

# Speech Chunks in Conversation: Syntactic and Prosodic Aspects

Eva Strangert

Department of Philosophy and Linguistics, Phonetics

Umeå University

eva.strangert@ling.umu.se

## Abstract

The paper reports on syntactic and prosodic analyses of speech chunks – the sequences of speech between perceived boundaries – in conversation. Comparisons with read-aloud speech are also included. The data analysis shows that of the total number of chunks in the conversation, almost 80% had endings coinciding with a syntactic boundary, while about 20% violated syntactic continuity by the occurrence of a boundary occurring in syntactically unmotivated positions. Suspension mainly occurred after initial function words close to the beginning of the constituent in accordance with the hypothesis of initial commitment. The results are discussed in terms of the commit-and-restore model developed by Clark & Wasow (1998), offering a linguistic-cognitive approach to speech processing.

## 1. Introduction

There is a wealth of research on boundaries and pausing in spontaneous and read speech, but less so on the sequences of speech between boundaries. These sequences, here referred to as *chunks*, are the focus of this study undertaken within the framework of the project “Boundaries and groupings – the structuring of speech in different communicative situations” [1].

“Chunks” is chosen as a neutral term to reflect the status of these units as the result of combined linguistic-prosodic and cognitive processing. That is, chunks have lots in common with, but are not identical to, either syntactic, semantic or prosodic constituents (Cf. units of performance structure as assumed by [6]). Chunking, moreover, is shaped by restrictions on planning and execution. For example, the demands on the speaker in on-line planning sometimes lead to breakdowns resulting in restarts, repetitions and pausing. Also, situational demands – adjustments to the listener – are a factor in the shaping of speech.

The interactions between the linguistic-prosodic and cognitive processing in spontaneous speech have been dealt with by Clark and colleagues in a number of studies, for example [3], [5]. These studies all lean on the concept of *the ideal delivery*, that is, the way an utterance would be produced, were there no problems [4]. One hypothesis in the *commit-and-restore model* [5] reflecting the notion of the ideal delivery is that “speakers prefer to produce constituents with a continuous delivery” [5]. (“Constituents” in this model primarily refer to noun, verb and prepositional phrases as well as to clauses and sentences.) That is, they aim at producing entire constituents without interrupting themselves. Another hypothesis accounts for the cases where continuity is violated, which happens when speakers suspend speech within a constituent (as a result of planning problems such as lexical search problems etc). According to this model, speakers make an initial commitment to what will follow, that is, they initiate

the constituent before having decided on all of it. In doing so, they give clues to the listener about what kind of syntactic form the following message will have. This syntactic signaling occurs combined with pauses – silent or filled – as well as lengthened durations of the initial word(s). By such commitments the speaker signals that he/she is going to continue speaking.

Clark & Wasow [5] found support for their hypotheses when tested on two large databases of English spontaneous speech. Also, some recent work on hesitations involving clause-initial conjunctions in spontaneous Swedish, demonstrated prosodic as well as segmental adjustments conforming to these hypotheses [8].

In the present study, the same database of Swedish conversational speech was used as in a previous effort on boundary characteristics [7] undertaken within the project [1]. After describing the general characteristics of the material, a characterization of the chunks will be given in terms of the features included in the database using the commit-and-restore model as a theoretical framework. For comparisons with read speech, some comparable data on chunking were extracted from a database primarily used for studies of pausing in read speech [10].

## 2. The database

The spontaneous speech material has the form of a radio interview, about 25 minutes long, with a female politician answering questions from two interviewers, one male and one female. The interview consists of approximately 4100 words (including hesitations, disfluencies etc) of which about 80 % were produced by the interviewee and 20% by the two interviewers. The material includes passages of interactive dialogue but longer stretches of monologue-like speech as well.

The complete material was annotated for perceived boundaries by three experienced transcribers marking each word as being followed by a strong, a weak or no boundary. Applying a majority-vote-decision procedure the result was 211 strong, 407 weak and 3459 no boundaries and 25 cases of total disagreement.

The material was automatically segmented using an orthographically transcribed speech file as input, and temporal data were extracted to capture boundary and pre-boundary characteristics. These data include, among other measurements, word durations and durations of word final rhymes as well as silent interval durations at boundary positions. The durations are given as absolute values and also, to be able to compare different words, as calculated average z-score normalized durations. The normalization procedure involved computing the difference between the observed duration for each segment and the mean duration of all segments of the same type expressed as standard deviation units. Thus, the normalized durations of words, for example,

express the summation of differences from the mean of all included segments. (See, for example [11] for methodological considerations concerning normalized durations.)

The database, moreover, includes linguistic descriptions of the transcribed conversation. The linguistic features used to classify the words are: content-function word, part of speech and phrase structure.

