



The linguistic role of hesitation disfluencies: evidence from Hebrew and Japanese

Vered Silber-Varod¹ & Takehiko Maruyama²

¹The Research Center for Innovation in Learning Technologies, The Open University, Israel

²National Institute for Japanese Language and Linguistics, Japan

Abstract

In this paper we examine a certain aspect of prosody-syntax interface, that of hesitation disfluencies (HD) that occur intra-phrases or intra-morphemes. Such cases were found in two spontaneous corpora of two syntactically distinct languages – Israeli Hebrew (IH) and Japanese. It was found that intra-phrasal hesitations in the two languages calls for different explanations, since in Japanese the noun (e.g., in NP) precedes the case marking particle while in IH the preposition (e.g., in PP) precedes the noun. In this paper we will present qualitative findings and suggest a unified view of the phenomenon of intra-phrasal HDs.

Index Terms: Hesitation disfluency, prosody-syntax interface, Israeli Hebrew, Japanese.

1. Introduction

In the present research we investigate the phenomenon of hesitation disfluencies (HD) in spontaneous Israeli Hebrew (IH) and Japanese. HDs are defined as prosodic manipulation of the speaker, produced by excessive elongation of a word final syllable (yet, see [1:59–70] for a detailed definition of the phonological realization of HDs in IH). For example:

Hesitation Disfluencies in spontaneous speech (1)

a. IH

[ani ani yexola lehavin et **ha:** et **ha:** tiskul]

I I can to.understand Acc. **the:** Acc. **the:** frustration.

‘I can understand the frustration’

b. Japanese

[**juutigatu:** no atama goro da to omou]

November: of beginning about COP QUOT think

‘I think it was about the beginning of November.’

The cognitive function of HD has been dealt within several theories, with respect to the part that the mental lexicon plays in the speech process. In the study of disfluencies such as ‘filled pauses’ (FPs), a major approach views them as indicators of increased cognitive processing. Shriberg [2], for example, claims that disfluency rates depend on the length and complexity of the sentence [2:157]; Clark and Wasow [3] showed that in American English, complex syntactic structures predict repetitions of function words. Their findings led them to formulate the commit-and-restore model of repeated words [3:203] and to propose the Complexity Hypothesis. For example, they showed that speakers repeat the definite article *the* when it precedes a complex noun phrase; Roll et al. [4] concluded that, in Swedish, a disfluent *att* ‘that’ is evidence of cognitive processing of more complex syntactic structures.

The complexity theory was also used as explanation for FPs in Japanese. Watanabe et al. [5] examined spontaneous Japanese speech corpora and showed that constituents tend to be longer or more complex when they are immediately preceded by FPs than when they are not, they found that

FPs cause listeners to expect that the speaker is going to refer to something that is likely to be expressed by a relatively long or complex constituent. Den [6] reported that clause-initial Japanese conjunctive *de* ‘and/then’ tends to be prolonged when a pause or fillers succeeds it, whereas clause complexity does not affect the duration of clause-initial ‘*de*’.

Such a perspective was also adopted for written text by Drescher [7], who showed that a kind of complexity approach can also be implemented on the Tiberian Hebrew accent system (*te’amim*). He demonstrated two cases of long subjects and relatively short verb phrases, where the main break (pause in Drescher’s terminology) falls between the subject and the verb; and two other cases with short subjects, where the main break falls after the verb. Drescher [7] concluded that “The difference between the two types of cases has to do with the length and prosodic complexity (i.e. number of phrases) of the subject relative to the verb phrase” [7:25].

The syntactic complexity, though, is only one possibility to explain why speakers elongate certain linguistic increment. Givón [8] talks about possible mapping relations between linguistic and cognitive complexity as “Coding: More complex mentally-represented events are coded by more complex linguistic/syntactic structures; Processing-I: More complex mentally-represented events require more complex mental processing operations. ... Processing-II: More complex syntactic structures require more complex mental processing operations.” [8:283].

