Ambiguous sentences in which a difference in meaning was not correlated with a difference in surface constituent structure were generally not disambiguated. There were, however, considerable differences between speakers and between listeners concerning success in disambiguation. Nonetheless, it seems clear that when available, intonation, final lengthening and pauses in the acoustic signal can impinge on the syntactic structure of the sentence.

One question in this respect is to what extent listeners can make use of the acoustic signal to aid in the processing of syntax. The goal of this paper is to investigate relationships between the use of prosody in sentence perception and the nature of the sentence structure in terms of complexity and frequency of occurrence.

A reasonable hypothesis would be that generally the use of prosodic cues in perception increases as a function of syntactic complexity with certain reservations for frequency differences between structures of varying complexity. A correlate to this hypothesis would then be that the use of such cues in perception decreases as a function of syntactic cues available through lexical subcategorization frames (Ford, Bresnan, & Kaplan, 1982).

To test the hypothesis, listeners could be presented with sentences having a reduced relative structure where the first verb is locally ambiguous between the simple sentence structure and the more complex reduced relative structure. An example of this type of sentence in English is the classic "garden path" sentence:

1.) The horse raced past the barn fell.

where the listener is led down "the garden path" until the

final verb is understood at which point the sentence must be reparsed.

Clark & Clark (1977) propose a complexity strategy by which a listener will assume the first clause to be a main clause provided it is not marked at or prior to the main verb as something other than a main clause. In accordance with this strategy, if sentence prosody is also ambiguous, listeners would be expected to begin processing by assuming the first clause to be a main clause thus incorrectly resolving the ambiguity of the verb by interpreting it as active. The listener would therefore end up with a structure which does not match the input upon input of the final verb. The sentence would then have to be reparsed lengthening processing time. If, however, markedly appropriate prosodic cues are supplied in the acoustic input, e.g. a final fall-rise in intonation at the end of the subject combined with a pause following the subject marking the second clause as a relative clause prior to the verb, the correct structure should be projected earlier. This would in turn lead to a shorter processing time. Reaction times to the point at which the sentence is understood could be used as a rough gauge of processing time.

To test the correlate hypothesis, i.e. that the use of prosodic information decreases as more syntactic information is available from lexical redundancy rules, "subcategorization features" or morphological restrictions, a second sentence type such as:

2.) The horse driven past the barn fell.

could be used. Here, the syntactic structure can be determined by lexical information and the listener would be able to tell upon processing the verb that it belongs to a reduced relative clause. Thus the listener should be able to parse the sentence correctly considerably earlier solely on the basis of segmental cues and linguistic competence. In this case, marked prosodic information should not contribute to the speeding-up of processing time as much as when the verb is ambiguous if at

all.

The differences in processing time between the two sentence types could be manifested experimentally by differences in reaction times. Reaction times to sentence type 1. with marked prosodic cues should be considerably shorter than without marked prosody. Reaction times to sentence type 2. with marked prosodic cues should only be slighty shorter than without marked prosody.

There is then a proposed interplay between sentence complexity, prosodic features, and lexical-morphological syntactic information. This interplay should be manifested by processing time for sentence parsing during speech perception.

2. SPEECH MATERIAL AND METHOD

Swedish sentences were chosen for the test such that plural subject-verb agreement corresponded to sentence type 1. (ambiguous verb), and singular subject-verb agreement corresponded to sentence type 2. (unambiguous verb) by virtue of the fact that the plural morpheme for passive in weak verbs is identical to the imperfect, while for singular they are differentiated. Two different sentence structures were chosen. The first structure contained a single passive verb, the preposition and the object followed by the active verb:

3a) Männen, ritade av flickan, rökte.
(The men, drawn by the girl, smoked.)

3b) Mannen, ritad av flickan, rökte.
(The man, drawn by the girl, smoked.)

In this structure, the preposition "av" can be misparsed as a particle which follows an active verb. This would alter the meaning of sentence 3a to "The men drew the girl, smoked". Sentence 3b should not be misparsed because "ritad" can only be passive. When "av" is used as a particle, however, it takes primary stress manifested prosodically by rising intonation,

lengthening and increased intensity. It was feared that without these prosodic cues, "av" would be immediately parsed as a preposition leading to a correct reparsing earlier than the final verb.