For a more detailed description of the database, see [7].

### 3. Analyses

#### 3.1. Number of words/chunk

As a starting point, the number of words/chunk was extracted. This is shown in Figure 1 containing the distribution for the entire conversation (618 chunks) with chunks ending with perceived strong (//) and weak boundaries (/) given separately.

The skewed distribution with a preponderance for chunks with just a few words should be attributed to the spontaneous character of the speech being analyzed. The same kind of analysis of read speech (10 speakers, each reading a text of 810 words) in a material used to analyze pausing [10] presents a different picture. In the read speech the distribution has its maximum at chunks with 7 words and also a preponderance for chunks in the range of 3-9 words. This contrasts with the maximum at chunks with 3 words in the conversation as well as the frequency of occurrence of chunks 1-2 words long. These very short chunks are particularly interesting as they set the conversational speech aside from the read. The shorter chunks in the spontaneous conversation and the longer in the read should be looked upon in a continuity vs. violations-of-continuity perspective [5]. Therefore, the next section will be devoted to a broad analysis of the syntactic form of words at chunk endings.

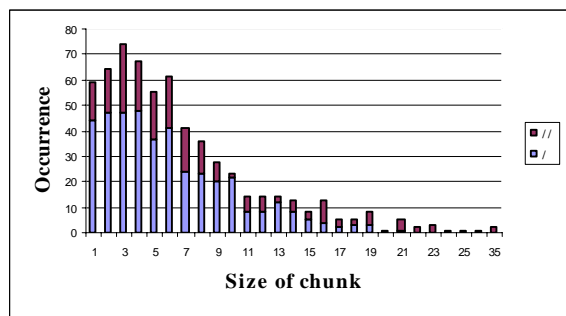


Figure 1: The distribution of size of chunk (number of words/chunk) separated for chunks ending with a strong (//) and weak boundary (/). Total number of chunks 418.

#### 3.2. Syntax at chunk endings

Perceived boundary positions were matched with parts-of-speech and phrase category markings to find out to which extent the boundaries occurred in “syntactically motivated” positions, that is, whether they had a continuous delivery or not (see discussion in introduction). Here syntactically motivated (= continuous delivery) means (a) occurrence

between, rather than, within, phrases and (b) before clause initial words rather than after.

The data analysis shows that of the total of 618 chunks, almost 80% had endings coinciding with a syntactic boundary, while 117, that is, slightly more than 20%, violated continuity by the occurrence of a boundary occurring in syntactically unmotivated positions. These positions and the frequency of occurrence of each appear from Table 1.

Table 1: Positions of perceived boundaries in relation to syntactic structuring.

	Grammatical category	Occurrence
<b>Within</b>	Prepositional phrase	27
	Noun phrase	27
	Verb cluster	9
<b>After</b>	Subjunction	14
	Conjunction	11
	Infinitive mark	9
	Pronoun	8
	Adverbs	7
	Other	5

Clearly, in most cases the speakers manage to deliver their utterances without suspending speech within syntactic constituents. Also, when suspension occurs, it happens close to the beginning of the constituent in accordance with the hypothesis of initial commitment (already discussed in the introduction). Suspension mainly occurs after initial function words, the most frequent being prepositions, clause-initial conjunctions and subjunctions. Also, almost all cases of violation occurred at boundaries judged as weak, 112 out of the 117 cases.

Comparing these observations with read-aloud data based on the material in [10] the differences are striking. In the read text comprising for all ten speakers together 8100 words, only 28 cases of perceived boundaries at syntactically unmotivated positions occurred, when employing the same criteria for “syntactically unmotivated” as in the present study.

More detailed conversation data are shown in Table 2.

Table 2: Number of chunks violating continuity in % of total number for each size of chunk. Number of violations involving a function word also given.

Size of chunk	Occurrence, violation	Occurrence, funct word	Chunks, total	Occurrence violation in % of total
1	23	18	59	39
2	18	14	64	28
3	15	11	74	20
4	11	8	67	16
5	10	6	55	18
6	8	5	61	13
7	5	4	41	12
8	3	3	36	8
9	7	7	28	25
10	5	5	23	22
>10	12	11	110	11

Treating each size of chunk separately, the table shows that almost half of the chunks are very short, consisting of either one, two or three words, and of the 59 one-word chunks, 23 fall into the “ungrammatical” (violation) category. The high incidence of function words involved in violation is clearly demonstrated.

### 3.3. Prosody at chunk endings

Next, a prosodic-acoustic analysis of the differently-sized chunks was undertaken. The prosodic data were also matched with the results of the analysis of linguistic structuring in 3.2.

