Indicating Dependency between Spoken Sentences by Prosodic Means

Mari Wiklund
Indicating Dependency between Spoken Sentences by Prosodic Means

Mari Wiklund
University of Helsinki


URL : http://journals.openedition.org/discours/9675

Titre du numéro : Varia
Coordination : Benjamin Fagard & Denis Vigier
Date de réception de l'article : 15/01/2018
Date d'acceptation de l'article : 24/04/2018
Indicating Dependency between Spoken Sentences by Prosodic Means

Mari Wiklund
University of Helsinki

In certain languages and in certain types of data, prosodic features are used to group prosodic units called by some scholars spoken sentences into larger units, often called speech paragraphs. The onsets of the first spoken sentences of these units are marked by a raised pitch level compared to the beginning of the preceding spoken sentence. The prosodic units that are formed typically correspond to topical entities. Prosodic means are also used to indicate relationships between spoken sentences belonging to the same speech paragraph. A new spoken sentence generally starts at a higher pitch compared to the end of the preceding spoken sentence. However, if two consecutive spoken sentences are closely related, the second begins at a lower pitch vis-à-vis the end of the previous spoken sentence, but the subsequent syllable of the new spoken sentence displays raised pitch. This phenomenon of sentence-initial lowered pitch indicates, on one hand, a close discourse relationship with the preceding spoken sentence but, on the other, the beginning of a grammatically independent spoken sentence. The data are in Finnish, and they come from conference-like, monologous presentations that are translated by speech-to-text interpreters to allow the speech to be accessed by the deaf and hard-of-hearing. The analyses are performed using instrumental phonetics methodology.

Keywords: prosody, Finnish intonation, paragraph intonation, Finnish spoken language, discourse analysis, dependency

1. Introduction

It is well-known that prosodic features play an important role in the structuring of discourse. Prosodic means are used, for example, in marking boundaries between units of different sizes as well as in grouping units of speech into larger wholes. This grouping is often referred to as “paragraph intonation”, because of its basic unit, termed a “speech paragraph” (or a “spoken paragraph”, Wichmann, 2000). The pitch on the first accented syllable of a speech paragraph is extra high. This phenomenon has been called “topic reset”, even if it does not always indicate a clear change of topic (Wichmann, 2000: 25). The phenomenon seems, however, to be typically related to the indication of topical transitions, such as a change of topic or subtopic (Barth-Weingarten, 2009; Chafe, 1994). Sometimes the phenomenon can also indicate the beginning of a new action (Couper-Kuhlen, 2004). I will refer to this phenomenon in this study simply by the term “sentence-initial high pitch” (SIHP). Nafâ Waasaf (2007) has shown that in simultaneous interpreting both the speakers

---

1. The notion of the *paratone* has also been used to signify a paragraph-like unit of speech (Brown, 1977; Brown et al., 1980; Yule, 1980).
and the interpreters treat a SIHP as a sign that indicates the beginning of a new topical unit. The same applies to speech-to-text interpreting (Wiklund, 2014). Unaccented syllables before the first accent also tend to be high (Couper-Kuhlen, 2006; Wichmann, 2000). The end of a speech paragraph is generally indicated by an extra-low pitch, close to the speaker’s baseline, and often by a noticeable pause (Couper-Kuhlen, 1986 and 2006; Wiklund, 2014).

The term “spoken sentence” refers generally speaking to a unit which, despite its name, is not in the first instance a syntactic unit but rather a prosodic one (Couper-Kuhlen, 2006; Wichmann, 2000). The beginning of a spoken sentence is indicated by a high pitch on the first accented syllable of an intonation phrase. The end of a spoken sentence, in turn, is signaled by a falling pitch starting on or from the last accented syllable of an intonation phrase and reaching a low point in the speaker’s voice range (Couper-Kuhlen, 2006; Wichmann, 2000). Canonically, the accents in the spoken sentence form a pitch line that gradually descends, or declines, throughout the unit (Couper-Kuhlen, 1986 and 2006). Laver (1994) has postulated a similar line of declination for amplitude. Spoken sentences may consist of several intonation phrases: if there are several, they are linked by a single declination line for pitch and amplitude (Couper-Kuhlen, 2006; Wichmann, 2000). Spoken sentences do not always correspond to a syntactic/orthographic sentence; the same prosodic pattern is also used for example in titles and other noun phrases.

In this study, a “spoken sentence” refers to a prosodic and a syntactic unit, which most often ends in a silent pause, but its end can also be indicated by a breathing (during which the informant does not speak) or a filled pause. If a silent pause, a filled pause or a breathing occurs inside a syntactic unit, it does not interrupt the spoken sentence. Thus, in this study, both syntactic and prosodic criteria were used in the definition of the boundaries of a spoken sentence.

Sometimes it was rather difficult to define if the boundary between two spoken sentences was located before or after a discourse particle. In these cases both prosodic and syntactic criteria were used to define the location of the boundary: basically, if the discourse particle was preceded by a silent pause, a filled pause or a breathing, it was considered to start a new spoken sentence. However, if the preceding unit did not constitute a complete syntactic whole (for example, if it consisted only of other discourse particles), the silent pause (or the filled pause or the breathing) did not interrupt the spoken sentence. If, on the other hand, the pause (or the filled

2. In my own previous study (Wiklund, 2014), however, speech paragraphs also rather often ended in a pitch rise, when the speaker was female. The data of this study, too, come from Finnish speech-to-text interpreting situations.
3. According to Szczepk Reed (2009: 351), an intonation phrase is generally speaking "a spate of talk delivered as one recognisable overall pitch movement".
4. Naturally, this rule does not apply to all languages and to all types of data.
5. The perception of sentence and paragraph boundaries is approached in Kreiman (1982).
pause or the breathing) occurred after the discourse particle and the preceding unit constituted a complete syntactic whole, it was considered to interrupt the spoken sentence; that is, a new spoken sentence started after it.