As a control of the possible effects of this stressed-unstressed distinction in the first structure, a second structure was also included in the test. This second structure contained a compound passive verb with the subject implied followed by the active verb:

4a) Killarna, rakade och tvättade, sjöng.
(The boys, shaved and washed, sang.)

4b) Killen, rakad och tvättad, sjöng.
(The boy, shaved and washed, sang.)

In this structure, correct reparsing would not be expected until the final verb is perceived.

A listening test was constructed comprising three pairs of each sentence type plus four pairs of unambiguous filler sentences of the following type:

5a) Killarna sa, att bilen skulle tvättas.
(The boys said, that the car should be washed.)

5b) Killen sa, att bilen skulle tvättas.
(The boy said, that the car should be washed.)

Each sentence was recorded by a native female speaker of Southern Swedish in two versions, one version with marked intonation and pauses and one version without pauses and with unmarked intonation (Fig. 1 and 2). The forty test sentences were then ordered systematically so that different versions of different sentences appeared before either version or sentence type was repeated. (See appendix for test sentences and sentence order.) This type of test construction would enable different test versions to be presented to different listeners with the results being compiled across listeners to eliminate learning effects. Four filler sentences were presented as a buffer before the test began.

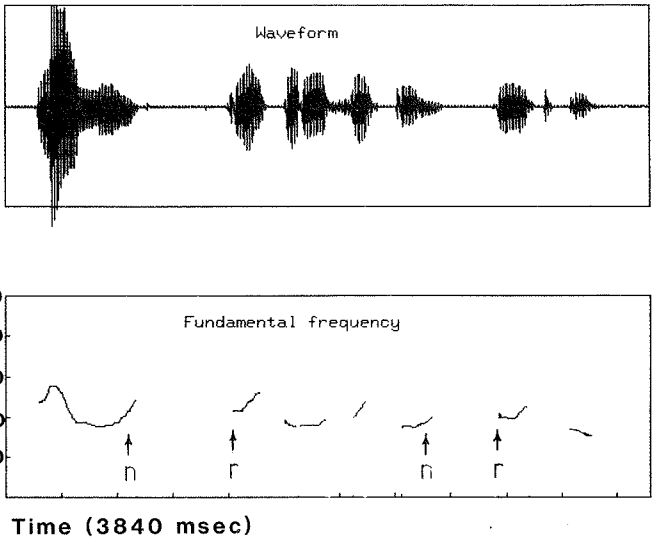

Figure 1. Waveform and fundamental frequency contour of test sentence 1A (marked intonation and pauses).

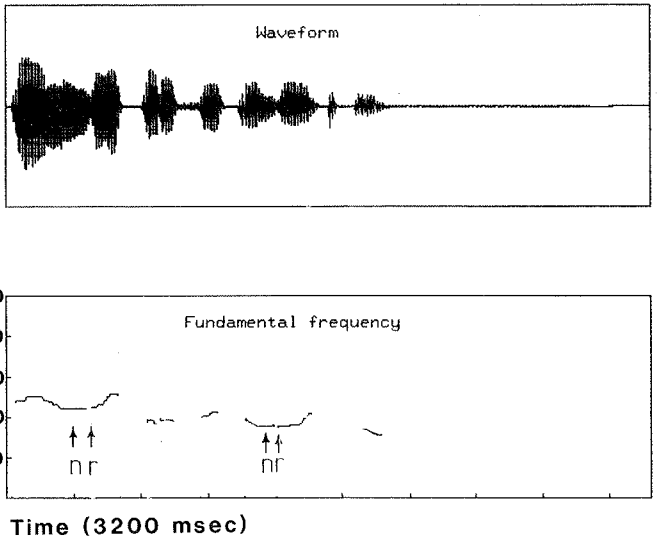

Figure 2. Waveform and fundamental frequency contour of test sentence 1C (unmarked intonation, no pauses).

Two listeners trained in linguistics and two naive listeners participated in a pilot version of the test. The sentences were presented over a loudspeaker in a sound-treated room where the listeners were tested individually. They were instructed to rest their index finger on a button and to press it as soon as they had understood the sentence. They were requested to paraphrase each sentence after pressing the button.