Table 3 and 4 both show mean word and word-final-rhyme durations at the end of chunks as well as following silent interval durations. All figures are given as average z-score normalized durations. (See 2. and [7] for further details of the calculations).

The generally positive z-score values indicate longer than average durations, that is, lengthening, at chunk endings. Though reflected in the total duration of words, the lengthen-

Table 3: Mean duration of words and word final rhymes before perceived boundaries and silent intervals for chunks of different size. Data (z-score normalized durations) given separately for weak (/) and strong (//) boundaries.

	Chunk size	Mean word dur	Mean word fin rhyme	Mean silence after	Occurrence, total
<b>Before /</b>	1	0,83	1,76	0,20	44
	2	0,51	0,95	0,15	47
	3	0,53	1,02	0,16	46
	4	0,70	1,18	0,20	48
	5	0,43	0,77	0,19	37
	6	0,61	1,14	0,21	41
	7	0,46	1,03	0,21	24
	8	0,34	0,69	0,21	24
	9	0,38	0,56	0,17	20
	10	0,25	0,64	0,24	22
	11	0,30	0,64	0,20	8
	12	0,21	0,73	0,27	8
	13	0,03	0,43	0,15	12
	14	0,21	0,16	0,13	9
	15	0,09	0,34	0,15	4
	>15	0,06	0,48	0,17	14
<b>Before //</b>	1	0,21	0,67	0,37	16
	2	0,25	0,59	0,39	17
	3	0,37	0,99	0,37	27
	4	0,48	0,71	0,35	20
	5	0,03	0,47	0,34	18
	6	0,11	0,36	0,45	21
	7	0,14	0,35	0,34	18
	8	0,13	0,31	0,34	11
	9	0,06	0,15	0,44	8
	10	-0,51	-0,73	0,03	1
	11	0,27	0,38	0,32	7
	12	0,19	0,61	0,41	6
	13	-0,11	0,37	0,35	3
	14	0,13	0,11	0,29	5
	15	-0,51	-0,34	0,20	3
	>15	0,02	0,22	0,34	30

ing has its primary source in the final part of the word, as shown by the word-final-rhyme durations.

Table 3 shows the durations at the end of the differently-sized chunks. Chunks before weak and strong boundaries, respectively, are presented at the top and bottom of the table.

For chunks ending before a weak boundary, there is a tendency of decreasing word and word final rhyme duration when the size of the chunk increases. One-word chunks, in particular, stand out as having considerably longer durations than the final words in chunks containing more words. For chunks before a strong boundary, there is a similar tendency, although much weaker, and the one-word chunks do not have similar excessively long durations. In addition, the durations of words and word-final rhymes are generally longer before weak boundaries than before strong. Thus, size of chunk as well as type of the following boundary affect the temporal structuring before the boundary. Silent intervals, on the other hand, appear to be unaffected by the size of chunks. Yet they differ consistently between strong and weak boundaries, being about half as long at weak as compared to strong boundaries. (See also [7].)

Next, the extent to which the syntactic structuring affected prosody was investigated. To that end, the durations for cases violating continuity were compared with cases of non-violation. Table 4 gives the respective figures for mean word, word final rhyme and silent interval durations across all sizes of chunks with chunks violating and those not violating continuity separated.

Table 4: Mean duration of words and word final rhymes before perceived boundaries and silent intervals given separately for chunks with violation and non-violation of continuity. Data given as z-score normalized durations.

	Mean word dur	Mean word fin rhyme	Mean silence after	Occurrence
<b>Violation</b>	.83	1.26	.24	117
<b>Non-violation</b>	.28	.69	.25	501

The speakers quite obviously behave differently in the cases where chunk endings coincide with a syntactic boundary and when they do not. Word durations, and word-final-rhyme durations in particular, are longer in cases of violation. Silent intervals, on the other hand, are very similar irrespective of the syntactic form at chunk endings.

## 4. Discussion

The preponderance of chunks consisting of just a few words is characteristic for the material analyzed. In particular the frequency of very short chunks (1-2 words) appears to be a typical feature, setting conversational speech aside from read speech. This difference without doubt should be ascribed to the heavier demands on on-line planning in conversational speech as compared to read.

Yet most of the chunks have the ideal continuous delivery assumed to be what speakers generally aim for [5]. However, when violations of continuity occur (in approximately 20% of the total number of chunks) they do not appear haphazardly, but rather in accordance with the predictions in [5].

Most violations occur in chunks of 1-4 words, with an overrepresentation of one-word chunks, and more than a third of all one-word chunks violate continuity. The violations, in addition, involve suspensions of speech in accordance with the hypothesis of initial commitment, that is, suspensions primarily occur after the initial part of what is to be produced. In the material analyzed, suspensions primarily occurred after initial subordinations and conjunctions and in the initial part of phrases, primarily prepositional phrases and noun phrases.