Spoken sentences of the same speech paragraph are typically linked internally via an overriding declination line (e.g., Sluijter & Terken, 1993; Wichmann, 2000). That is, there is a “supradeclination” between the beginnings of spoken sentences occurring inside the same speech paragraph: a new spoken sentence generally starts on a lower pitch level than the preceding one. This applies especially to data coming from highly controlled experimental settings (e.g., Sluijter & Terken, 1993). In data coming from naturally occurring situations, the supradeclination is not systematic (Wichmann, 2000: 121). Indeed, according to Wichmann (2000: 121), the supradeclination constitutes an “envelope” within which linguistically motivated variation operates. For example, the information structure of a speech paragraph plays a role in determining on which level each spoken sentence starts. It is possible that, for instance, reformulations and precisions start on a rather low level, and after this, a spoken sentence that continues the actual topic start again on a slightly higher level (Wichmann, 2000). It is noteworthy that in dialogues and in spontaneous speech the phenomenon may take very different forms, which should be further studied.

Indeed, prosodic means are not only used to mark the boundaries between speech paragraphs; they also work to indicate relationships between spoken sentences belonging to the same speech paragraph. A new spoken sentence generally starts at a higher pitch level compared to the end of the preceding spoken sentence. This phenomenon is called “pitch reset” (Wichmann, 2000: 24). If two consecutive spoken sentences are closely related, however, the second begins at lower pitch than the end of the previous spoken sentence. That is, the unaccented syllables occurring in the beginning of the spoken sentence are produced on a remarkably low pitch level6. The first accented syllable – which is the second or a subsequent syllable of the spoken sentence – then carries the pitch reset phenomenon. This phenomenon indicates, on one hand, a close relationship between the spoken sentences, but, on the other, the beginning of a grammatically independent spoken sentence (Wichmann, 2000: 75-78). The first syllable carrying the lowered pitch is most often a discourse particle7 or a grammatical word (such as a conjunction or pronoun). However, the phenomenon does not concern all sentence-initial discourse particles and grammatical words. Indeed, the spoken sentences in these data rarely start with a content word; however, the majority begin with raised pitch rather than lowered pitch compared to the end of the preceding spoken sentence. That is, even if the

---

6. In this study, the threshold for defining a low pitch is −1 semitone compared with the end of the preceding spoken sentence, because this difference is perceivable.

7. In this study, “discourse particles” are words that are uttered not because of their contribution to propositional content, but because of some pragmatic function for the ongoing discourse (Stede & Schmitz, 2000: 126).
so-called phenomenon of “sentence-initial lowered pitch” (SILP) is typical of discourse particles and grammatical words, most of the spoken sentences beginning with these words do not display the phenomenon in question.

The fact that the SILP phenomenon concerns the initial position of spoken sentences is essential. Indeed, previous studies have shown that this position plays a strategic role in the structuration of discourse (Degand & Simon, 2009; Martin et al., 2014); it typically has textual functions as well as functions related to coherence and topical management (Martin et al., 2014). In other words, it constitutes an anchor point where the contents of the discourse can be connected in a coherent manner to the preceding contents (Virtanen, 2004: 80-81). According to Ho-Dac (2007), this position is also strategic for the development of discourse.

The main objectives of the study consist in showing 1) that the SILP phenomenon can be found both in male and female informants’ speech in Finnish data consisting of conference-like, monologous presentations, and 2) that the phenomenon in question plays a role in the indication of dependency between spoken sentences in these data. My hypothesis is that in these data, the SILP phenomenon acts as a discourse-structuring tool that cooperates with the pragmatic and the semantic features of discourse in the indication of dependency. The way how the phenomenon is marked in the printed text typed by the speech-to-text interpreters is also taken into account in this study.

Aho (2010) has developed a model for the prosodic segmentation of spontaneous speech. This model, called the “IJ-model of intonational chunking”, is based on Finnish and Finland Swedish data, and it is inspired by Sadeniemi’s (1949), Sovijärvi’s (1946 and 1954), Bruce’s (1994 and 1998), Bolinger’s (1972, 1986 and 1998) and Chafe’s (1993 and 1994) works. The smallest unit of the IJ-model is a prosodic word (Aho, 2010: 35-36 and 42-43). A prosodic word corresponds most often to a lexical word, but it can also be shorter or longer (Bruce, 1998: 80 and 124-126). For example, a long compound word can belong to two different prosodic words. A prosodic word is most of all a tonal unit, and it normally includes one clear melodic peak. A prosodic word consists of one accented syllable, and one or several unaccented syllables. In all prosodic words, one part is normally more prominent than the others, but prominence cannot always be measured acoustically. A minor intonation unit, in turn, is generally based on audition (Aho, 2010: 39-42). Its boundaries can be marked with changes in pitch, volume, speech rate or quality of voice, or with other phonetic features occurring alone or together. The duration of a minor intonation unit is usually 1-2 seconds. This might be related to rhythmic

---

8. I have written previously a brief conference article on the same phenomenon (Wiklund, 2013a). It is also treated in a Finnish article (Wiklund, 2013b), which summarizes my studies carried out within the framework of an interdisciplinary research project, titled “Print interpreting: process, comprehensibility, and technology”.