Taking the coding principle into account, HDs, assumed to be considered as “more coding material”, might therefore be used for “less predictable” and “more important” information, and not always due to “more syntactically complex”. Ariel [9] and Hudson [10] suggest a “larger chunk” includes *phonologically* larger, and not syntactically larger. Ariel [9] argues that the ease with which a piece of given information is processed reflects its degree of mental accessibility, and that representations of linguistic material and physically salient objects are assumed to be in the short-term working memory, as opposed to representations of encyclopedic knowledge, which are assumed to be in long-term memory. Two principal criteria of Accessibility Theory (AT) associated with specific degrees of accessibility may be relevant to explaining the predicted linguistic element after HDs: The first is *informativity*. Accessibility markers representing a low degree of accessibility incorporate more lexical information than those representing a high degree of accessibility (e.g. open lexical categories vs. closed categories). Second, the *attenuation* criterion (i.e., phonological size) states that all things being equal, the less accessible an entity referred to by an expression is, the larger the expression is phonologically. This criterion also refers to the difference between stressed and unstressed forms. Shorter and unstressed forms have a higher degree of accessibility (e.g., function words with CV syllabic structure, as /be/ ‘in’ in IH or /ga/ ‘but’ in Japanese) than longer and stressed forms (e.g., verbs, proper names, etc.).

Hudson [10] argues that Dependency Grammar theory allows us to count the number of active dependencies, defining

a dependency as active if either the head or the dependent are still awaited. An active dependency is satisfied as soon as the word concerned is encountered [10:275–279]. At that point, the burden on the working memory decreases and more space remains for continuous processing of information.

2. Research goal

Considering the different theories that were introduced above, the research goal of this paper is to find out the linguistic role of HDs, by taking an account of the *syntagmatic aspect* and the *function* HD plays. We will focus on specific syntactic structures – phrases, and their interface with HDs. Since phrases consist, both in Hebrew and Japanese, as well as in other languages, on a head and a nominal increment, a study of phrases' interface with HDs can shed light on the relations between *form*, the prolonged material, and *function*, the syntactic increment that is prolonged and the following increment(s). To sum up, in this paper we provide evidence from spontaneous spoken IH and Japanese corpora to tackle the question how theory can explain HDs on varied syntactical increment, since when we consider disfluent phenomena of prosody and syntax, “explicit procedure must be designed to deal with these problems”, as mentioned before in Maruyama [11:782].

3. Data

3.1. Israeli Hebrew spontaneous corpus

The findings of IH presented in this paper are taken from 19 audio segments from 19 different recordings that were selected from CoSIH – Corpus of Spoken Israeli Hebrew [12]. The recordings, which were taken during 2001–2002, are of authentic Israeli Hebrew everyday conversations. Each dialogue consists of conversations between one core speaker and various interlocutors with whom the speaker interacted on that day. The research corpus consists of 31,760 word-tokens (over 6 hours of speech) of which 4,289 are word-types. 44% of the examined material is one-side telephone conversations; while 56% is face-to-face dialogues. The prosodic boundary tone inventory consists of 9,400 annotated boundary tones. The present research focuses on the 764 hesitation disfluencies that consist of 10.72% of all detected prosodic boundaries.

3.2. Japanese spontaneous corpus

The findings of Japanese presented in this paper are taken from the sample S01F0183 of CSJ (Corpus of Spontaneous Japanese) [13]. In this casual monologue a young woman makes her speech about the experience of travel to Hokkaido. In CSJ, elongated points are tagged by “<H>” (in this paper they will be tagged:⚡). These are carefully heard and transcribed by transcribers who are well-trained. Final Boundary Tone (FBT) is also transcribed at each end of phrase. FBT tags include L%, H%, HL%, and so on.

Japanese is a pitch accent language, and each accentual phrase (AP) basically ends with low pitch (L%). But sometimes AP ends with high pitch, in that case H% is tagged. And sometimes (or more frequently by some speakers) low pitch comes after H% with elongation, which is tagged by HL%. HL% can be regarded as not the case of elongation with hesitation, but particular speaking style. So ⚡ tags with HL% are excluded from the analysis. Other excluded cases are of ⚡ at sentence-final particles, since they mark the end of S/S' and it is often elongated. These elongations do not reflect the hesitations.

The sample S01F0183 consists of 2,129 words, including total of 179 elongations. The examples discussed here regard the 46 cases of elongations, with pre- and post- context.

4. Word order

4.1. Israeli Hebrew word order

Among the “basic orders” found in languages of the world, Hebrew is said to prefer a SVO word order. Nevertheless, Israeli Hebrew word order is relatively free and all possible alternatives can appear in specific contexts, e.g. literature and poetry. In the verb system, Israeli Hebrew morphology is characterized by the non-concatenative Semitic type structure. One of the relevant issues concerning verbs is that verbs are also accompanied by affixes indicating tense, person, number, and gender. Rosén [14] suggested considering the preposition as forming one constituent together with the verb: “The preposition constitutes the government properties of the verb” [14:169–170]. Rosén presented an example of the prepositions /e/ ‘to’, /be/ ‘in’ and /al/ ‘on’, and noted that, with the occurrence of certain verbs, these prepositions have no substitution, and function as cases (such as the accusative case marker /et/ ‘Acc.’). Nevertheless, Hebrew, as a “non-strict word-order” language, does not allow clitics and affixes at the phrase final position. Thus, the preposition stranding phenomenon does not occur in Hebrew. This characteristic of Hebrew means that we will not find prepositions in clause final position or in phrase final position (although this syntactic constraint is overruled in case of few coined idioms).