The frequent involvement of prepositional phrases is worth notice. Analyses of pausing in read speech [10] revealed that most often phrases were not separated by pauses. But, when that happened, prepositional phrases were among the most frequent. The assumption was that the underlying factor was a high semantic load in this kind of phrases, often giving important information about dates, places etc. And pausing, by delimiting this information from the rest, would be the means to signal this importance. Given that this is the case, the high incidence of suspensions in prepositional phrases in conversation would not be surprising.

The analysis in [7] showed weak boundaries to be characteristically different from strong boundaries in that they had shorter silent intervals but at the same time longer word-final rhymes. (This can be interpreted in terms of a trading relationship between the final lengthening and the following silent interval, cf. [9] reporting previous work on Swedish.) This same pattern is evident across the different sizes of chunks. However, while silent intervals do not vary across the different sizes of chunks – although being consistently longer at strong as compared to weak boundaries – the lengthening of (final parts of) words are strongly affected by the size of chunk. There is a general trend, in particular at boundaries judged as weak, of increasing lengthening the less words in the chunk. Accordingly the one-word chunks again stand out from the rest, in this case by having the longest durations (most lengthening). Before weak boundaries, the one-word chunks even have extremely long durations.

Cases of violation almost exclusively involved boundaries judged as weak, that is, boundaries with relatively short silent intervals but considerable final lengthening. Violations of continuity, moreover, predominated in chunks with just a few words, the chunks characterized by the most extreme lengthening. Thus planning problems resulting in suspensions of speech within, rather than between, constituents appeared to be characteristically signaled to the listener through excessively long durations of segments before the suspension. Similar observations were made by [8] studying the two conjunctions *att* 'that' and *och* 'and' in Swedish in positions of hesitation. In addition to longer vowels, they found significantly longer final aspiration phases before hesitations than in fluent speech.

## 5. Conclusions

These results [8], similarly as the ones presented in this paper, support the assumptions of the commit-and-restore model [5] of specific strategies to handle planning problems in speech production. These strategies are both structural – suspensions revealing the syntactic form of the message to be produced – and prosodic – slowing down on the final word before the suspension.

In the commit-and-restore model, the perspective is twofold; it takes both the speaker and the listener into account. Thus, the strategies used by the speaker are assumed

to be favorable to the listener. Investigating this lies outside the scope of the present study. Recent work [2] within the “Boundaries and groupings” project [1] however demonstrates that listeners can decide on the type of boundary on the basis of information before the boundary itself.

## 6. Acknowledgement

This work was carried out within the GROG project “Boundaries and groupings – the structuring of speech in different communicative situations” supported by The Swedish Research Council (VR).

## 7. References

- [1] Carlson, R., Granström, B., Heldner, M., House, D., Megyesi, B., Strangert, E., Swerts, M., 2002. Boundaries and groupings – the structuring of speech in different communicative situations: A description of the GROG project. *Proceedings of Fonetik 2002, TMH-QPSR* 44, 65-68.
- [2] Carlson, R., Swerts, M., 2003. Perceptually based prediction of upcoming prosodic breaks in spontaneous Swedish speech materials. *Proceedings 15<sup>th</sup> International Congress of Phonetic Sciences, Barcelona*, 507-510.
- [3] Clark, H.H., 2002. Speaking in time. *Speech Communication* 36, 5-13.
- [4] Clark, H.H., Clark, E.V., 1977. *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- [5] Clark, H.H., Wasow, T., 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37, 201-242.
- [6] Gee, J.P. & Grosjean, F., 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15, 411-458.
- [7] Heldner, M., Megyesi, B., 2003. Exploring the prosody-syntax interface in conversations. *Proceedings 15<sup>th</sup> International Congress of Phonetic Sciences, Barcelona*, 2501-2504.
- [8] Horne, M., Frid, J., Lastow, B., Bruce, G., Svensson, A., 2003. Hesitation disfluencies in Swedish: prosodic and segmental correlates. *Proceedings 15<sup>th</sup> International Congress of Phonetic Sciences, Barcelona*, 2429-2432.
- [9] Horne, M., Strangert, E., Heldner, M., 1995. Prosodic boundary strength in Swedish: Final lengthening and silent interval duration. *Proceedings 13<sup>th</sup> International Congress of Phonetic Sciences, Stockholm*, 170-173.
- [10] Strangert, E., 1991. Pausing in texts read aloud. *Proceedings XIIth International Congress of Phonetic Sciences, Aix-en Provence*.
- [11] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 9 (3), 1707-1717.