9. The data will be presented in more detail in Section 2.0.

10. IJ is shortened from intonaatiojaksomalli (“model of intonational chunking”).
factors as well as to the limited capacity of working memory (Aho & Yli-Luukko, 2005: 210). Indeed, Baddeley (1997: 21-22) claims that the working memory is able to store information only for 2-20 seconds. A minor intonation unit does not necessarily always have semantic content 11, but it typically has a pragmatic value 12. Rhythm is an important factor in the definition of the boundaries between minor intonation units. A major intonation unit includes one or generally several minor intonation units (Aho, 2010: 35-39). Its duration varies, but it is normally 5-10 seconds. It typically contains related material, and it almost always ends in a pause. It constitutes a semantically coherent unit. Pauses occurring between major intonation units are usually rather long and clearly indicate the boundaries between the units. If there is a pause inside a major intonation unit, it either constitutes a minor intonation unit in its own right or it belongs to another minor intonation unit. Generally a major intonation unit starts on a high pitch level, and the pitch curve gradually declines towards the end of the unit. In the beginning of a major intonation unit the amplitude usually gets remarkably larger, and then it gets smaller towards the end (Aho, 2010: 36). Sometimes only the shape of the amplitude may be enough to signal the boundaries between intonation units. In my study, a spoken sentence corresponds to a “major intonation unit” in Aho’s study (2010). Chafe’s “sentence” (Chafe, 1994: 140) or “center of interest” (Chafe, 1980: 26) are also reminiscent of Wichmann’s (2000) spoken sentence. Different units of Aho’s (2010) model are summarized in Table 1 below. The table has been quoted from Aho (2010: 46) 13.

<table>
<thead>
<tr>
<th>Prosodic unit</th>
<th>Strong criteria</th>
<th>Weak criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major intonation unit</td>
<td>- Clear acoustic boundaries (for example a pause)</td>
<td>- Contents</td>
</tr>
<tr>
<td></td>
<td>- Global shape of the f0 curve</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Global shape of the waveform</td>
<td></td>
</tr>
<tr>
<td>Minor intonation unit</td>
<td>- Rhythm (duration usually less than 2 seconds)</td>
<td>- Contents</td>
</tr>
<tr>
<td></td>
<td>- Global shape of the f0</td>
<td>- Acoustic boundaries (change in the f0 curve and/or the waveform and/or a pause)</td>
</tr>
<tr>
<td></td>
<td>- Global shape of the waveform</td>
<td></td>
</tr>
<tr>
<td>Prosodic word</td>
<td>- Rhythm (one tonal rhythm figure which includes an accented syllable and unaccented syllables related to it)</td>
<td>- Acoustic boundaries</td>
</tr>
<tr>
<td></td>
<td>- Local shape of the f0 curve</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 – Different units of Aho’s (2010) model (quoted from Aho, 2010: 46)

11. The “semantic content” refers here to the meanings of words. Thus, in a semantically coherent unit the meanings of the words are related to each other.
12. The “pragmatic value” refers to the use of language.
13. The contents of the table have been translated from Finnish into English by the author of this article.
2. **Data and methods**

The data are in Finnish, and they come from four situations (with four different speakers) where conference-like, monologous presentations are interpreted by professional speech-to-text interpreters. In speech-to-text interpreting, also called print interpreting, speech is simultaneously transformed into written format and displayed on a screen so it can be accessed by the deaf and hard-of-hearing (Borg, 2008; Mäkiranta, 2006; Ohrankämmen, 2008; Salakari, 2008; Tanhuamäki, 2008; Tiittula, 2006 and 2009). In such situations, the role of prosody is particularly important, as prosodic features help speech-to-text interpreters understand the structure of the spoken source text and transmit it to the written target text. Speech-to-text interpreters are trying to include all audible information in the written output, but since speaking is much faster than writing, in spite of a high typing rate, errors and omissions occur. The text appears on the screen letter by letter as the speech-to-text interpreter writes, and recipients see the real-time writing process, including pauses and corrections (Sharmin et al., 2016: 120).

All four speakers are native speakers of Finnish. Two of them are female, and two are male. The duration of one of the female informant’s talk is 9:30 minutes (101 spoken sentences), and it deals with autism spectrum disorders. It is the first part of two presentations given by the same speaker. The other female speaker talks about the empowerment of the deaf and hard-of-hearing. The duration of the talk is 8:13 minutes (87 spoken sentences), and it is the beginning of a long lecture. The presentation of one of the male speakers is 8:30 minutes long (60 spoken sentences). It is the beginning of a long lecture dealing with the activities of the Finnish Federation of Hard of Hearing. The other male informant talks about seeing and lighting. The passage that is studied is 7:54 minutes in duration (66 spoken sentences), and it is also the beginning of a long lecture. Thus, the total duration of the data in which all 314 spoken sentences are included is 34:07 minutes. The content of the talks was prepared in advance. The speakers nevertheless give their presentations rather spontaneously, without reading from script. The description of the data is summarized in Table 2 below.

