

Edited by:  
Jan Heegård  
Peter Juel Henriksen

# New Perspectives on Speech in Action

## Proceedings of the 2nd SJUSK Conference on Contemporary Speech Habits

43

Copenhagen Studies in Language

# **New Perspectives on Speech in Action**



# **New Perspectives on Speech in Action**

## **Proceedings of the 2nd SJUSK Conference on Contemporary Speech Habits**

Edited by:

Jan Heegård & Peter Juel Henriksen

**Copenhagen Studies in Language 43**

**Samfundslitteratur Press**

*New Perspectives on Speech in Action*  
*Proceedings of the 2nd SJUSK Conference on Contemporary Speech Habits*

**Copenhagen Studies in Language 43**

© Samfundslitteratur and the authors 2013

ISSN 0905-09857  
ISBN 978-87-593-1957-4

Cover by: Klahr Grafik  
Layout by the editors

Copenhagen Business School  
Handelshøjskolen i København

Published by:  
Samfundslitteratur Press  
Rosenørns Alle 9  
DK-1970 Frederiksberg C

Phone: +45 3815 3880

slforlagene@samfundslitteratur.dk

samfundslitteratur.dk  
csl.samfundslitteratur.dk

Editors: Jan Heegård & Peter Juel Henriksen

## Contents

<b>List of contents</b> .....	v
<b>Foreword</b> .....	vii
<b>Jens Allwood and Elisabeth Ahlsén:</b> On language production as expansion, reduction or multiple representations.....	1
<b>Lars Brink:</b> Phonetic reduction.....	13
<b>Andrea Deme:</b> Pragmatic functions of lengthenings and filled pauses in the adult-directed speech of Hungarian children.....	23
<b>Slobodanka Dimova and Christian Jensen:</b> Reduction in language testing.....	41
<b>Jan Heegård:</b> Functional phonetic reduction? The case of the Danish adverb <i>faktisk</i> ('actually').....	59
<b>Peter Juel Henriksen:</b> Sloppiness and distinctness in the realm of body language.....	79
<b>Eva Skafte Jensen:</b> Stress distribution in phrases with predeterminant <i>sådan</i> in Danish.....	89
<b>Tilda Neuberger and András Beke:</b> Automatic laughter detection in Hungarian spontaneous speech using GMM/ANN hybrid method.....	113
<b>Yulia Nigmatulina:</b> Sound contraction in Russian spontaneous speech and its implications for spoken word recognition.....	127
<b>Olga Raeva:</b> The role of grammatical and semantic context in spoken word recognition.....	141
<b>Elena Riekhakaynen:</b> Reduction in spontaneous speech: How to survive.....	153
<b>Ruben Schachtenhaufen:</b> Phonetic reductions and linguistic factors.....	167
<b>Marina Snesareva:</b> Pausing in Modern Irish: spontaneity and intention.....	181
<b>Jakob Steensig et al.:</b> The DanTIN project – creating a platform for describing the grammar of Danish talk-in-interaction.....	195
<b>Jacob Thøgersen and Martin Kristiansen:</b> Between 'maintenance' and 'development'. Popular perceptions of the spoken language of the Danish National Broadcasting Corporation.....	227
<b>Richard Todd:</b> Ethnic speaker identification with limited data: examining acoustic correlates of grouping outcomes.....	243

## Foreword

We are proud to present these New Perspectives on Speech in Action from the SJUSK 2013 conference, held at Copenhagen Business School in March 2013 ([www.cphspeech2013.dk](http://www.cphspeech2013.dk)). With the SJUSK 2013 conference we again put natural speech habits on the agenda, continuing the research interest that arose from the SJUSK 2011 conference. However, the 2013 conference included not only research contributions on phonological theory and language technology but also perspectives from other disciplines that deal with spoken language phenomena: Hearing science, second language acquisition, language attitudes, modern speech technology and conversation analysis.

This publication presents a selection of the papers and posters that were presented at the conference. It describes spoken language phenomena from many different languages, including Modern Irish, Russian, Hungarian, Latin American Spanish, European Spanish, Portuguese, Danish, Swedish, Dutch, and British English. It is our hope that the papers may contribute to the reader's understanding of the challenging diversity of speech in action, and we shall look forward to suggestions as to workshops on more new perspectives of spoken language at the SJUSK 2015 conference.

We would like to thank Fulvio Benedetto Rizzollo for his enthusiastic and competent work in the formatting of the articles. And we would also like to thank the peer-reviewers for their careful reading and suggestions as to improvements of the contributions.

Commemorative words: In November 2013 the Danish linguist Jan Katlev suddenly passed away. Jan Katlev was the first Danish linguist to do research in allegro phonology and it was only natural that he participated in the SJUSK 2011 and SJUSK 2013 conferences, as well as contributed to the proceedings of the SJUSK 2011 conference. Jan was a flamboyant linguist, a rare type in the Nordic scene. The world has become sadly quieter without him.

Jan Heegård  
DanCAST  
Copenhagen Business School

Peter Juel Henriksen  
DanCAST  
Copenhagen Business School

# On language production as expansion, reduction or multiple representations

Jens Allwood

[SCCIIL Interdisciplinary Center](#)

[University of Gothenburg](#)

[jens@ling.gu.se](mailto:jens@ling.gu.se)

Elisabeth Ahlsén

[SCCIIL Interdisciplinary Center](#)

[University of Gothenburg](#)

[elisabeth.ahlsen@gu.se](mailto:elisabeth.ahlsen@gu.se)

## Abstract

*There are three main options in reflecting on the fact that written language word forms often are longer than corresponding spoken language word forms. The first is to use a "full form" or "written language" standard and describe shorter spoken language forms as reductions. The second is to see early acquired spoken language forms as the basis from which words are later expanded to more "written language" like forms. A third option is to consider spoken and longer language forms as parallel forms. Expansion seems more likely than reduction, in language acquisition. A result could be multiple alternative possibilities of pronunciation, i.e., reinforced by written language, developing into the third option. In line with this, some phonological or prosodic distinctions should be described more in terms of distribution of occurrences of one or the other form of a particular word (taken from an "expression potential"), the occurrence of which would be influenced by factors like social activity type or social group characteristics.*

**Keywords:** *Spoken versus written language; reduction in language production; expansion in language production; word reduction in speech; multiple word forms; written language word forms.*

## 1. Introduction

### 1.1 Reduction, expansion or multiple forms

Written language word forms often are longer than spoken language word forms, see, for example, the Swedish expressions "jag kan inte" -> *jakatte* ('I can not' -> 'I can't'). In this paper, we explore some of the ways in which we can understand this phenomenon. We will be considering three main alternative perspectives on the relation between longer (often written language like) and shorter (often spoken language like) forms of expression.

1. Reduction
2. Expansion
3. Multiple forms

There are many spoken language conventionalized phrases that are often referred to as "contractions" or "reductions". Compare the three phrases in example 1, which exemplify a pattern of negation found with auxiliaries in Swedish. (Similar examples can be found in many other languages.)

Example 1.

- A. *jag kan inte* ('I can not')
- B. *ja kan inte*
- C. *ja kante*
- D. *ja kant*
- E. *jakatte*
- F. *jakat*
- G. *jakte*

It is common to regard the forms 1B-1G as reductions of 1A, reflected in written forms like *ka'nte*. But this is hardly reasonable considering the 1B and 1C above (i.e., expansion and multiple forms). We would like to question the assumption that the form *jag kan inte* or *ja kan inte* is the unmarked case both in speech and writing. This form is used when speech, for some reason, is very formal, linked to writing or especially emphasized. In most spoken language activities, it is probable that forms 1C and 1D are more frequent and that form 1A should rather be seen as an expansion of the spoken language form.

From example 1, it could be reasonable to assume that it is the auxiliary that is inflected for negation with a suffix like *-te*. If, however, we look at example 2, where the word order is reversed, this seems less likely, since we would then have to say that it was a discontinuous suffix, allowing for pronominal insertion or that it could possibly also be used to inflect pronouns negatively (which would be a rare alternative from a typological perspective).

Example 2.

- A. *det kan jag inte* ('that I can not')
- B. *de kan ja inte*
- C. *de kan jante*
- D. *de kanjate*
- E. *de kajate*
- F. *de kate*

Considering both example 1 and example 2, it seems better to regard *-te* as a negation particle which can be used as a suffix, but also as a free standing morpheme (*inte*), depending on context and writing conventions.