Reaction times were measured by a micro-computer triggered at the beginning of each sentence by an envelope detector. The paraphrased versions of each sentence were recorded by the experimenter.

3. PILOT TEST RESULTS

A. Reaction time measurements

The expected reaction time results, i.e. that sentence versions A and B would elicit faster reaction times than sentence versions C and D and that the difference between reaction times for A and C would be greater than for B and D were not obtained. Instead, reaction times varied from around 500 msec. preceding the end of the sentence to around 500 msec. following the end of the sentence. The variation for all four listeners appeared to be completely random. This is most probably due to the rather coarse-grained nature of the task. To press a button when "you have understood the sentence" does not appear to be a very precise measure of reaction time, and probably does not correlate well at all with actual processing time.

B. Sentence interpretations

Results obtained from the listeners' paraphrases were much more illuminating. The two listeners with training in linguistics (listeners 1 and 2) were immediately aware of the structures involved and paraphrased each sentence correctly. The naive listeners (listeners 3 and 4), on the other hand, differed in their interpretations of the sentences. Both listeners,

however, correctly parsed the sentences in more instances in the marked prosody versions (42%) than in the unmarked versions (25%) (Fig. 3).

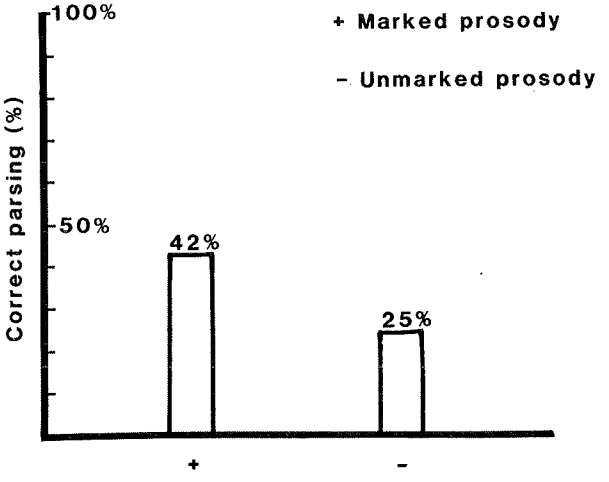

Figure 3. Correct parsing in percent for listeners 3 and 4 (combined results).

Misparsings of the sentences by both naive listeners conformed to the complexity strategies outlined above. Sentences having the structure as in 3a and 3b (sentences 1 to 3 in the test, see appendix) were most often parsed as two NP-VP's joined by "och" (and). Thus test sentence 1A "Männen, ritade av flickan, rökte," (The men, drawn by the girl, smoked) was interpreted as "Männen ritade och flickan rökte," (The men drew and the girl smoked). Listener 4, however, parsed several of these sentences as "Männen ritade av flickan och hon rökte," (The men drew the girl and she smoked). In these instances, listener 4 disregarded the unstressed prosodic features of the preposition

"av". Even sentences where the syntax was specified by lexical subcategorization frames were similarly misparsed.

Sentences having the structure of 4a and 4b (test sentences 4 to 6) were misparsed by both naive listeners as NP-compound VP by adding an additional "and" between the final two verbs. It is interesting to note, however, that both sentences containing the phonetically more salient allomorph /-at/ (test sentences 5B and 5D) were correctly parsed by both listeners, and that these two sentences were the only two with the exception of 6B that were correctly interpreted by listener 4 (Fig. 4).

Correct parsings, listener 3

A	B	C	D
1A			
	2B		
3A	3B		
4A	4B	4C	4D
	5B	5C	5D
6A		6C	

Correct parsings, listener 4

A	B	C	D
	5B		5D
	6B		

Figure 4. Sentences in which correct parsing was performed (naive listeners, 3 and 4)

4. DISCUSSION

A. Reaction times

The difficulties involved in obtaining reaction times which could be correlated to actual sentence processing times rendered impossible the testing of those aspects of the hypothesis concerning the relationships between speed-of-processing, preliminary misparsing and reparsing. The fact that many of the sentences were not reparsed upon comprehension of the final verb, but rather the misparsings were maintained to conform to a simple structure, seems to indicate that in those sentences processing time might have been faster than in the sentences which received correct parsing.