<table>
<thead>
<tr>
<th>Informant</th>
<th>Duration (minutes)</th>
<th>Number of spoken sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female informant 1</td>
<td>9:30</td>
<td>101</td>
</tr>
<tr>
<td>Female informant 2</td>
<td>8:13</td>
<td>87</td>
</tr>
<tr>
<td>Male informant 1</td>
<td>8:30</td>
<td>60</td>
</tr>
<tr>
<td>Male informant 2</td>
<td>7:54</td>
<td>66</td>
</tr>
<tr>
<td>Whole data</td>
<td>34:07</td>
<td>314</td>
</tr>
</tbody>
</table>

Table 2 – Description of the data
Theoretically, the study is primarily based on Wichmann’s (2000) discourse analysis approach to intonation, and the IJ-model of intonational chunking developed by Aho (2010). Pitch level (f0) is measured on the last syllable of the preceding spoken sentence, the first syllable of the “new” spoken sentence, and the subsequent syllable carrying raised pitch. In Finnish words, primary stress always falls on the first syllable of the word (Iivonen, 1998: 315; Iivonen & Aulanko, 2001: 26; Hakulinen et al., 2004: § 13; Sadeniemi, 1949). Consequently, the first accented syllable (onset) of a spoken sentence is generally – but not always – its first syllable. The discourse-structuring function of the pitch reset seems to override sentence stress. That is, a word that may not be important at the level of one spoken sentence can be accented, for example, in order to indicate relationships between spoken sentences. Most of the spoken sentences in these data begin with an accented or an unaccented discourse particle carrying raised pitch vis-à-vis the end of the preceding spoken sentence. In spoken sentences displaying the phenomenon of SILP, the first syllable of the spoken sentence carries a remarkably low pitch compared to the end of the previous spoken sentence, and the second or a subsequent syllable of the spoken sentence carries a raised pitch. In terms of Mertens (2008), the SILP phenomenon constitutes a “prosodic signified” (signifié prosodique) the function of which is related to the structuration of discourse.

According to Iivonen (1998: 316), Finnish stress is relatively weakly expressed. The prominence of the default stress depends on the type of word: content words typically carry more prominent stress than function words (Iivonen & Aulanko, 2001: 26-27; Sadeniemi, 1949). However, as the spoken sentences in these data rarely start with a content word, differences related to parts of speech are not salient here.

In my data, a “spoken sentence” corresponds to a prosodic unit which most often consists of one grammatical clause, but it can also include several clauses, if they constitute only one prosodic unit. The end of a spoken sentence is indicated by a pause, a breathing (during which the informant does not speak) or a filled pause. A “pitch reset” occurs when the sentence-initial pitch rises on a higher level than at the end of the preceding spoken sentence. The SILP phenomenon occurs when a spoken sentence begins at lower pitch than the end of the previous spoken sentence. That is, the unaccented syllables occurring in the beginning of the spoken sentence are produced on a remarkably low pitch level. The first accented syllable – which is the second or a subsequent syllable of the spoken sentence – then carries the pitch reset phenomenon. Due to the background noise constantly present in these data, automatic boundary detection systems could not be used in this study. The

Notes:

12. The information structure and the structure of speech paragraphs have also been approached from different points of view for example in the following studies: Fezza (2016), Grobet & Simon (2003), Kreiman (1982), Morel (2011), Pickering (2004), Simon & Grobet (2003).

13. Martin (2004: 44) refers to a similar phenomenon of French by the term "melodic neutralization" (neutralisation mélodique); in this phenomenon, the first syllable of an utterance carries a flat pitch.

14. For automatic boundary detection, cf. for example Suni et al. (2016).
The phenomenon of SILP is rather frequent in the data. Indeed, 87 (27.7%) of the 314 spoken sentences in the data display this prosodic phenomenon. The Chart 1 below illustrates the proportion of the spoken sentences carrying the SILP phenomenon in the whole data corpus.

SILP is, however, not as frequent in all parts of the data. Indeed, in the data from the presentation on autism spectrum disorders (female speaker 1), the proportion of spoken sentences displaying SILP is as much as 46.5%, whereas in the other female informant’s talk (female speaker 2) the corresponding percentage is only 12.6%. In the talk concerning seeing and lighting (male speaker 1) the frequency of SILP is 27.3%. In the other male informant’s speech (male speaker 2), 18.3% of spoken sentences display the phenomenon in question. The variation between the speakers can be explained by the different lengths of the speech paragraphs: while the female speaker 1 – whose speech includes a lot of spoken sentences starting with the SILP phenomenon (46.5%) – typically produces long speech paragraphs, the female speaker 2 – whose speech includes much less SILP sentences (12.6%) – produces mainly rather brief speech paragraphs. The frequency seems to be related to the length of the paragraphs. But at the same time, it is related to the occurrences of discourse particles: if the paragraph is longer, there are also more sentence-initial discourse particles. The Chart 2 represents below the proportions of the spoken sentences carrying the SILP phenomenon in the different parts of the data corpus.

![Chart 1](http://journals.openedition.org/discours/9675)  
Chart 1 – Proportion of the spoken sentences carrying the SILP phenomenon in the whole data corpus

---

17. In Finnish, the letters correspond to phonemes (apart from some rare exceptions). Therefore separate phonological transcriptions were not necessary.
The following example illustrates two occurrences of SILP (lines 04–06). Here, the second spoken sentence, which displays SILP (ja se rokote annetaan samaan aikaan, “and the vaccine is given at the same time”, lines 04–05), elaborates on the first spoken sentence (Ja tota öhm .mth se: noh ehkä se rokote väärinnäröys jobituu osittain sitähän että usein se autismi puhkee niin kun puoltoistavuotiaana, “well er mm mth it well maybe the misunderstanding with vaccines is partly caused by the fact that autism often emerges at 1.5 years of age”, lines 01–04), and the third spoken sentence — which is also a SILP sentence — elaborates on the second spoken sentence (niin siinä näyttää olevan ikään kun kausaliteetti mitä ei sit otkeesti olekaan, “so there seems to be a causality that doesn’t actually exist”, lines 06–07). This type of spoken sentence, which elaborates on the previous one, is very typical of SILP. The extract comes from the presentation of female speaker 1. The speaker is describing the causes of autism spectrum disorders, after having just stated that they are not caused by vaccines. Now she is providing a hypothesis on the causes of the misunderstanding related to vaccines.