If forms like, for example, 1E, 2D and 2E increase in frequency generally in a language, it would be reasonable to see this as a language change in which alternative pronunciations are becoming more frequent, rather than as reductions of forms like 1A and 2A into the other forms. We could then seek the reasons for such a change in different factors that could affect the choice of form, like types of communication, media of communication, frequency of communicative activity types, status and fashion of different speech modes etc. In the following, we will explore the presence of long and short forms some more. We therefore turn first to the three alternative perspectives on the occurrence of short and long word forms.

### **Option 1. Reduction.**

The first option, *reduction*, is to use a 'full form' or 'written language standard' form as starting point (presumably also in our mind and brain) and describe (possibly using reduction rules) shorter spoken language forms as reductions (cf., for example, Schachtenhaufen 2010). In this case, one would say that *jakatte* is a contraction or reduction of *jag kan inte*.

### Option 2. Expansion.

The second option, *expansion*, is to see early acquired spoken language forms as the basis from which words are later expanded (or possibly reduced in some cases) to more "written language" like forms, for example, due to experience of listening to text being read or learning to read and write. The basic form might not necessarily be the shortest, but probably a very salient and often heard form. Other longer (and shorter) spoken language forms can then be seen as expansions (or reductions) of a basic form, possibly using expansion (and reduction) rules.

In this case, perhaps *jakatte* (rather than *jakte*, which is less salient and frequent for small children) could be considered as the basis. Later, when learning to read and write, the child learns the expanded form *jag kan inte*, which is regarded as the correct form in written language. This implies also learning to separate the morphemes/words and to identify the expanded forms with the earlier acquired forms, which can take place before or during learning written language forms. Since writing has high status, the children are taught that the written forms are really the correct forms of the word and that spoken words are just sloppy versions for this ("sjusk", the Danish word for sloven language). In spite of this ideology, the shorter spoken language forms are usually retained as correct in spoken language. In this way, children are trained in a double standard, i.e. to think of written language forms as correct, while implicitly retaining and using the spoken language forms as correct in spoken language.

### Option 3. Multiple forms.

A third option is to consider spoken language and longer written language – like forms as *multiple potential forms* of the "same" word or phrase, some of them with shorter and simpler pronunciations, others with more elaborated and written language like pronunciations. Actualized forms are then described, long as well as short, with the help of memorized collocations and rules, which could be described as rules of contextual selection, rules of contextual determination or rules of adaptation. All potential forms come with related affordances, which favor the word forms in particular contexts.

In this third case, like in the second, *katte*, *kante* and other spoken forms are assumed to be learned early by children, while *kan inte* is learned

when learning to read and write, as another variant, among the potential realizations of a morpheme, word or phrase.

Processes of expansion and reduction and/or shifts between different forms take place all the time both as contextual adaptation on an individual level and as development of individual language or on a collective level as language change, for example, with general trends for drifting in some direction, for example toward shorter word or longer forms.

The question is now which of the three accounts is most realistic, simple and consistent with other phenomena in language production.

### ***1.2 From parallel or multiple forms to expression potentials***

We believe the arguments presented above favor a view of linguistic units as connected with multiple realizations. Thus, morphemes, words and phrases can, from a production point of view, be regarded as being connected with an expression (or pronunciation) potential, which, depending on context, actualizes different forms. Linguistic units (like morphemes, words or phrases) are, in this way, seen as more abstract than the actual forms realizing them and written and spoken language are seen as special cases of context.

### ***1.3 Selecting particular forms of expression***

#### ***1.3.1 Expansion and reduction***

Processes of expansion and reduction and/or choice of different forms take place all the time (i) as contextual determination/adaptation in language production, (ii) as ontogenetic development of an individual's language, increasing the expression potentials of linguistic units, and (iii) as language change, for example, with a trend drifting in some direction, like toward shorter word forms.

#### ***1.3.2 Contextual actualization***

All of the alternative approaches described above (reduction, expansion and multiple forms), together with an account of how the distribution of forms is influenced by factors like social activity type of social group characteristics, can provide a fruitful description of the variation of spoken forms.

In the example above, we might use the form *jag kan inte* ('I can not') in speech in a number of specific contexts.

If shorter forms are favored or not, depends on the context and affordances, i.e. what is needed, what is more efficient, what requires more or less effort etc. We will now consider some of the contextual factors influencing which alternative forms of expression are used.

## **2. Some examples of multiple realizations of linguistic units in spoken language**

### ***2.1 Reading aloud a written text***

When reading a written text aloud, depending on the purpose of the activity, the pronunciation can be more or less written-like and in this context, this is often not seen as unnatural. However, the purpose of the reading activity will also play a role, so that reading a story aloud to children might include shorter speech forms, while reading an official document in order to check the content, might include more "written-like" variants.

### ***2.2 Speaking with emphasis and using contrastive stress***

A person speaking with emphasis, for example in a 'read my lips' context or when using contrastive stress often slows down his/her speech, which leaves room for pronunciation of more sounds, possibly increases the intensity, and also deliberately articulates words in a 'hypercorrect' way, thus, ending up with a 'written-like' form.

Contrastive stress on one or more words has the same effects locally and this effect usually also spreads to the adjacent parts of the utterance, giving them an increase in stress and duration, as a kind of 'assimilation', i.e. if emphasizing 'jag' in our example, it is more likely that the versions 'JAG kan inte' or 'JAG kante' occurs, than, for example, 'JAG katte'. If contrastive stress is put on 'kan', the result would most likely be something like 'ja KAN inte', rather than 'ja KANte' and 'jakatte' becomes impossible. Contrastive stress on 'inte' also prevents spoken forms like 'jakante' and 'jakatte' and necessitates a form like 'ja kan INTE'.

### ***2.3 Speaking in a thoughtful and/or hesitant way***

Thoughtfulness and hesitation are factors that mainly slow speech down, they can perhaps also reduce intensity. The increased duration would here

be the main reason for a longer, more written-like speech form to occur, like *ja kan inte*.

#### ***2.4 Speaking to persons with comprehension problems***

Another situation when long expanded forms can be used is in speaking to persons who, for some reason, have difficulties understanding speech. This could be language learners or person with communication disorders or cognitive disorders. Also here, slow speech and more use of emphatic stress and hypercorrect articulation are common.

#### ***2.5 Speaking to a speech recognition program***

The increasing use of speech recognition programs and the design of such programs, have led to an increased focus on the difference between more written-like speech forms and other, shorter, speech forms. The use of more 'written-like' forms is, for example, today often required for a successful result using dictation programs.

Since successful speech recognition usually involves producing a written language form from a spoken input, this task is easier if the spoken input is as "written like" as possible (cf., for example, Jurafsky 2000, Schneiderman 2000). Speech recognition has, therefore, often taken over the notion of the written language form as the correct form also for speech. However, this is of course dependent on how speech recognition is designed and whether it is highly dependent on recognizing written-like units, like letters or phonemes, from the speech signal. An interesting question is whether more holistic, pattern recognition methods connected to semantic representations could be more successful for covering a wider range of speech forms. This could make 'natural connected speech' easier to use with speech recognition systems. If this is not possible, speech recognition will probably continue to be dependent on written-like speech forms also in the future.

#### ***2.6 Everyday speech***

Above, we have enumerated and exemplified specific contexts for written-like forms of speech. However, the unmarked case of everyday speech in most social activities does not require these variants, but rather fast and short forms that are mostly produced and perceived/understood in a context

with affordances providing disambiguation and interpretation of short and fast speech forms in connected speech (Allwood 2001). Since this is how most people in most situations speak, the alternative short forms should be of great interest also for speech recognition, if it is to acquire a wider use in more social everyday activities. This is a challenge to be met.

All the forms in examples 1 and 2 above can occur in everyday speech, the shorter ones especially in fast speech. There are a number of well-established spoken forms, in Swedish for example, the spoken form *de*, which can correspond to many written forms, such as *de* 'they', *det* 'it', *de är* 'they are' and *det är* 'it is'. The spoken word *jaga* ('hunt') can correspond to four different written forms (*jagade* – *jagar* – *jagat* – *jaga*). Another form is *å*, which occurs on the one hand in phrases like *han gick å la sej* ('he went and put himself to bed/he went to bed') and *de e kul å sjunga* ('it is fun to sing'). The corresponding written forms are *han gick och la(de) sig* and *det är kul att sjunga*. The spoken word form *å*, thus, corresponds to the written forms *och* ('and') and *att* ('to'). In many contexts, many speakers are uncertain about which of the written corresponding forms should be used. This does not lead to any problems in spoken interaction, but it does lead to a mix-up of the two written forms in some expressions and it is a phenomenon that speech recognition has to handle.