4.2. Japanese word order

Since Japanese is an agglutinative language, particles (case-marking particles, topic-marking particles, conjunctive particles, etc.) are put *after* noun or S'. It means that the function words of ‘to’ and ‘that’ are located *after* the content words, for example a noun or S' (2).

[*boku* *wa* *sueeden* *ni* *ikitai* *to* *omotta*] (2)
 I topic Sweden DIR want to go **QUOT** think-PAST
 ‘I thought that I want to go to Sweden.’

The structure of the sentence above is analyzed as (3).

[[*boku wa*] [[[*sueeden ni*] *ikitai*] *to*] *omotta*] (3)

One can see /*wa*/ ‘topic’, /*ni*/ ‘to’, and /*to*/ ‘that’ are located after the noun/S'. No matter how the quoted S' becomes long and complicated, /*to*/ ‘that’ always comes after the whole structure of S'. Even so, as we will show the data later, particles are sometimes elongated.

5. Results

5.1. The elongated words

5.1.1. Israeli Hebrew results

The results of the present research on spontaneous Israeli Hebrew showed that HDs occur within syntactic units, even intra-phrases or intra-morphemes, for example: [le: -sader] ‘to: arrange’, and basically that these hesitated increments are function words. Moreover, according to the results, the elongated words vary in terms of complexity. For example, elongated definite article /*ha*/ ‘the’, that was found as the second most probable ($p=0.938$) elongated word in the corpus (149 such occurrences, as shown in Table 1), predicts a noun (i.e. simple structure) as its complement, while the subordinate conjunction /*ʃe*/ ‘that’ which predict a complex S' structure, was also found among the five most probable ($p=0.875$) elongated words (42 such occurrences, as shown

in Table 1). In brief, HDs were observed to split four types of dependency relationships. These are outlined below and ranked from the weakest to the strongest dependency:

The weakest type of dependency is of course when HD does not split a syntactic dependency, that is, when the word carrying the tone and the following word are not in any direct syntactic relationship (not in the same syntactic phrase or clause). HDs typically *follow* discourse markers, which are considered a “no dependency” type.

Coordinate structure dependencies occur between a conjunction and the adjacent syntactic structures (from increments to sentences). HDs are more likely to occur *after* the conjunction.

The subject-predicate dependency represents the cases where HDs occur between personal pronouns and verbs. This subject-predicate dependency is considered “stronger” than the two mentioned above, since in spontaneous IH they imply that HDs split intra-clausal structures.

Intra-phrasal sequences are the “strongest” type of dependencies that are broken by HDs, and indeed HDs were found to be the most likely to occur intra-phrasal.

The results of the distribution of elongated POS are presented in Table 1. The cases in lines 2, 4, 6, and 7 are considered intra-phrasal in IH.

5.1.2. Japanese results

In Japanese it was found that elongations in S01F0183 can be classified mainly into two types: A. elongations at the end of function words; B. elongations at the end of content words. Type A consists of conjunctives and particles, and type B mainly consists of nouns.

Type A-1) elongation of conjunctives:

The most frequent type was the case where a conjunctive was elongated at the beginning of an utterance [6]. Conjunctive /*de*/ ‘then’ appeared 38 times in the target data, and almost 40% of them (15 times) were elongated as ‘*de:*’. Examining the context, these elongations can be considered that the speaker utters ‘*de:*’ at the beginning of an utterance as a discourse marker. Since this talk is a monologue, the speaker as to keep on producing her message. So this kind of elongation can be classified as a preventive device of silence; it signals that the speaker is under processing and has more to say. For example.

[*atatta toka itte: de: Hokkaido ryoko no tikketo*] (4)

won QUOT said and Hokkaido travel of ticket
‘she said she won the quiz prize, and she had tickets to Hokkaido.’

Type A-2) elongation of particles:

Particles are sometimes elongated. There are some categories in Japanese particles, and major types are “conjunctive particle”, “case-marking particle”, and “topic-marking particle”. In the case of “case-marking” and “topic-marking” particles, the elongations occur in the middle of the utterance, which seem to indicate that the further part of ongoing utterance is now under processing. “Conjunctive” particles are sometimes elongated too, but these particles appeared at the end of clauses, so it signals that the speaker will continue her next utterance, as the case of conjunctives.