Example 1 18:

01 Ja tota öhm (1.7) .mth se: noh ehkä se rokote
     well er mm       mth it well maybe

18. A question mark in the transcriptions signifies rising pitch at the end of a spoken sentence.
02 väärinymmärrys johtuu osittain
the misunderstanding with vaccines is partly
03 siitäkin että usein se autismi puhkee niin
caused by the fact that autism often emerges
→ 04 kun puoltoistavuotiaana ? (. ) ↓ ja se ↓ and it
  at 1.5 years of age and
05 ↑ rokote # annetaan samaan aikaan. # (0.6)
↑ vaccine
the vaccine is given at the same time
→ 06 ↓ Niin ↑ sinä (0.6) / näyttää / olevan ikään kun
↓ so ↓ there
so there seems to be a kind of
07 kausaliteetti mitä sit ei / oikeesti olekaan /
causality that doesn’t actually exist

The second spoken sentence (ja se rokote annetaan samaan aikaan, “and the vaccine is given at the same time”, lines 04-05) starts with the word ja (“and”), which acts in this context as an additive discourse particle (Hakulinen et al., 2004: § 804); that is, it works to indicate that the spoken sentence that starts adds something to the preceding unit. Ja is followed by the demonstrative pronoun se (“it”) which refers typically to something that belongs to the “attention field” of the listener (Hakulinen et al., 2004: § 720). In this context, se acts as a kind of definite article related to the following word, rokote (“vaccine”), which has already been mentioned earlier (Larjavaara, 2001). Indeed, even if articles do not basically exist in Finnish, the word se is sometimes used as a kind of definite article referring to something that has already been mentioned earlier or something that is for another reason supposed to be known by the recipient (Larjavaara, 2001). These two words (ja and se) carry a low pitch, which emphasizes their above-mentioned pragmatic roles as well as the fact that their semantic value is low. Thus, this example shows how prosodic features often “cooperate” with pragmatic and semantic factors. The same applies to the third spoken sentence (niin sinä näyttää olevan ikään kun kausaliteetti mitä ei sit oikeesti olekaan, “so there seems to be a kind of causality that doesn’t actually exist”, lines 06-07), where the low pitch is carried by the discourse particle niin (“so”). Niin works here to connect two consecutive main clauses and to indicate a causal relationship between them (Hakulinen et al., 2004: § 811). The low pitch carried by the sentence-initial particle niin emphasizes this relationship.

The first spoken sentence of Example 1 (ja tota öhm . mth se: nob ebkä se rokote väärinymmärrys johtaa osittain siitäkin että usein se autismi puhkee niin kun
Indicating Dependency between Spoken Sentences by Prosodic Means

puoltoistavuotiaana, “well er mm mth it well maybe the misunderstanding with vaccines is partly caused by the fact that autism often emerges at 1.5 years of age”, lines 01-04) starts a new speech paragraph. The second spoken sentence (ja se rokote annetaan samaan aikaan, “and the vaccine is given at the same time”, lines 04-05) begins 1.6 semitones lower than the first spoken sentence. Thus, it belongs to the same speech paragraph. The first syllable of the second spoken sentence is produced 2.8 semitones lower than the last syllable of the preceding spoken sentence. Consequently, the phenomenon of SILP occurs here. Figure 1 illustrates the pitch curve during the production of the last minor intonation unit (vuotiaana, “at the age of”) of the first spoken sentence and the first minor intonation unit of the second spoken sentence (ja se rokote, “and the vaccine”). As the curve shows, the pitch rises slightly (2.0 semitones) during the third syllable ([ro]) of the second spoken sentence.

SILP indicates here, on one hand, the beginning of a new grammatically independent spoken sentence. On the other hand, there is a need to indicate the close anaphoric relationship between the two spoken sentences. Indeed, as the adverb samaan aikaan (“at the same time”) refers to puoltoistavuotiaana (“at 1.5 years of age”), which occurs at the end of the first spoken sentence, the second spoken sentence is asymmetrically dependent on the first.

The third spoken sentence (niin siinä näyttää olevan ikään kun kausaliteetti mitä ei sit oikeesti olekaan, “so there seems to be a causality that doesn’t actually exist”, lines 06-07) also begins with SILP: its first syllable ([niin]) is produced as much as 15.0 semitones lower than the end (syllable [kaan]) of the preceding spoken sentence. However, during the production of the second syllable ([sii]), the pitch rises considerably (10.4 semitones). Thus, the second word, siinä, carries the pitch reset in this case rather than the first word, niin. During the first syllable ([näyt]) of the next word, the pitch still rises 1.4 semitones, and the stress is more strongly expressed than during the production of the word siinä.

Figure 1 – SILP at the beginning of the second spoken sentence (Example 1)
Figure 2 – SILP at the beginning of the third spoken sentence (Example 1)

Figure 2 shows the pitch curve during the production of the last minor intonation unit of the second spoken sentence, samaan aikaan (“at the same time”), and the two first minor intonation units (niin sitä, “so there”, and näyttäis olevan, “seems to be”) of the third spoken sentence.

As in the case of the first occurrence, SILP indicates here, on one hand, the beginning of a new, grammatically independent spoken sentence. On the other hand, there is a need to indicate the close anaphoric relationship between the two spoken sentences, as the pronoun siinä (“in it” → “there”) refers to what has been said in the second spoken sentence of the speech paragraph. As the third spoken sentence could not be understood without the second, there is a relationship of asymmetric dependency between the two spoken sentences.