### 3. When are multiple forms a problem?

If we assume that multiple forms are represented and used in speech, this seems to be unproblematic in most spoken interaction, as exemplified above. So an interesting question, related to our examples above of contexts where written-like forms are often used, is when (if ever) multiple forms are a problem?

The short answer is that they are a problem when writing/spelling, reading in a normative context, directed to monolingual-monocultural standardization.

As we have seen, they are a problem for automatic speech recognition, which is usually aimed at some type of written language output. Here, our question was if there could be other types of speech recognition, for instance, with an output of pictures, meaning representations or direct response utterances (which do not have as a recognition stage a written language type representation).

Are multiple forms a problem when we talk? This is mostly not the case for accomplished first language speakers. But they can be a problem for (second) language learners, who have to learn several different forms and often are misled by the emphasis on written forms in traditional language teaching. Attempts to use different spoken variants can often result in stylistic mistakes, due to lack of experience regarding how contextual factors determine which form can be used when.

#### **4. Some questions inspired by Danish "sjusk"**

In the light of discussions about increasing reduction of spoken forms, for example, in Danish (cf., for example, Schachtenhaufen 2010), some questions can be asked.

The first question is if multiple forms of expression are a new phenomenon? This is probably not the case, more likely it is only considered a problem in a normative context directed to monolingual-monocultural standardization. This is surprisingly also often the case for speech recognition.

A second question concerns how we talked earlier in history. Did people speak more clearly, with longer word forms earlier? The speech recordings from earlier times that are available cannot really provide an answer to this, since the mere fact that the speech was recorded provided a formal context with a strong normative pressure. This was reinforced by the fact that recordings were new and that most recorded speech represented specific formal or theatrical contexts, such as political speeches, news readings, sermons and theatre plays. All these circumstances create an illusion of a written-form-like pronunciation, partly obscuring the fact that we know very little about ordinary speech in earlier times.

A third question concerns differences between different languages regarding the extent to which spoken language forms are represented as written forms. If we compare, for example French, English, Danish, Swedish, Norwegian and Finnish, we know that the writing system is quite close to spoken forms in general in Finnish, slightly less, but also quite close in Swedish and Norwegian and considerable less close in French, English and Danish. These typological differences affect the approach to handling the multiple speech forms and how much this comes into focus, for example, in

language teaching and in research and development concerning speech recognition.

## 5. Conclusions

We have argued that ontogenetic language development points to learning short forms, rather than long written language forms first. We do not start to learn the forms by learning reduction rules from longer forms that, in fact, are not learned before we learn to read and write.

From an acquisition point of view, expansion of naturalistic short and primary speech forms to longer, more elaborated and written language like forms seems more plausible than a reduction of such written-like forms to speech forms.

However, the fact that we are able to understand and use multiple forms quickly, makes it more plausible that we learn several forms rather than learning to generate forms by context sensitive expansion rules. As a result of this, our third option - multiple potential possibilities of pronunciation, seems the most likely one.

An attractive feature of assuming multiple expression forms constituting an "expression potential" is that it can make the content side and the expression side of language more similar. This can be done by combining an account of expression potentials with an account of semantics based on "meaning potentials" (cf. Allwood 1999), i.e., that morphemes and words carry multiple potential meanings and the context determines which of these meanings is realized. Assuming an expression potential provides a parallel perspective on forms of expression.

## References

- Allwood, Jens. 1999. Semantics as meaning determination with semantic-epistemic operations. In Jens Allwood and Peter Gärdenfors (eds.) *Cognitive Semantics*. Amsterdam: John Benjamins. 1–8.
- Allwood, Jens. 2001. Corpus based spoken language and computational research. In Henrik Holmboe (ed.) *Nordisk Sprogteknologi: Nordic Language Technology. Årbog for Nordisk Sprogteknologisk Forskningsprogram 2002-2004*. København: Museum Tusulanums Forlag. 59–68.

- Jurafsky, Daniel. 2000. *Speech and Language Processing, An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, N.J.: Prentice Hall.
- Schachtenhaufen, Ruben. 2010. Looking for lost syllables in Danish spontaneous speech. In Peter Juel Henriksen (ed.) 61–88. *Linguistic Theory and Raw Sound* [Copenhagen Studies in Language 40]. Copenhagen: Samfundsliteratur.
- Schneiderman, Ben. 2000. The limits of speech recognition. *Communications of the ACM* 43: 63–65.



# Phonetic reduction

Lars Brink

[labri@webspeed.dk](mailto:labri@webspeed.dk)

## 1. Definition

The following definition is meant to delimit, and state more precisely, our not too sharp concept of 'phonetic reduction'. Phonetic reduction is

articulation energy loss related to the energy maximum of the expression concerned in natural spontaneous speech, apart from volume, intonation, and speed.

Reductions take place according to fixed, traditional rules of the language, but may also be due to individually lowered motor control. Examples of reduction in Modern Unlocalisable Danish and High British are:

Danish: [ˈkr̥ðhus ˌb̥l̥æsŋ] → [ˈkr̥ðüs ˌb̥l̥æsŋ] → [ˈkr̥ðs ˌb̥l̥æsŋ]  
*Rådhuspladsen* 'The City Hall Square'.

High British: [ˌɹɒbɪn ˈhʊd] → [ˌɹɒbɪ ˈhʊd] → [ˌɹɒmn ˈhʊd] *Robin Hood*.

Between upper articulation and the phenomena of glottis amplitude, frequency, and speed there is the essential difference that the former has a natural maximum. *Konge* 'king' is pronounced [ˈg̊hɒŋə] and these sounds cannot be supplied with more upper articulation energy than their well-known positions require. Nor can more energy requiring sounds be added. Whereas loudness, intonation and speed have no clear energy maximum. One may speak as loud and as slowly as one is capable of and with as big intervals as one's mood dictates.

*Volume* (primarily vocal cord amplitude), of course, serves the same purpose as reduction: comprehensibility respectively energy saving within the limits of comprehensibility. It is correlated (negatively) with reduction: the louder, the less reduction. But only rather weakly so. If you are asked to

speak more clearly by a normal person sitting 1 m away, you do not raise your voice much whereas your reductions decrease drastically.

*Intonation.* Utterance intonation is even farther away from reduction albeit they are correlated. Only interval size is relevant since bigger intervals require higher, more 'muscle tense' notes. But interval size has another purpose than upper articulation energy in that it does not serve comprehensibility; rather, it signals eagerness, inquisitiveness etc. This is clearly seen from the fact that a maximal interval reduction – as long as some interval remains – in itself does not make speech unintelligible. (It just sounds dull). While a drastic reduction of the individual sounds makes the speech quite blurred.

*Speech speed* can physically be defined as organ movement speed, but is perceived as the number of actual sounds (or actual syllables) per second. Speed is narrowly correlated with reduction: If you speed up, you are bound to have more reductions, and vice versa. (For instance, the common pronunciation of *hende* (her) [ən] is completely 'unphonetic' at a slow speed).

Speed, however, has a complicated relation to energy. If certain body movements are to be performed, you save nothing by performing them quicker. More speed requires more energy, but the cost, then, is regained with the time saving. When you reduce the upper articulations, your sounds approach the organ position of rest: Vowels are centralised, plosives lose their closure, sounds are devoiced etc. So the organ movements are shortened whereby more sounds are said per second at the same organ speed. This, of course, is no real speed increase. (As certain sounds are omitted, the number of sounds per second actually goes down on this account!).

What people exploit by a high speed is the 'lingering' of the vowels (and the other syllabic sounds and possible long consonants). Syllables are syllables because one of their sounds has a non-minimal duration (is a bit more lingering than the others), most obvious, of course, in the *long* vowels, but it also applies to the short ones. Here length can be 'eaten' in an energy cut. A long vowel like in *Lise* ['li:sə] may easily take 0.25 sec., but at a normal speed less than half of it.

One might consider including speed in the definition of reduction, but it must be admitted that this kind of energy saving (the time during which

energy is spent) is fundamentally different from the articulation savings i.e. reductions of the very routes of the organs.

*Comprehensibility.* It is tempting to define reduction by comprehensibility, cf. German *Sprich deutlich!* or Danish *Tal tydeligt!* 'Speak clearly!'. The goal of speech (first time round) is comprehensibility. You climb up the comprehensibility ladder to be understood – and downwards to save energy within the limits of comprehensibility. However, comprehensibility – correlating narrowly with reduction – is a *consequence* of reduction, not a *kind* of reduction. You *intend* to talk more comprehensibly *by means of* approaching the articulation energy maximum.