Factors other than syntax and prosody might also have influenced reaction times. As previously mentioned, measuring reaction times where comprehension of an entire sentence is concerned, at least judging from this pilot test, is a very rough tool for investigating the complex process of sentence comprehension. The results indicate that perhaps more success can be attained by constructing material which can be misinterpreted and then using listener interpretations to investigate the processing of syntactic structures.

B. Syntactic complexity strategies and listener differences

The awareness of structure on the part of listeners 1 and 2 (linguistically trained) reflect the necessity of using naive listeners if misinterpretation results are to be used to investigate syntax and prosody. The difference in interpretations between listeners 3 and 4 (naive) reflect the individual nature of syntactic competence when processing complex structures.

Basically, however, these results support the complexity strategy described above. The naive listeners tended to interpret the first verb as the logical verb of the initial NP

even in cases where this interpretation conflicted with phonetic form. In other words, the processing mechanism can override phonetic form to make it fit a simple sentence syntax. There seem to be, however, certain limitations depending on the phonetic saliency of the morphology (see D. below). There was not a strict priority order for choosing simple syntax over phonetic form or linguistic competence, but rather if the phonetic form of the sentence and linguistic competence would allow the misinterpretation, the simple syntactic structure was adhered to.

C. Effects of prosody on syntax

The preliminary results of the pilot test indicate that marked prosody can be used in speech perception to parse a sentence in accordance with a complex syntactic structure. Prosody then fits into the complexity strategy as it supplies information concerning sentence structure (a fall-rise in Fo and a pause) prior to the main verb.

The results also indicate, however, that such prosodic cues are not always needed for correct parsing depending on the linguistic competence of the listener and the saliency of the segmental phonetic aspects of the sentence. In a like manner, prosody is not always used leading to misparsings of sentences.

D. Morphological saliency and phonetic form

There seems to be a rather complex interplay between phonetic form (both segmental and prosodic), morphology as manifested by mapping between lexical items, and syntax. As evidenced by the domination of the allomorph /-at/ over complexity strategies and unmarked prosody, the use of information specified by the lexicon as discussed by Bresnan (1978) and Wasow (1977) can depend greatly on the phonetic saliency of the surface form. The suffix /-at/ with its clearly perceived high-frequency burst is immediately distinguished from the suffix /-ade/ whereas /-ad/+V is much more easily interpreted as /-ade/+V.

The results of this experiment can be seen as preliminary evidence for a model of syntactic processing which operates by simultaneously integrating aspects of the acoustic signal, syntactic complexity strategies involving linguistic competence, and morphological restrictions in the lexicon. This model of syntax perception would be characterized by a mechanism which latches on to the most prominent aspect or cue in the sentence which can determine syntax. The important point here is that cue dominance is not determined by the linear order in which the cues appear in the sentence, but rather dominance is affected by a number of different factors such as phonetic saliency, linguistic competence, and relative frequency of structure. It could also be conjectured that cue dominance leads to a certain kind of perceptual masking of other cues in an intricate interplay of processing.

E. Further experiments with synthetic stimuli

A more satisfactory test of the hypothesis and model proposed in this paper would entail the synthetic manipulation of the prosodic parameters of intonation and pauses. A much larger number of listeners would of course also be needed to support the trends indicated by the pilot study.

An informal experiment was carried out in which the intonation contour of test sentence 1A "Männen, ritade av flickan, rökte" was replaced by the intonation contour of 1C "Männen ritade av flickan rökte", i.e. unmarked intonation replaced marked intonation and vice versa (Fig. 5 and 6). Simply switching intonation, however, without changing the timing aspects produced extremely artificial sounding sentences. A systematic alteration of both intonation and pauses would need to be undertaken to more successfully test the degree of prosodic representation necessary to serve as the dominating cue in the perception of sentence syntax.