The speech-to-text interpreter who transformed this talk into written text omitted the word se (“it”) in the first occurrence of the SILP (line 04) but maintained the word ja (“and”). (S)he did not put a punctuation mark between the spoken sentence starting with the SILP and the previous spoken sentence. In the case of the second occurrence of the SILP (line 06), a comma is added before the beginning of the spoken sentence starting with the SILP. The word niin (“so”) is replaced by the word joten (“thus”), which emphasizes the causal relationship between these two spoken sentences.

The second example includes two occurrences of SILP. It is also taken from the talk of female informant 1. The speaker is describing sudden changes in the development of autistic children.

Example 2:\n
01 Toi yks (1.0) amerikkalainen autismiturkija

That one American autism researcher calls

---

19. When a word or a part of a word is produced more quietly than the surrounding speech, it is marked with °°-signs. A question mark preceded by a comma (,) stands for slightly rising pitch.
Indicating Dependency between Spoken Sentences by Prosodic Means

The example includes four spoken sentences (toit yksi amerikkalainen autismi- tutki joka räjähtää sitä yhtäkkiä / kaikki tiedot mitä on opittu siihen mennessä niin kun taantuu ja poistuu kentä / mut se kehittyy aivojen kehitykseen / et siinä kohta tulee tietty uus kehitystehtävä aivoille). All these spoken sentences belong to the same speech paragraph, and the third and the fourth spoken sentences begin with SILP.

The third spoken sentence, mut se kehittyy aivojen kehitykseen (“but it develops to the development of the brain”, lines 06-07), lacks coherence. The speaker probably means to say “but it is related to the development of the brain” instead of “but it develops to the development of the brain”. This is, however, irrelevant here, as the meaning of the wrong word can be easily deduced from the context.

Nevertheless, what is interesting is that the discourse particle mut starts 5.2 semitones lower than the last syllable ([ties]) of the preceding word. Thus, there is no pitch reset at the beginning of this spoken sentence. The part of speech to which the first word belongs does not seem to be a relevant factor.

---

20 The fact that the speech-to-text interpreter translates the word kehittyy (“develops”) as liityy (“is related to”) in the written target text is a good proof of this.
Indeed, at least in these data, particles and other function words also typically carry a raised pitch when they begin a spoken sentence. Here, however, it is the third syllable of the spoken sentence ([ke]) that carries a raised pitch: this syllable is produced 4.1 semitones higher than the first syllable ([mut]) of the spoken sentence (cf. Figure 3).

*Mut* (“but”, shortened from *mutta*), carrying the low pitch, acts here as an adversative-concessive discourse particle, which indicates a change of point of view (Hakulinen et al., 2004: § 1034, § 1103). The following word, the demonstrative pronoun *se* (“it”), refers – as already mentioned – typically to something that belongs to the “attention field” of the listener (Hakulinen et al., 2004: § 720). In this context, *se* refers to the compound *autismipommi* (“autism bomb”) which has been mentioned earlier (line 02). The SILP phenomenon emphasizes here the above-mentioned pragmatic functions of the words *mut* and *se*, and contributes to indicate that their semantic value is low. Thus, in this context as well, prosodic features seem to cooperate with the pragmatic and semantic levels of speech.

The SILP phenomenon indicates here, on the one hand, the beginning of a new, grammatically independent spoken sentence. On the other hand, this spoken sentence could not be understood without the preceding spoken sentences, due to the anaphoric relationship between the pronoun *se* (“it”) and the compound *autismipommi* (“autism bomb”) occurring in the first spoken sentence of the extract.

In the second occurrence, the spoken sentence starts with the particle *et* (“so”). The spoken sentence begins as much as 20.3 semitones lower than the end of the preceding spoken sentence. The second word, *siinä* (“at that point”), begins, however, as high as the end of the preceding spoken sentence (cf. Figure 4).

---

![Figure 3 – SILP at the beginning of the third spoken sentence (Example 2)](image-url)

---

21. The f0 curve has been corrected manually due to the background noise.

URL: http://journals.openedition.org/discours/9675
The word *et* (“that”, shortened from *että*) acts here as an additive discourse particle (Hakulinen et al., 2004: § 801); it indicates that the spoken sentence that starts will add information to what has been said in the previous spoken sentence. The SILP phenomenon emphasizes this pragmatic function and contributes to indicate that the semantic value of the word carrying the phenomenon is low.

As in the first occurrence, the spoken sentence carrying SILP is grammatically independent here. It could not, however, be understood without the preceding spoken sentence, as *siinä kohtaa* (“at that point”) refers to *aivojen kehitykseen* (“to the development of the brain”), which was mentioned in the preceding spoken sentence.

The first occurrence of the SILP phenomenon is taken into account by a full stop in the written target text, which is quite rare in these data; a comma would be a more typical punctuation mark in this type of context. The words *mut* (“but”) and *se* (“it”) carrying the low pitch have been reduced, which is very typical of these data. The second spoken sentence starting with the SILP phenomenon has been completely omitted in the written output. The fact of reducing the whole spoken sentence carrying the SILP phenomenon is not exceptional in these data.

Example 3 has been drawn from the talk of male informant 1. The speaker is just about to start a PowerPoint presentation, and he is talking about his new USB flash drive.