That comprehensibility is not directly built into the concept of reduction, is seen from the fact that comprehensibility *can* go down when articulation energy goes up. If you use very distinct (energetic) pronunciations in normal conversation, your partner may be confused. If, for example, one says ['kʊʂdnə] *rustne* (rusty) instead of the normal ['kʊʂnə], this will confuse the receiver ("Was it *rustende* (rusting)?" – in this word [d] cannot be elided). An American-English super-distinct ['strɛjtən 'æqt] *straighten out* instead of ['strɛjt̩ 'æwɪ] will puzzle him ("...*Turn out?*").

Note that in this section I have pointed to some rather undisputed phonetic phenomena without documentation. My above contribution, however, is a concept analysis and, therefore, not directly subject to empirical evidence.

## 2. The pedagogical problem

All languages have their rules of reduction – though many of them are shared. They make up a serious pedagogical problem. At school you must of necessity learn the most distinct pronunciation, but, for a long time, you will still not understand foreigners speaking normally together. (Having learnt French for 3 years and achieved a reasonable ability to read and some volubility, I did not in Paris understand *fnet*. I even thought there could not be such a word, the phonotactics were wrong. It had to be repeated to me several times before the window opened: *Oh* [fə'ne:tʁə]!)

The pedagogical solution is not to drop distinctness (you cannot; as soon as you mention a word materially you are sure to get high up in the hierarchy), but to make more of listening to genuine everyday speech.

### 3. Distinctness level vs reduction level

It is important to distinguish between *distinctness level* and *reduction level*. Each word has a series of reduction levels down to a faint buzz. For example, *sten* 'stone' before a pause:

*sten* 'stone' ['sd̥e:ʔ<sup>e</sup>n → 'sd̥e:ʔn → 'sd̥e:ʔŋ]<sup>1</sup>

*tiger* 'tiger' ['d̥<sup>shi</sup>i:jɑ → 'd̥<sup>shi</sup>i:ɑ → 'd̥<sup>shi</sup>iɐ]

The reduction level 0 (maximally distinct) does not simply correspond to the 0-level of another word. You may very well combine *sten*'s no. 0 with *tiger*'s no. 1 in the same bit of a second. – The single utterance parts are normally placed on a certain *general distinctness level* (i.e., can normally only just be understood at a certain distance). As an example, a high distinction level requires the lowest (0th) reduction level of the word *sprække* 'crack', but is easily combined with reduction level no. 1 of *ikke* 'not': ['ɛ̝ŋ]. The standard-distinct level of a certain word (i.e. neither especially clear nor unclear) will require a certain reduction level apt for this specific word. (See Brink and Lund 1975, § 48).

### 4. Historical change vs reduction

Reductions must be kept apart from historical changes. The reductions do not offer *new* pronunciations. The reductions existed in older and extinct generations too. *When they disturb, it is only because they are carried to excess (e.g., due to sloppiness) so that the receiver does not understand the sender.* All normal, sober etc. people can speak clearly. (I cannot prove it, but I have checked it several times). The historical changes, however, show quite new pronunciations. When, for example, all youngsters today say ['f̥ɔ̝ɑ.mʔ] (forward) instead of ['f̥ɔ̝ɑ.æmʔ], it is not reduction. And it can be said – and is said – just as slowly and loudly as you wish. Whereas a reduction like [minj 'vɛŋ 'g̥ɔ:ʔ] *min ven går* 'my friend goes' cannot be said slowly.

But it is more complicated than that. All sound laws – whether they imply saving articulation energy or not – show *vascillation* in the beginning, where the new form 'prefers' quick and subconscious passages (Brink

<sup>1</sup> In the phonetic transcriptions the symbol [ʔ] represents the Danish *stød*.

and Lund 1975, § 134). For example, before everyone said [g̊k̥ɑːw] *krage* 'crow', this form competed with the old one, [g̊k̥ɑːɣ̥], which tended to appear in more distinct passages. During the transition period the new form was both a reduction result and a historically new form. (Paradoxically, [w] with its double articulation requires *more* energy than [ɣ̥]. I have not adjusted my definition of reduction to this very special circumstance).

### 5. Three Danish reductions

1) All vowels outside maximally distinct speech are centralised, i.e. relaxed and approaching the position of rest of the tongue and lips. The vowels being closest to the centre, viz. [ɐ, ɛ̥], are rather easily reduced all the way to the lax neutral vowel, [ə], provided they are short.: [min 'vən g̊ʰʌm 'əg̊ 'jəm'] *min ven kom ikke hjem* 'my friend did not get home'; in unaccented positions the reduction is even more "willing": [hæn ɑ̥ səð 'f̥k̥ɑ̥æm' d̥ʰə 'd̥ɛ mə mədmɑ̥.'xi'i] *han har set frem til det med Mette Marie* 'he has been looking forward to the Mette-Marie thing'. Other vowels are less willingly reduced all the way to [ə], the velar vowels almost not at all. So, the starting-point distance to [ə] is decisive, together with the articulation speed which is much higher in unaccented syllables.

It is highly interesting, however, that also the frequency of the word has an effect on its pronunciation: ['b̥h̥ɔ̥ ləg̊s̥'ɔ̊g̊l̥, ɛðð] *pålægsschokolade* 'chocolate in thin slices' is a normal form that only your receiver's hardness of hearing or a blizzard would force you to make clearer, but the unestablished *vadsækskolonnade* 'colonnade where young people place their duffel bags' will not exhibit corresponding reductions – it would not be understood. Likewise, [d̥ən'] *dén* 'that' is quite common, whereas ['mən'] *mænd* 'men' demands fatigue or depression.

As you see, not only the general distinction level of the *sentences* are adjusted to the situation, but the reduction level of every *word* is adjusted against the hearer's understanding: You choose – depending on distance, shyness, dedication, personal habits etc. a general *distinction level* – characterised by the average sounds per second, correlating well with the time distance between the primary stresses. But within this level, the *reduction level* for each word is very different: high in the unaccented syllables and the most trivial words, low in the more rare ones.

The rarity applies to the sender's belief about the hearer. The head doctor will pronounce the word *pankreas-akylion* with a high reduction level to the nurse, e.g., not to the patient, it is to be hoped. The politicians in the Danish Parliament say ['sɔ̃ɛ: 'sm̩, nisɔ̃ɽɿ] *statsministeren* 'the prime minister', we laymen prefer fewer reductions.

2) A seemingly uncomplicated reduction is the law of h-loss after s and in unaccented syllables: *Skovs'oved* (Danish place name), 'Kurt ar 'tabt 'Kurt has lost'. But it is not as simple as that. At the very start of the utterance *h* is not dropped: *Har Kurt tabt?* 'Has Kurt lost?', *Hotellet var godt*, cf. *et godt otel* 'a good hotel'. It might be due to some general decrease in distinctness from the beginning of an utterance (or even a word) to its end.

But even with such detailed rules, something important is missing: *H* may drop, but of course it has not got to. It depends, as we saw, on the situation and the individual. But what is quite independent of the situation and speaker, is the *relative strength* of the reduction:

Some reductions almost always apply, others are weak and rare. Consequently, some reductions can only appear when certain others have taken place. *Dit herbarium* 'your herbarium' sounds maximally distinct like [ɔ̃ɽɿ hæ̃'ɔ̃ɽ: 'j, ɔ̃ m] with older people and like [ɔ̃ɽɿ hæ̃'ɔ̃ɽ: 'i, ɔ̃ m] with younger people. Everybody can reduce these to:

[ɔ̃ɽɿ hæ̃'ɔ̃ɽ: 'j, ɔ̃ m]

[ɔ̃ɽɿ hæ̃'ɔ̃ɽ: 'jɔ̃ m]

[ɔ̃ɽɿ hæ'ɔ̃ɽ: 'jɔ̃ m]

[ɔ̃ɽɿ æ'ɔ̃ɽ: 'jɔ̃ m]

[ɔ̃ɽɿ æ'ɔ̃ɽ: 'jm] and younger people further to:

[ɔ̃ɽɿ æ'ɔ̃ɽ'jm].

(The variety *herbarie* is ignored). The combination [ɔ̃ɽɿ æ̃'-] does not exist since ɣ is dropped before h. This phenomenon is called reduction harmony.