Type B) elongation of nouns:

Nouns are sometimes elongated at the final syllable (rarely inside the noun). These cases are elongation within noun phrases (noun:+particle) or predicates (noun:+copula marker). This type can be considered as a signal of difficulty

of fluent production of the ongoing utterance. The speaker confirms the validity of information.

In a subset of whole CSJ with 443,732 words that was retrieved for this research, we confirmed almost the same tendency of elongated POS as the cases observed in S01F0183.

The results of IH and Japanese distribution of elongated POS are presented in Table 1.

Table 1: *Distribution of the most probable elongated (HD) POS in IH and Japanese.*

	Elongated POS in CoSIH	Occurrences	Elongated POS in CSJ	Occurrences
1	Conjunction	149	Noun	18
2	Definite article	149	Conjunction	16
3	Personal pronoun	111	Particle	9
4	Preposition	100	Adjective	2
5	Subordinate conjunction	42		1
6	Modifier	36		
7	Possessive particle	22		
8	the lexeme /hjj/ ‘be’	17		
9	Existential lexeme (Auxiliary)	11		
10	Modal lexeme (Auxiliary)	10		

5.2. What follows HDs?

Results in IH show that there is wide variation in the *following* elements to the HDs, and this for itself is a simple argument that rejects complexity as an ultimate explanation for the HD phenomenon. Results showed no priority to syntactically complex structures in the expected following structures. Another argument are cases of elongated construct state nouns (5a–b) and infinitive prefixes (6), with the prediction of only simple following structures, a noun and a gerund, respectively.

Construct-state nouns in IH (5)

a. [ze jihje be m sof e: ogust].
‘it will be at ehm end of August.’

b. [milxemet e: xamiʃim ve ʃeʃ].
war-of: fiftv and six.
‘the 1956 war.’

Infinitives in IH (6)

[holex li: -krot]?
‘going to: happen?’

Such cases were also found in Japanese (7). In this case the speaker remembered the name of place *Nemuro*, and then she repeated the word within a noun phrase (*nemuro made*) with elongation.

Intra-phrasal HDs in Japanese (7)

[nemuro da nemuro: made kuruma de itte:]
‘It was Nemuro, we went to Nemuro: by car.’

6. Discussion

The explanation that is taken to apply to the findings of IH corpus is that of syntactic planning coming before lexical planning [15, 16]. To this statement, another term – the placeholder [16] is added. In spontaneous speech, placeholders “mainly have a pronominal origin and serve as a preparatory substitute for a delayed constituent” [16:11] and placeholders are “among other lexical and grammatical resources that allow the speaker to refer to object and events for which the speaker fails to retrieve the exact name, or simply finds the exact name to be unnecessary or inappropriate” [16:11]. Both [15] and [16] assume the pronominal nature of disfluencies or placeholders. In this respect, HDs are prosodic morphemes which also have a pronominal nature. This is to say that IH speakers first utter the syntactic frame with the elongated word, the lexical element that “carries” the HD with its pronominal nature. The elongated word is expected to be followed by a syntactic increment or a target word. It is suggested in [1] that by the elongated production of function words the expected syntactic structure is already indicated, meaning the syntactic structure is thus complete. In other words, the following lexeme(s) to the HD does not contribute a fundamental increment to the *structure*, only to the *content*. This mechanism reduces the burden of the working memory, and thus enables processing of new information.

The question is whether the pronominal nature of HDs should be adapted to Japanese as well. In case of Japanese, not only function words but also content words (nouns) are sometimes elongated. In (8), three cases of elongation appeared; *de:*, elongation of the conjunctive (discourse marker) at the beginning of utterance, *sore:* *ga*, elongation at the final syllable of noun in noun phrase (noun+case marking particle), and *nido:* *me*, elongation within the compound noun.

[*de:* *watasi wa* *sore:* *ga* *hokkaido wa* (8)
and: I TOP **it:** SUBJ Hokkaido TOP
nido: *me* *datta n desu keredomo*]
second times Copula-PAST though
 “and it was the second time for me to go to Hokkaido.”

In this situation the speaker is under the constraint to produce a series of episodes linearly in real time (cf. linearization problem, [17]). Elongations at the end and within the nouns can be considered to be used; 1) to make a plan of further incremental utterance which is roughly designed at the beginning, or 2) to confirm the validity of information just mentioned by the noun. As shown in the cases of IH, these elongations also reduce the burden of the working memory.