**Example 3**

01 Sen verran olen investoinut että (1.7) poika

That much I have invested that my son

---

22. A full stop appearing in the transcriptions signifies a clearly descending pitch. A colon indicates a lengthened sound.
This example includes four spoken sentences (Sen verran olen investoinut että / poika neuvo ostamaan tommosen muistitikun / Ja hän sanoi että siinä mahtuu koko sinun elämäsi / se on kaheksangigainen tikku). They all belong to the same speech paragraph, begun by the first spoken sentence in the example. There is an occurrence of SILP at the beginning of the fourth spoken sentence, se on kaheksangigainen tikku (“it is an 8GB flash drive”, lines 04-05). The first syllable ([se]) of this spoken sentence is produced 4.8 semitones lower than the last syllable ([si]) of the preceding spoken sentence (Ja hän sanoi että siinä mahtuu koko sinun elämäsi, “And he said that your whole life will fit onto it”, line 02). The third syllable ([ka]) in turn, is produced 10.5 semitones higher than the first syllable. Figure 5 illustrates the prosodic shape of the occurrence.

The word se (“it”) refers here to the compound muistitikun (“a USB flash drive” + GEN), which has been mentioned earlier in the speech paragraph (line 02), and which is therefore supposed to be known by the listeners (Hakulinen et al., 2004: § 720). The word on (“is”) is the third-person singular of the verb “to be”. Thus, the semantic value of these elements is low; they simply indicate that the spoken sentence that starts adds information to what has been said earlier. The SILP phenomenon works here to emphasize these pragmatic and semantic roles.

Figure 5 – SILP at the beginning of the fourth spoken sentence (Example 3)
As in the preceding examples, SILP also works here to emphasize a close anaphoric relationship between spoken sentences. As already mentioned, the pronoun *se* (“it”) at the beginning of the fourth spoken sentence (line 04) refers to the compound *muistitikun* (“a USB flash drive” + GEN) appearing in the second spoken sentence of the speech paragraph (line 02). The same anaphoric pronoun occurs in the third spoken sentence (*se → siiben, “onto it”, line 03). Thus, the anaphoric relationship emphasized here takes place not only between two consecutive spoken sentences but also between three consecutive spoken sentences of a speech paragraph.

This occurrence of the SILP phenomenon is taken into account by a full stop in the printed text. As already mentioned, this is however not typical of these data.

Example 4 comes from the same male informant’s talk as Example 3. Here, the speaker is talking about the logo and the upcoming anniversary of the Finnish Federation of Hard of Hearing.

Example 4:

<table>
<thead>
<tr>
<th>Line</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Sininen väri (1.6) kertoo (0.2) liiton</td>
</tr>
<tr>
<td></td>
<td><em>The blue color symbolizes</em></td>
</tr>
<tr>
<td>02</td>
<td>pitkästä historiasta. (1.9) Meillähän</td>
</tr>
<tr>
<td></td>
<td><em>the long history of the federation. Our</em></td>
</tr>
<tr>
<td>03</td>
<td>liitto täyttää (0.9) ensi vuonna (0.7) 80</td>
</tr>
<tr>
<td></td>
<td><em>federation will be 80 years old</em></td>
</tr>
<tr>
<td></td>
<td><em>(next year)</em></td>
</tr>
<tr>
<td></td>
<td><em>(and we plan to)</em></td>
</tr>
<tr>
<td>04</td>
<td>vuotta (1.1) ↓ ↓ ja ↑ on tarkotus sitä sitten</td>
</tr>
<tr>
<td></td>
<td><em>(years old)</em></td>
</tr>
<tr>
<td></td>
<td><em>(next year)</em></td>
</tr>
<tr>
<td></td>
<td><em>(and we plan)</em></td>
</tr>
<tr>
<td></td>
<td>↓ and ↑ is</td>
</tr>
<tr>
<td></td>
<td><em>(years old)</em></td>
</tr>
<tr>
<td></td>
<td><em>(next year)</em></td>
</tr>
<tr>
<td></td>
<td><em>(and we plan)</em></td>
</tr>
<tr>
<td>05</td>
<td>juhlistaa (1.0) monellakin eri tavoin.</td>
</tr>
<tr>
<td></td>
<td><em>celebrate it in many ways</em></td>
</tr>
</tbody>
</table>

This example consists of three spoken sentences (*Sininen väri kertoo liiton pitkästä historiasta / Meillähän liitto täyttää ensi vuonna 80 vuotta / ja on tarkotus sitä sitten juhlistaa monellakin eri tavoin*), which all belong to the same speech paragraph. The third is marked with SILP. The first syllable ([ja]) of this spoken sentence is produced 6.3 semitones lower than the last syllable ([ta]) of the preceding spoken sentence. The second syllable ([on]) of the third spoken sentence is produced 8.2 semitones higher than the first syllable. The third syllable ([tar]) is produced a further 1.5 semitones higher than the second syllable ([on]) and the stress is more strongly expressed during its production than during [on]. The prosodic shape of the occurrence in question is presented in Figure 6.
The word *ja* ("and") that initiates the spoken sentence carrying SILP acts here as an additive particle (Hakulinen et al., 2004: § 804); that is, it works to indicate that the new spoken sentence adds something to the preceding spoken sentence and is therefore closely related to it (Hakulinen et al., 2004: § 804). SILP emphasizes this additive function, as well as the fact that the semantic value of the word is low. In the printed text, this occurrence of the SILP phenomenon corresponds to the presence of a full stop. The first word *ja* ("and") carrying the low pitch is reduced, which is typical of these data.

The last example is taken from the talk of male informant 2. The speaker has just begun his presentation, and here he is describing his own background.