3) Let us look at the reduction of the plosives [ɔ̃, ɔ̃] between vowels. They are, here, reduced → [ɔ̃, ɔ̃], i.e. they are not being completely closed, so that a minimal noise arises. Since [ɔ̃, ɔ̃] are convenience-determined as regards voice (i.e., they have no fixed glottal position but assume ever max-

imally distinctly, the easiest one), they must, here, be reduced to voiced sounds, for example, [ $'\varepsilon:\text{b}\text{p}\text{i} \text{'b}\text{h}\text{q}\text{g}\text{p}$ ] *aber i pakker* 'monkeys in packs'. The reduction, however, is very weak in accented anlaut: *Få bader, Det går ikke* 'few bathe; it does not go' owing to the fact that all sounds are slightly more energetic and longer in accented anlaut than in other positions. On the other hand, the reduction is a bit stronger if a neighbouring syllable contains a [ $\text{b}, \text{g}$ ] too: [ $'\text{b}\text{i}\text{b}\text{ə}, \text{h}\text{p}\text{l}'\text{ə}$ ] *bibeholde* 'retain', [ $\text{d}\text{i} \text{g}\text{i}\text{g} \text{'g}\text{l}\text{e}:\text{ð} \text{'j}\text{e}\text{m}'$ ] *de gik glade hjem* 'they went home happy'.

This rule is interesting, dissimilatoric as it seems, because a repetition is not especially energy requiring. It might be caused by the well-known fact that repetitions are irritating, so that the motivation for reduction increases.

Finally, I would like to mention that some highly frequent words have their own word-specific reductions. *Sku', ka', ve', ve', te'* instead of *skulle, kan, vil, ved, til* 'should, can, will, at, to' are examples (all common, even accented) as well as [ $'\text{g}\text{e}\text{m}'\text{m}$ ] *gennem* 'through' and, with some, [ $\text{æ}'\text{m}\text{e}\text{n}'\text{li}$ ] *almindelig* 'common'. Cf. English *not* → *n't* with stress reduction. These reductions do not follow the general and productive rules, but are independently handed down. (Accordingly, it is impossible to say \* $['\text{g}\text{h}\text{u}]$  *kulde* 'cold (sb.)' or \* $[\text{æ}'\text{m}\text{e}\text{g}\text{d}\text{i}]$  *almægtig* 'almighty'.) An amusing example from English is [ $'\text{s}\text{i} \text{'r}\text{æ}\text{ɔ}\text{n}$ ] *Sit down!* with [ $\text{d}\text{d} \rightarrow \text{d}\text{d} \rightarrow \text{d} \rightarrow \text{r}$ ], a reduction succession not found elsewhere (Jones 1966[1909]: 73).

The field of phonetic reduction – or the study of allegro forms – may sound dull. However, the attraction lies not in the fact that sounds reduce and may fuse to a 'phonetic mush' which is hard to distinguish segmentally. Rather, what attracts this author is the huge, strongly automatised rule system controlling it. But also the sender's considerations about his receiver are interesting. They go far beyond linguistics, since judgement of distance, storm, and the receiver's expected perception and horizon is not a matter of language.

## 6. Sloppiness

The reduction system, of course, has nothing to do with sloppiness. All languages have such a system, its purpose being to save needless articulation energy. Indeed, there is a lot to save – compared to a scenery where we all speak maximally distinctly to each other. You may, for instance, realize

this in an oral, mocking phenomenon. In many languages it is a traditional teasing to speak maximally distinctly where the situation does not require it at all. A kind of read-my-lips speech. For example, a husband under suspicion may say: "'I 'HAVE 'NEVER 'KISSED 'MONICA!!" in such an exaggerated way, including an overspending of articulation energy, that, thereby, he co-signals: "I've told you 7 times. Now you must have grasped it. Don't expect me to discuss it anymore!". If I take it down in accurate phonetic transcription, everyone can see that it is super distinct (pauses and primary stresses), nobody, however, sees that it is *hyper*-distinct. The founding fathers of IPA, as well as the Danish Otto Jespersen, have not given us means to render speech which – in a certain situation – is excessively distinct. For instance, they have not enriched us with signs for speed and loudness.

No pronunciation is sloppy per se. A faint mutter may be addressed to oneself as a reminder of buying cheese, whereas a pronunciation with few reductions may be sloppy because the speaker disregarded a foul radio in the background.

Phonetic sloppiness is: not being *sufficiently* clear (or distinct, or easy to understand) in a particular situation although one could easily have avoided the indistinctness.

Many people today criticize the youth for speaking in a sloppy way. With justice, I *suspect*. For it is my *impression* that young people ask each other to repeat their words more often than we did in 1950–1970. I sense that a certain bravery about speaking quickly has crept in. But it is a phenomenon that calls for systematic research. It would not be too difficult to go through the many recordings of the Danish Radio Company with young people talking freely in the 1960s and compare them with youngsters of today in similar situations (typically informal round table discussions). Suspensions are not enough.

## References

- Brink, Lars and Jørn Lund. 1975. *Dansk Rigsmål I-II*. Copenhagen: Gyldendal.
- Brink, Lars, Jørn Lund, Steffen Heger and Jens Normann Jørgensen (eds.). 1991. *Den Store Danske Udtaleordbog*. Copenhagen: Munksgaard.

Jones, Daniel. 1966[1909]. *The Pronunciation of English*. Cambridge: Cambridge University Press.

### Afterwords

*Dansk Rigsmål* (Brink and Lund 1975) contains mentioning of reduction in many places, especially in § 48 and § 134. Our material and method is described in § 3. In *Den Store Danske Udtaleordbog* (Brink et al. 1991) each word is shown in a distinct and indistinct version (determined by general or individual rule). Besides, all general reductions known to the authors are listed on pp. 1573–1631. Before the publication of *Dansk Rigsmål* in 1975 there was no serious treatment of Danish reduction. But in recent years the field has received much attention. The following list is almost exhaustive.

Grønnum, Nina. 2007. *Rødgrød med fløde, en lille bog om dansk fonetik*, Section 13.2. Copenhagen: Akademisk Forlag.

Heegård, Jan and Peter Juel Henriksen (eds.). 2012. Thematic section in *Danske Talesprog* 12.

Katlev, Jan. 1981. Mangt om meget. *SAML* 8.33–51.

Katlev, Jan. 1989. En drøftelse af 'sjuskefonologi' i moderne københavnsk rigsmål. In *2. Møde om Udforskning af Dansk Sprog*, Mette Kunøe and Erik Vive Larsen (eds.), 190–198.

Pharao, Nicolai. 2009. *Consonant reduction in Copenhagen Danish. A study of linguistic and extra-linguistic factors in phonetic variation and change*. PhD thesis, University of Copenhagen.

Schachtenhaufen, Ruben. 2007. Realiseringen af schwa i spontan dansk tale. MA thesis, University of Copenhagen.

Schachtenhaufen, Ruben. 2010. Looking for lost syllables in Danish spontaneous speech. In *Linguistic Theory and Raw Sound*, ed. Peter Juel Henriksen [Copenhagen Studies in Language 40], 61–88. Frederiksberg: Samfundslitteratur.

Schachtenhaufen, Ruben. 2012. Reduction of word final schwa in Danish – rhythmic and syntactic factors. In *Speech in Action. Proceedings of the 1<sup>st</sup> SJUSK Conference on Contemporary Speech Habits*, eds. Jan Heegård and Peter Juel Henriksen [Copenhagen Studies in Language 42], 53–71. Frederiksberg: Samfundslitteratur Press.

Schachtenhaufen, Ruben. 2013. *Fonetisk reduktion i dansk*. PhD dissertation, Copenhagen Business School.