As we have shown in Table 1, the distribution of elongated POS between IH and Japanese is different in some aspects, especially content words in Japanese. The difference may be caused by syntactic and/or pragmatic reason, which indicates that a typological study on disfluencies is possible.

7. Conclusions

To conclude, the comparative evaluation of the two languages which are different in their word order characteristics had shown that HD phenomenon is a matter of “processing time, cognitive complexity or mental effort”. Syntactic complexity, if exists, can be explained as reflecting the “coding principle” [8] of the cognitive process.

According to the approach adopted here, what is common to all elongated words is the fact that they imply

continuity, regardless of whether they are dependents of heads (i.e., nouns in NP) or heads of dependents (i.e., case marking particle in NP). What should be stressed here is that they share a [+dependency] syntactic feature. It can be said that what is actually elongated is not the word itself (or a syllable of the word), but this syntactic feature. The [+dependency] feature shows that “there is more to come”, i.e. what are the communicative intentions of the speaker. It allows the speaker to think, if either the function word or the content word of a phrase are still awaited, by elongating structures. In our view, what is common to elongated grammatical elements is a [+dependency] feature.

We tried to connect between syntagmatic aspect and the linguistic function of HD, and to propose a pronominal approach to the phenomenon of HDs, which we explained in both languages. Nevertheless, it must be also related to cognitive process, as mentioned, or can be considered as a signal of difficulty of fluent production of the ongoing utterance.

8. References

- [1] V. Silber-Varod, *The SpeeCHain Perspective: Prosodic–Syntactic Interface in Spontaneous Spoken Hebrew*, PhD dissertation, Tel Aviv University, 2011.
- [2] E. Shriberg, “To ‘errrr’ is human: Ecology and acoustics of speech disfluencies”. *JIPA* 31(1), pp. 154–169, 2001.
- [3] H. H. Clark & T. Wasow, “Repeating words in spontaneous speech”. *Cognitive Psychology* 37, pp. 201–242, 1998.
- [4] M. Roll, J. Frid and M. Horne, “Measuring syntactic complexity in spontaneous spoken Swedish”. *Language and Speech* 50 (2), pp. 227–245, 2007.
- [5] M. Watanabe, K. Hirose, Y. Den and N. Minematsu, “Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners”. *Speech Communication* 50(2), pp. 81–94, 2008.
- [6] Y. Den, 2009. “Prolongation of clause-initial mono-word phrases in Japanese”. In S.-C. Tseng (ed.), *Linguistic patterns in spontaneous speech*, Taiwan: Institute of Linguistics, Academia Sinica, pp. 167–192, 2009.
- [7] B. E. Dresher, “The prosodic basis of the Tiberian Hebrew system of accents”. *Language* 70(1), pp. 1–52, 1994.
- [8] T. Givón, *The Genesis of Syntactic Complexity*. Amsterdam: John Benjamins Publishing, 2009.
- [9] M. Ariel, “Accessibility theory: An overview”. In T. Sanders, J. Schliperoord, and W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects*, Amsterdam: John Benjamins, pp. 29–87, 2001.
- [10] R. A. Hudson, “Do we have heads in our minds?” In G.G. Corbett, N.M. Fraser and S. McGlashan (Eds.), *Heads in grammatical theory*, Cambridge: Cambridge University Press, pp. 266–291, 1993.
- [11] T. Maruyama, “Speech segmentation by clausal and non-clausal boundaries in Japanese”. *The Fifth International Conference on Cognitive Science*, Abstracts Volume 2: Workshop “Spoken discourse corpora as a window on cognitive mechanisms of speech production”, Kaliningrad, Russia, pp. 782–783, 2012.
- [12] The Corpus of Spoken Israeli Hebrew (CoSIH), available at: <http://humanities.tau.ac.il/~cosih/english/>
- [13] K. Maekawa, “Corpus of spontaneous Japanese: Its design and evaluation”. *Proceedings of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, p. 712, 2003.
- [14] H.B. Rosén, *Contemporary Hebrew*. The Hague: Mouton, 1977.
- [15] C. Blanche-Benveniste, “Linguistic analysis of spoken language: The case of French language”. In Y. Kawaguchi, S. Zaima and T. Takagaki (eds.) *Spoken Language Corpus and Linguistic Informatics*. Amsterdam: John Benjamins, 2006.
- [16] V.I. Podlesskaya, 2010. “Parameters for typological variations of placeholders”. In N. Amiridze, B. H. Davis, and M. Maclagan (Eds.), *Fillers, pauses and placeholders. Typological Studies in Language*, Amsterdam: John Benjamins. pp. 11–32, 2010.
- [17] W.J.M Levelt, *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press. 1989.