**Example 5:**

01 Öö (0.2) ITse (0.2) olen opiskellut valaistusta,  
   *Er myself I have studied lighting*

02 (0.4) ja: (0.3) ööee Helsingissä (.)  
   *and er in Helsinki*

03 teknillisessä korkkeakoulussa ja (0.6) öö (.)  
   *at the University of Technology and er*

04 olen tehnyt (.) diplomityöni (.)  
   *I did my Master’s thesis*

05 palvelutalojen valaistuksesta, (0.4)  
   *on lighting in sheltered accommodation*

→ 06 ↓eli (.) ↑millainen (.) valaistus (.) palvelutaloissa  
   *that is to say ↑what kind of  
   *that is to say what kind of lighting*
The example includes four spoken sentences: “they should have in sheltered accommodation and at the same time we studied what kind of lighting” begins with SILP. The first syllable ([eli]) of the third spoken sentence commences as much as 13.0 semitones lower than the end ([ta]) of the preceding spoken sentence. The second syllable ([mil]), in turn, is produced with a raised pitch: the pitch level is 16.9 semitones higher during the production of the syllable [mil] than during the production of the syllable [eli]. The prosodic shape of the occurrence is given in Figure 7.

The word *eli* (“that is to say” / “in other words”) is an additive discourse particle that is used here to signal the transition to a paraphrase of what has been said before (Hakulinen et al., 2004: §1031): the speaker has just told the audience that he had written his Master’s thesis on lighting in sheltered accommodation (lines 04-05), and then he explains what that meant in practice. As the new spoken sentence elaborates on the preceding one, there is a relationship of asymmetric dependency between these spoken sentences. SILP emphasizes this relationship and the transition to the paraphrase, as well as the fact that the semantic value of this word is low; that is, it has no actual *meaning* but rather a discursive role. Thus, here too, prosodic features cooperate with pragmatic and semantic levels of speech.

![Figure 7 - SILP at the beginning of the third spoken sentence (Example 5)](image-url)
In the written target text, this occurrence of the SILP phenomenon is marked with a full stop. The first word *eli* (“thus”) is not omitted, which is rather exceptional.

4. Discussion

Prosodic means are used to group spoken sentences together into larger units, termed speech paragraphs (Wichmann, 2000). The beginning of a new speech paragraph is marked by a melodic upstep carried by the onset of the first spoken sentences (Brazil et al., 1980; Brown et al., 1980; Couper-Kuhlen, 1986; Lehiste, 1975; Wichmann, 2000; Wiklund, 2013b and 2014; Yule, 1980). Prosodic features also indicate relationships between spoken sentences belonging to the same speech paragraph. In the unmarked pitch pattern, the new spoken sentence begins at a higher pitch than the end of the preceding spoken sentence. This phenomenon is called “pitch reset” (Wichmann, 2000: 24). This article has focused on a phenomenon called “sentence-initial lowered pitch” (SILP). The phenomenon generally appears inside a speech paragraph when two consecutive spoken sentences are closely related, for example because of an anaphoric relationship. In these cases the first syllable of the new spoken sentence is produced at a lower pitch compared to the end of the previous spoken sentence. The second or subsequent syllable, in turn, carries raised pitch. The phenomenon indicates, on one hand, a relationship of asymmetric dependency between spoken sentences, but, on the other hand, the beginning of a grammatically independent spoken sentence (Wichmann, 2000: 75-78; Wiklund, 2013a and 2013b). SILP is rather frequent in these data. Indeed, 27.7% of spoken sentences in the data carry this prosodic feature. Thus, the article shows that the phenomenon can be found both in male and female informants’ speech in Finnish data consisting of conference-like, monologous presentations. This phenomenon constitutes also a good example of how prosodic features cooperate with pragmatic and semantic levels of speech.

As already mentioned (cf. Section 2), the data of this study come from speech–to–text interpreting situations, where speech is simultaneously transformed into written format and displayed on a screen so that it can be accessed by the deaf and hard-of-hearing (Sharmin et al., 2016; Borg, 2008; Mäkiranta, 2006; Öhrankämmen, 2008; Salakari, 2008; Tanhuamäki, 2008; Tiittula, 2006 and 2009). In the printed text (that is to say, in the printed output produced by the speech–to–text interpreters), the phenomenon of SILP most often entails (80.4%) the reduction of the first unaccented word of the spoken sentence (Wiklund, 2013a and 2013b). It usually occurs (79.3%) in the middle of a paragraph of written text. As the phenomenon also almost always occurs in the middle of a speech paragraph and not at the beginning, this does not seem surprising. The spoken sentences beginning with SILP are only preceded by a full stop in the printed text in 15.2% of cases. This also seems natural, because the phenomenon typically indicates dependency between spoken

---

23. The “topic reset” phenomenon – which starts a speech paragraph – usually also entails (91.7%) a paragraph division in the written target text (Wiklund, 2013b and 2014).
sentences. Most often a spoken sentence beginning with SILP is linked to the preceding spoken sentence either with a comma (32.6%) or without a punctuation mark (65.2%) (Wiklund, 2013a and 2013b).

In the future, it would be interesting to study SILP in a larger data set in order to see if the phenomenon occurs in Finnish also in different types of speaking situations. In addition to monologous, conference-like presentations, spontaneous speech and dialogues should be studied from this point of view. Statistical analyses would also be interesting to carry out. After these studies, a systematic model explaining the phenomenon in Finnish could be constructed. A systematic model, based on a variety of speaking situations, would also allow comparing the possible occurrences of the SILP phenomenon in different languages.

References


Indicating Dependency between Spoken Sentences by Prosodic Means