# Pragmatic functions of lengthenings and filled pauses in the adult-directed speech of Hungarian children

Andrea Deme

[Research Institute for Linguistics, Hungarian Academy of Sciences](#)

[deme.andrea@nytud.mta.hu](mailto:deme.andrea@nytud.mta.hu)

## Abstract

*Two most common disfluencies of spontaneous speech, vowel lengthenings (VLE) and non-lexicalized filled pauses (NLFP) were investigated in the adult-directed speech of eight Hungarian children. Though VLE and NLFP might seem to be similar vocalizations, recent investigations have shown that their occurrences might differ remarkably in child speech and may also change as a function of age. Based on these findings, in the present study the functional analysis of VLEs and NLFPs was performed. It was hypothesized that in child speech the two phenomena have roles not only in speech planning, but also in discourse management, and that they show functional distribution. The analysis provided evidence that VLE is more common than NLFP. VLE often tends to mark discourse events and may play a role in turn-final floor-holding strategies, while NLFP is mostly connected to speech planning, and occasionally, it may also participate in turn-taking gestures, as well.*

**Keywords:** *Child speech; vowel lengthening; filled pause; floor-holding; floor-ceding; turn-taking; turn-allocation.*

## 1. Introduction

Lengthenings (hereafter, LE) and filled pauses (hereafter, FP) are probably the two most common phenomena of spontaneous speech (see e.g. Ecklund 2001, Horváth 2004). However, while FP has been in limelight for a relatively long time, and is a topic of ongoing dispute, LE is less in the scope of interest. Regarded both as disfluencies, LEs are often treated similarly to FPs. Due to this prevailing tendency, further analysis or detailed comparison of the two phenomena is rarely elaborated on. The objective of the pre-

sent study is to continue the analysis and comparison of LE and FP initiated and proposed by Deme (2012) and Deme and Markó (2013) (see Section 1.3) through addressing the question of their functional distribution. The analysis is limited to the domain of child speech since it has been given less attention so far (regarding LEs and FPs). Comparing the results of research in child speech to previous findings on adults can reveal important aspects regarding not only the functions, but the acquisition of the two phenomena, as well. For the sake of a more extensive analysis, distinguishing between lexicalized and non-lexicalized FPs on the one hand, and differentiating among many kinds of segment LEs (vowel LE, consonant LE, syllable LE, etc.) on the other, the investigation is narrowed down to *vowel lengthenings* (hereafter, VLE), as in *répalevest* [re:palevɛʃt] 'carrot soup', and *non-lexicalized (vocal) filled pauses* (hereafter, NLFP): *ő* [ø:], *mm* [m] and *öm* [øm]. In order to see whether there is enough evidence for the rough generalization treating LEs and FPs or VLEs and NLFPs as one category, we start with a brief overview of some examples of the available literature.

### ***1.1 LEs and FPs in adult speech***

In the majority of the fundamental studies of the issue, FPs are defined as disfluencies: artifacts of planning processes which help the speaker obtain time to plan his/her speech, and to execute processes of lexical access; therefore, FPs are in connection with cognitive load and cognitive effort or the difficulty of the task the speaker has to carry out (Maclay and Osgood 1959; Goldman-Eisler 1968; Clark and Fox-Tree 2002; Markó 2004; Gósy 2006; Corley and Stewart 2008). In other studies the function of floor-holding is also suggested. The authors propose that FPs help the interlocutors in turn-taking (especially in floor-holding) as well (Maclay and Osgood 1959; Schegloff 1982). In later works, Corley and Stewart (2008) argued, that FPs are important in attracting the listener's attention to the upcoming word, thus (in accordance with Clark and Fox-Tree 2002) FPs should be considered as intentionally planned and executed, independent items of language, namely words (not only as involuntary and automatic correlates of mental processes). To summarize, a wide range of functions for FPs have been proposed, including the domain of speech planning and pragmatics, as well, but there is no general agreement on their status.

LEs, however, seem to be the "dark horse of the disfluency stable" (Ecklund 2001). These phenomena are rarely mentioned or analyzed independently, and are generally regarded as simple disfluencies similar to FPs: vocalizations, which are the acoustic correlates of planning processes by signaling the difficulties the speaker encounters while speaking (see e.g., Horváth 2004; Giannini 2003; Esposito 2006). It is a telltale sign that the previous short description practically contains everything that is attributed to LEs in the literature. There is little amount of functions proposed and examined, also, there is no really accurate definition available. On the contrary, most definitions are not only non-extensive, but also fairly inconsistent which underlies the "underestimation" of LEs: most of the studies define LE based on to objective duration (using the terms "length" and "duration" in their descriptions), but apply the subjective auditory judgment of the researchers themselves in designating the phenomenon (see e.g. Ecklund 1999; Bell et al. 2000).

This unbalanced presence of LEs and FPs in the scientific discourse should warn the researcher to raise doubts about how well-founded the approach to handle LE and FP as two identical means of speech might be (NB no substantial evidence is provided). In addition, despite the proposed similarities of the two phenomena, there are some functions suggested for FPs which are not investigated regarding LEs at all. The discourse-related roles are a case in point – in the present proposal it will be suggested that these are the main functions of VLE in child speech.

### ***1.2 LEs and FPs in child speech***

In child speech less data on FP is provided than for adults, and practically no detailed studies on the analysis of occurrences and functions of LE have been conducted. The interpretation of FPs in children is similar to those found for adults, like "look to the novel referent" or "pay attention to the following word" (Kidd et al. 2009, in accordance with Corley and Stewart 2008). Independent analysis of LEs, however, are practically missing from the literature, thus no particular functions (typical of child speech, as well) are assigned to them.

As a matter of fact, items of discourse-management as such seem to be less frequently addressed in the child language literature. Items of discourse pragmatics may include discourse markers in the strict sense (here-

after, DMs; referring to lexical items that are imposing a relationship between discourse segments, see Fraser 1999), and other lexical and non-lexical means, as well, which can be used for controlling turn-allocation or can be part of floor-holding or floor-ceding strategies (e.g. marking turn starts or turn ends).

Broadly speaking, studies investigating the discourse-management of children can be divided into three main groups. The first group of works covers mainly the development of timing and pausing as cues of discourse management skills. They investigate gaps, overlaps, and the effectiveness of interruption in children's speech (Ervin-Tripp 1979; Esposito 2006; Tice 2010; Tice et al. 2011). The second group of papers investigates the development of the usage of DMs, purely in the strict sense (Kyratzis and Ervin-Tripp 1999; Choi 2007; Markó and Dér 2011). Discourse functions filled with (or supported by) non-verbal items are, on the other hand, less often in the focus of research. The third group deals with these non-lexical means of discourse-management. They demonstrated, for instance, various functions of *humming* in Hungarian child speech (Markó et al. 2010) and suggested that children might also use non-verbal response marking (response-initial) means, like *uh* and *um* aiming to express floor-holding intentions (Tice 2010).

### ***1.3 Novel initiatives in the study of LFs and FPs, hypothesis***

Recently it has been demonstrated that the frequency of occurrence and positions of NLFPs and VLEs might be different within a group of children at the same age and may also vary with age (studied in 8-year-olds by Deme 2012 and in a comparison between 8-year-olds and adults by Deme and Markó 2013). Based on the results, the authors proposed that although NLFPs can be regarded as disfluency phenomena and means of speech planning by default (as it is usually suggested), VLEs cannot be treated the same way – at least in case of child speech.

They base their argument on four striking observations that are briefly summarized below in order to establish the present paper's approach and methodological considerations.

1. It was found that in ADS VLEs appear mostly in content words (in more than 70%), whereas in adults it is mostly the content-word-preceding function words, which tend to contain them (in more than 60% of the cas-

es). 2. In addition, in ADS, it is mostly the end of these content words (the very last syllable) where the phenomenon appears (in more than 70% of the cases). 3. VLEs in ADS were found to be most frequent in speech-session final positions (i.e. before pauses, in about 70% of the cases), while in adults it was less (about 50%), whereas the counts of VLEs in the other positions were roughly equally distributed in their material. By contrast, in ADS a great amount of NLFPs occurred in isolation (in more than 50% of the cases), while in adults, session initial and session final positions were the most frequent (about 30%–30%, respectively, thus accounting for 60% of all cases). 4. Finally, in ADS there were more NLFPs neighbored by other disfluency phenomena than VLEs (NLFPs: approx. 35%, VLEs: approx. 10%), whereas in adults VLEs and NLFPs were involved with the same frequency (in approx. 20%).

According to the above findings, it seems reasonable to argue that NLFPs and VLEs in adult speech and NLFP in child speech might have functioned as "coverage" for the difficulties of the processes of lexical retrieval (or grammatical planning), thus helping the speakers to gain time for resolution. However, this argument does not hold for VLEs in ADS, since in that case VLEs were not followed but contained by the content word and were positioned word-finally and session-finally. As the usage of VLEs in adults and NLFPs in both groups seemed to be primarily triggered by the error of a mental process, VLEs in child speech (according to the data of 8 children) seemed to have no obvious connections to error resolution. Based on their data, the authors proposed that in many cases the primary function of VLEs were discourse-related.

Motivated by the above results, the aim of the present paper is to examine the possible functions of VLEs and NLFPs in adult-directed speech of children. It is hypothesized that these means have functions over and beyond signaling planning difficulties, namely in turn-allocation. Moreover, the functions of VLE and NLFP differ. Through the present analysis some insight to the discourse management skills of children can be obtained. Providing evidence for the usage of an element in child speech with the exclusive function of managing the conversation (as suggested for VLEs in child speech) would be an important step, since that would provide direct evidence for the turn-allocation intentions of children as well.