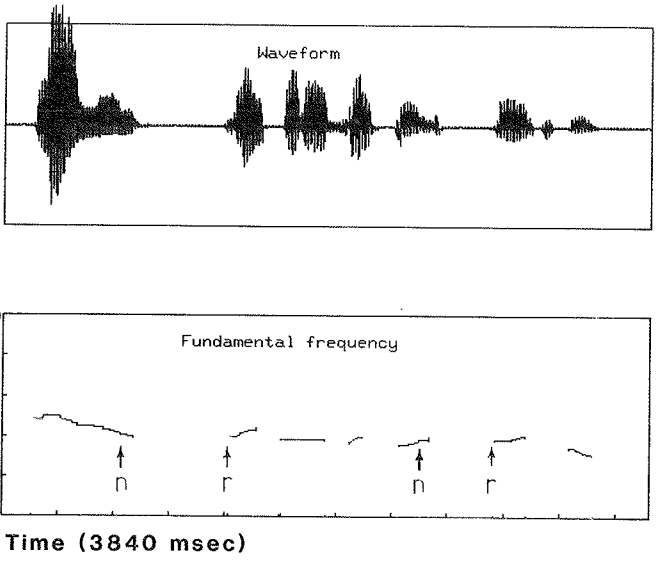

Figure 5. Waveform and fundamental frequency contour of test sentence 1A with pitch coefficients edited to conform to the contour of test sentence 1C (unmarked intonation).

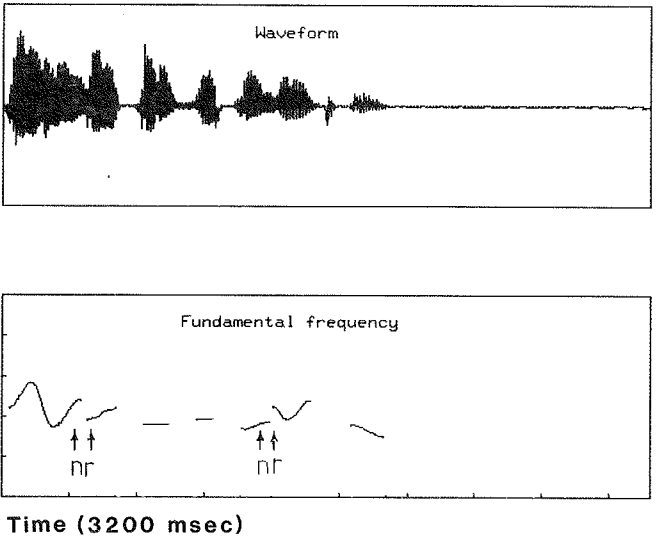

Figure 6. Waveform and fundamental frequency contour of test sentence 1C with pitch coefficients edited to conform to the contour of test sentence 1A (marked intonation).

ACKNOWLEDGEMENTS

I would like to thank Helen Goodluck for much help, discussion and comments. Thanks also to Gösta Bruce and Eva Gårding for discussion and comments and to Anders Löfqvist for help with the computer program.

REFERENCES

Bresnan,J. 1978. A Realistic Transformational Grammar, in Halle, Bresnan, Miller (Eds.) Linguistic Theory and Psychological Reality. The MIT Press. Cambridge, MA.

Clark H. & Clark E. 1977. Psychology and Language, Harcourt Brace Jovanovich, New York.

Ford, M., Bresnan, J., & Kaplan, R. 1982. A Competence- Based Theory of Syntactic Closure. In Bresnan, J. (Ed.) The Mental Representation of Grammatical Relations. The MIT Press, Cambridge, MA.

Lehiste, I. 1973. Phonetic disambiguation of syntactic ambiguity. Glossa, 7, 107-122.

Wasow, T. 1977. Transformations and the Lexicon. In Culicover, Wasow, Akmajian (Eds.) Formal Syntax. Academic Press, New York.

APPENDIX

Test sentences. Commas indicate marked prosody.

1A. Männen, ritade av flickan, rökte.
1B. Mannen, ritad av flickan, rökte.
1C. Männen ritade av flickan rökte.
1D. Mannen ritad av flickan rökte.

2A. Kvinnorna, målade av mannen, log.
2B. Kvinnan, målad av mannen, log.
2C. Kvinnorna målade av mannen log.
2D. Kvinnan målad av mannen log.