The study targets ADS for a very simple reason. ADS is less frequently dealt with in the literature than peer conversations, because in adult–children interactions it is more the adults, who dominantly control speech timing (as, for e.g. Ervin-Tripp 1979). Though Ervin-Tripp’s argument may be valid to a certain extent, it is also true that ADS is a typical and common speech situation. In life our first experience in discourse is adult or parent-directed speech. Furthermore, ADS is a dominant part of frontal school instruction scenarios, as well, and it is often the verbal performance (carried out with the teacher) that is evaluated in the classroom. Therefore, the investigation of ADS situations is of great relevance.

## **2. Methodology**

### ***2.1. Subjects and material***

The occurrences of VLEs and NLFPs were analyzed in 8 Hungarian children’s (4 boys and 4 girls, age: 7–8) 45 minute long speech material. The corpus was recorded in quasi-interview situations, the children were asked about their holiday activities and their days in kindergarten by the adult interviewer. Designation of VLEs was carried out with a perception test involving 10 linguists (introduced in Deme 2012). The subjects were asked to listen to the speech samples using headphones in a quiet room, and meanwhile to follow the transcription of the texts presented on a screen (the transcription was orthographic, but without punctuation). The subjects’ task was to mark those vowels in the written texts which they felt to be lengthened based on their subjective auditory judgment. Vowels marked by at least 6 linguists were defined as VLE. NLFPs were identified based on the auditory perception of the author, together with the visual confirmation of a spectrogram. To avoid blending the data, only VLES occurring in function or content words were considered as VLES, while those occurring in FPs, DMs or any kinds of disfluency phenomena were counted as FP, DM or a disfluency phenomenon (and disregarded as VLE).

### ***2.2 Methods***

To determine the possible functions of the two phenomena in turn-allocation, floor-holding and floor-ceding strategies had to be detected in children’s speech. However, it was important not to label these arbitrarily since it would have led to circular reasoning: the means and formal features

causing the impression of a certain function or strategy serving as the bases of detecting it would have been found to be the means and form of that particular strategy. Consequently, the places of analysis had to be defined not on functional but on formal basis, in hope of finding clusters of congruent attributes afterwards which can be labeled as functional groups or "strategies".

For this purpose, in the first approach two places of supposedly different "transition relevance" (Sacks et al. 1974) were selected for analysis: the "places of grammatical completion" (PGC) (derived from Denny 1985) and the "places of possible turn ends" (PTE). In fact, both PGC and PTE share the property of grammatical completion or grammatical finiteness (i.e. they are points at which the utterance can be regarded as syntactically complete). However, while PGC was defined as a point of finiteness, which is not followed by a silent pause (SP) of any length, PTE was determined as a point which is followed by an SP. According to the results of Tice (2010), who demonstrated the tendency of children using longer gaps in conversations, it was expected that PGC cannot serve as a real turn-end, while PTE can. After dividing points of grammatical completion to the classes of PGC and PTE it turned out that this suggestion holds: indeed speaker exchanges occurred only in PTE in our material. This means, that practically every speaker-exchange contained an SP, and every PGC was localized between the speech segments of children.

Based on these pilot results PGC and PTE were retained as the points of analysis and two suppositions were formulated. First, it was suggested that the properties of PGC (namely the lack of an SP, the intonation contour, etc.) serve as a means of compensation strategy for the transition relevance implied by the grammatical finiteness. Thus, inspection of these properties might help us to get closer to floor-holding strategies in children. Second, it was assumed that PTE (with an SP and somewhat greater transition relevance) can be the point of the realization of 1. floor-ceding intentions, in which case means to *compensate for* the transition relevance (probably including the SP, as well) have to be present, or 2. floor-holding intentions, in which case means to *enhance* the transition relevance of this position have to be present. Accordingly, at PTE two separate set of properties were predicted which supposedly attend on the two distinct functions

of floor-holding and floor-ceding, whereas at PGC there is only one dominating setting expected which supposedly coincides with case 1 at PTE.

The analysis involved the inspection of the  $f_0$ -contour (one of the most salient variables of intonation) and the presence of LEs and FPs was determined.

The distribution of the total number of LEs and FPs was also assessed with regard to the function of marking discourse events and their appearance in speech planning strategies (by using the additional criterion of the presence of adjacent disfluencies, DPh, as suggested by Deme and Markó 2013). In this phase of analysis, the terms "turn start" and "turn extension" were introduced. *Turn start* refers to the event at which speaker exchange occurs and the second interlocutor (the child) starts a new turn. *Turn extension* covers those events at which there is no speaker exchange after a PTE and the child is extending his/her turn. Adhering to the premise presented in Section 1.3, every NLFP was associated with speech planning functions and was defined as DPh by default, but was labeled so only in those cases in which there were no other (secondary) functions suspected. If there were other possible functions specified, NLFP got the label of this function. VLEs found at discourse marking positions and not accompanied by (or participating in) other DPhs were always labeled as discourse marking elements.

At this point, in order to provide a more complete picture of the discourse events occurring in ADS and to be able to relate the results obtained for discourse marking VLEs and NLFPs to discourse events of ADS in general, other lexical DMs (namely, DMs in the strict sense, see Fraser 1999) were also sorted out and analyzed. These DMs were only used to contextualize the children's discourse-management strategies more accurately.

### **3. Results**

In the speech material of children VLEs appeared more often than NLFPs on average (a VLE occurred once in every 56 vowels, while a NLFP was present every 32 seconds) with great variability across speakers (Figure 1): there is one child, who uses VLEs (G1), another one uses both VLEs and NLFPs with the same frequency (B1), a third one uses NLFPs more than twice as often as VLEs (B2), and a fourth one whose use of VLEs was al-

most equally rare (G2). The figures of frequency seem to be remarkably different from those reported in adult speakers. On the one hand, the average frequency of VLEs in children is twice as high as in adults; on the other hand, the ratio of VLEs to NLFPs in children is reverse that of adult's (cf. Deme and Markó 2013).

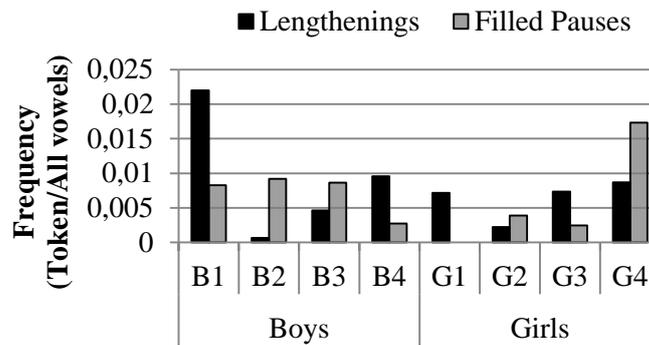


Figure 1. Frequency of occurrence of LEs and FPs per speaker.

As a general observation, it can be stated that no NLFPs were found in PGC and PTE marking functions, so possibly no significant role is played by NLFPs in turn-ending or turn-extending gestures in ADS. VLEs, however, have several appearances at these places (more than 10% in PGC and more than 24% in PTE) (Figure 2).

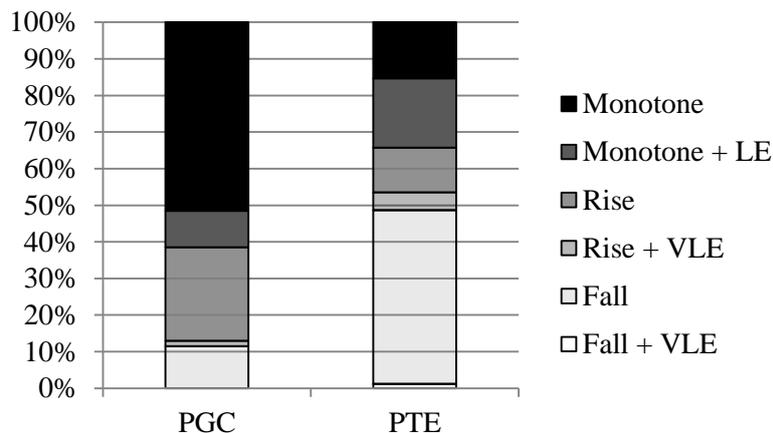


Figure 2. Summary of the properties observed at PGC and PTE.

The typical (and practically: only)  $f_0$ -contours found at PGC were monotonic (in 61.5%) and rising (in 26.5%). Monotonic and rising contours are interpreted as part of the compensatory strategies for the transition relevance of this position (cf. Section 2.2). Moreover, it is suggested that VLE might act as a means of floor-holding. The claim is supported by previous findings of the literature. As Varga (2002) states in his fundamental work on Hungarian intonation, the contour types of rise and monotone imply that the unit is not complete but a preparation for something complete or significant that follows, while fall suggests that the stretch of speech has come to an end. In her investigation of spontaneous speech Markó (2005) also found that rising contour is a common means at the end of speech sessions and clauses with the most probable indication of the willingness to continue. Nevertheless, intuitively, as well, one may suggest that monotonic and rising contour signify infiniteness in intonation, as opposed to fall, which is more likely to express finiteness. Since these contours are frequently accompanied by VLE (both at PGC and PTE), VLE is supposed to be a marker of this compensatory/floor-holding strategy as well. (However, the proportions of the appearance of VLE also suggest that in most cases the intonation-contour and the lack of an SP provide a probably sufficient signal, and VLE might only serve as a secondary marker; 10% of PGCs and 22% of PTEs are marked by VLEs.)

If the previously described cluster of monotonic and rising contours (with or without VLE) acts as one functional group, it can be concluded that there are two distinct and dominant clusters of properties occurring at PTE, as expected: 1. monotone and rise (with or without VLE) and 2. fall (with VLE, in a very few cases). Similarly to PGC, we suggest that the first group accounts for the strategy of compensation/floor-holding, while the second expresses floor-ceding intentions through giving the impression of finiteness conveyed by the descending ("final-sounding")  $f_0$ -contour (in accordance with the findings of Beattie et al. 1982 and the description of Varga 2002).

Furthermore, the suggestion regarding the role of VLE in compensatory/floor-holding strategies is supported by the closer evaluation of the properties observable at PTEs as well, since VLEs are almost never linked to the presumably "turn-passing" intentions and falling contours (as fall is accompanied by VLEs in less than 2%).

Figure 3 and 4 summarize the distribution of all of the occurrences of VLEs, NLFPs and the other DMs found in the material in terms of their function of marking discourse events and the participation in speech planning strategies.

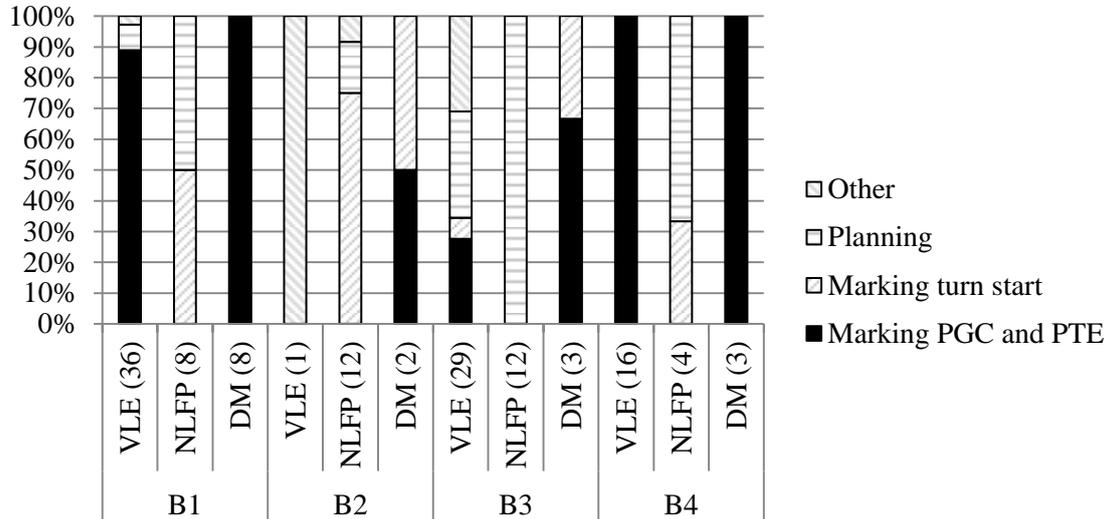


Figure 3. Occurrences of VLEs, NLFPs and other DMs in the boys' material. (The numbers in brackets indicate the total number of occurrences.)

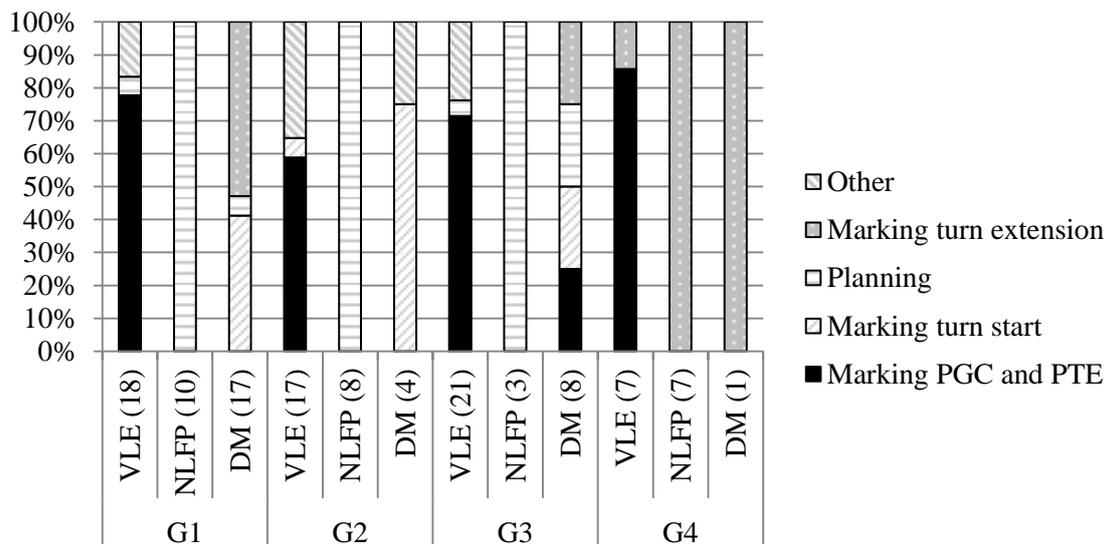


Figure 4. Occurrences of VLEs, NLFPs and other DMs in the girls' material. (The numbers in brackets indicate the total number of occurrences.)

Generally speaking, VLE appears most often in the material, as it is used by 6 children (B1, B3, B4, G1, G2, G3), while NLFPs are only used by 2 children (B2 and B3), and DMs (in the strict sense) are only used habitually by only one speaker (G1). Based on the number of adjacent DPhs, it can also be concluded that NFLPs seem to be more involved in speech disfluencies and speech planning processes than VLEs.

As far as VLEs are concerned, 5 out of 6 children use them as markers of the places of grammatical completion (or in a few cases, to mark turn start) with VLEs occurring in content words, and only one boy (B3) follows a different tendency. In his speech, the appearance of VLEs in function words (pronouns, conjunctions, articles) is more dominant. Interestingly, this pattern, which can be considered rare in children's material, is the pattern that Hungarian adults seem to follow, as well (according to Horváth 2004). Moreover, VLE is generally considered to be a time-gaining vocalization exactly due to these appearances, as it provides time for searching in the mental lexicon (if occurring in articles and pronouns before content words) or planning whole clauses (if occurring in conjunctions). However, this kind of usage is the most seldom in our material.

The case of NLFPs draws a more scattered picture. Those few speakers who use them quite frequently show divergent patterns: one of them uses NLFP for marking discourse events (turn starts), as well as signaling planning difficulties (B2), while others prefer to employ it only in the latter role (B3, G1) (along with the most children who use this means less often), and in a few cases, NLFPs also mark turn extensions (in G4). With regard to NLFP, in those children's speech, who use VLE in the function of signaling discourse events consistently (B1, B4, G1, G2, G3), 3 possible scenarios are observable: i) NLFP is only connected to speech planning difficulties (G2, G3), ii) NLFP has a role in planning as many times as in marking discourse events (B1, B4) or iii) NLFP is not present at all (G1). In other – non-prototypical – cases, in which VLE plays mostly the role of time-gaining, the general lack of discourse marking elements of any kind can be detected (B3).

There were 46 DMs found in the corpus. These (in an ascending order of frequency) are the following: *ilyen / ilyenek(et)* 'such' (28%) > *hát* 'well' (16%) > *és* 'and' + SP (14%) > *csak ennyi(t)* 'that's all' (14%) ~ *akkor* 'then' (14%) ~ *aztán* 'then'. In addition to these, other explicit forms were regis-