3A. Flickorna, tecknade av pojken, skrattade.
3B. Flickan, tecknad av pojken, skrattade.
3C. Flickorna tecknade av pojken skrattade.
3D. Flickan tecknad av pojken skrattade.

4A. Killarna, rakade och tvättade, sjöng.
4B. Killen, rakad och tvättad, sjöng.
4C. Killarna rakade och tvättade sjöng.
4D. Killen rakad och tvättad sjöng.

5A. Barnen, tvättade och torkade, lekte.
5B. Barnet, tvättat och torkat, lekte.
5C. Barnen tvättade och torkade lekte.
5D. Barnet tvättat och torkat lekte.

6A. Kvinnorna, tecknade och målade, log.
6B. Kvinnan, tecknad och målad, log.
6C. Kvinnorna tecknade och målade log.
6D. Kvinnan tecknad och målad log.

7A. Killarna sa, att bilen skulle tvättas.
7B. Killen sa, att bilen skulle tvättas.
7C. Killarna sa att bilen skulle tvättas.
7D. Killen sa att bilen skulle tvättas.

8A. Pojkarna sa, att flickorna skulle spela med.
8B. Pojken sa, att flickorna skulle spela med.
8C. Pojkarna sa att flickorna skulle spela med.
8D. Pojken sa att flickorna skulle spela med.

9A. Flickorna sa, att pojkarna skulle måla huset.
9B. Flickan sa, att pojkarna skulle måla huset.
9C. Flickorna sa att pojkarna skulle måla huset.
9D. Flickan sa att pojkarna skulle måla huset.

10A. Killarna sa, att huset skulle målas.
10B. Killen sa, att huset skulle målas.
10C. Killarna sa att huset skulle målas.
10D. Killen sa att huset skulle målas.

Test Sentences (Translation). Commas indicate marked prosody.

1A. The men, drawn by the girl, smoked.
1B. The man, drawn by the girl, smoked.
1C. The men drawn by the girl smoked.
1D. The man drawn by the girl smoked.

2A. The women, painted by the man, smiled.
2B. The woman, painted by the man, smiled.
2C. The women painted by the man smiled.
2D. The woman painted by the man smiled.

3A. The girls, sketched by the boy, laughed.
3B. The girl, sketched by the boy, laughed.
3C. The girls sketched by the boy laughed.
3D. The girl sketched by the boy laughed.

4A. The boys, shaved and washed, laughed.
4B. The boy, shaved and washed, laughed.
4C. The boys shaved and washed laughed.
4D. The boy shaved and washed laughed.

5A. The children, washed and dried, played.
5B. The child, washed and dried, played.
5C. The children washed and dried played.
5D. The child washed and dried played.

6A. The women, sketched and painted, smiled.
6B. The woman, sketched and painted, smiled.
6C. The women sketched and painted smiled.
6D. The woman sketched and painted smiled.

7A. The boys said, that the car should be washed.
7B. The boy said, that the car should be washed.
7C. The boys said that the car should be washed.
7D. The boy said that the car should be washed.

8A. The boys said, that the girls should play along.
8B. The boy said, that the girls should play along.
8C. The boys said that the girls should play along.
8D. The boy said that the girls should play along.

9A. The girls said, that the boys should paint the house.
9B. The girl said, that the boys should paint the house.
9C. The girls said that the boys should paint the house.
9D. The girl said that the boys should paint the house.

10A. The boys said, that the house should be painted.
10B. The boy said, that the house should be painted.
10C. The boys said that the house should be painted.
10D. The boy said that the house should be painted.

Sentence order in the listening test

1. (10A)	11. (3C)	21. (6A)	31. (9C)
2. (10B)	12. (6D)	22. (9B)	32. (1D)
3. (10C)	13. (9A)	23. (1C)	33. (4A)
4. (10D)	14. (1B)	24. (4D)	34. (7B)
5. (1A)	15. (4C)	25. (7A)	35. (2C)
6. (4B)	16. (7D)	26. (2B)	36. (5D)
7. (7C)	17. (2A)	27. (5C)	37. (8A)
8. (2D)	18. (5B)	28. (8D)	38. (3B)
9. (5A)	19. (8C)	29. (3A)	39. (6C)
10.(8B)	20. (3D)	30. (6B)	40. (9D)