Journal of Phonetics 39 (2011) 253–260

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Editorial

An introduction to reduced pronunciation variants

ARTICLE INFO

Keywords:
Reduction pronunciation variants
Casual speech
Speech comprehension
Speech production
Experimental research
Corpus research

ABSTRACT

Words are often pronounced very differently in formal speech than in everyday conversations. In conversational speech, they may contain weaker segments, fewer sounds, and even fewer syllables. The English word *yesterday*, for instance, may be pronounced as [jɛʃeɪ]. This article forms an introduction to the phenomenon of reduced pronunciation variants and to the eight research articles in this issue on the characteristics, production, and comprehension of these variants. We provide a description of the phenomenon, addressing its high frequency of occurrence in casual conversations in various languages, the gradient nature of many reduction processes, and the intelligibility of reduced variants to native listeners. We also describe the relevance of research on reduced variants for linguistic and psychological theories as well as for applications in speech technology and foreign language acquisition. Since reduced variants occur more often in spontaneous than in formal speech, they are hard to study in the laboratory under well controlled conditions. We discuss the advantages and disadvantages of possible solutions, including the research methods employed in the articles in this special issue, based on corpora and experiments. This article ends with a short overview of the articles in this issue.

1. Introduction

It is common knowledge within the linguistic and psycholinguistic literature that every token of a word is different from other tokens of that same word in many respects. Tokens differ, among other things, in the characteristics of the speaker's voice, in speech rate, in the exact qualities of their vowels, and in the acoustic details of the consonants. Thus, a token of the English word *tea* pronounced by a petite female speaker might have a vowel with a second formant at more than 3000 Hz, while a token of the same word by a large man might have a vowel with a second formant near 2000 Hz. Two tokens by the same speaker, even if pronounced in the same sentence, might differ in duration of the vowel by some tens of milliseconds, or in voice onset time by some smaller number of milliseconds.

Importantly, pronunciation variation is far more extensive in spontaneous conversations than in formal or read speech. To give some American English examples,[1] the /k/ can be pronounced as [ç] in the word *weekend*, /g/ as an approximant in *you guys*, and /t/ and /d/ as well as some other segments may be absent or merged in the word *yesterday*, resulting in [jɛʃeɪ]. Some words, especially function words, may be absent in their entirety, such as *have* in *do you have time*, realized as [djutɛm]. We define these variants, characterized by incomplete articulatory gestures or fewer segments compared to the variants typical of read speech, as reduced variants. Reduced variants are very frequent in casual speech, and accordingly they have received attention in many sociolinguistic studies, showing how often individual speakers produce particular variants in daily-life settings where they use their native dialects. These studies focused especially on pronunciation variants with missing segments, in particular the absence of word-final /t/, (e.g., Guy, 1991; Labov, 1972). In contrast, reduced pronunciation variants have so far received little attention in the phonological, phonetic, or psycholinguistic literature, even though a large number of studies in these fields have investigated naturally occurring speech. Nevertheless reduced pronunciation variants have far reaching consequences for psycholinguistic models of speech production and comprehension, for language acquisition, and for speech technology. The present article forms an introduction to the eight research articles incorporated in this special issue on reduced pronunciation variants. These eight articles study the characteristics and the production and perception of reduced variants in four different languages, from several different angles. We discuss the phenomenon of reduction, including its frequency of occurrence within languages and across languages, the gradient nature of many reduction processes, and the intelligibility of reduced variants (Section 2). We then describe why research on pronunciation variants is indispensable for the formulation of psycholinguistic models of speech comprehension and production, for second language teaching, and for the development of robust ASR models (Section 3). Investigating the characteristics, production, and comprehension of reduced pronunciation variants is not simple, since these variants occur most often in spontaneous speech, which means that researchers must develop creative, novel methods to study it in a controlled fashion. We discuss these problems and possible solutions, including the research methods

[1] See http://www.u.arizona.edu/~nwarner/reduction_examples.html for sound clips of reduced speech examples.

0095-4470/$ - see front matter
doi:10.1016/S0095-4470(11)00055-6

employed in the articles in this special issue (Section 4). This article ends with a short overview of the articles incorporated in this issue (Section 5).

2. The phenomenon

2.1. Frequency of occurrence

Reduced productions form an important characteristic of spontaneous conversations in many languages. Lists (1) and (2) contain twenty examples of reduced pronunciation variants from our recordings in American English and Dutch, respectively, as represented through phonetic segmental transcriptions (for more examples, see, for instance, Johnson (2004), Greenberg (1999), and Shattuck-Hufnagel and Veilleux (2007) for English, and Ernestus (2000) and Ernestus, Baayen, and Schreuder (2002) for Dutch). They show that reduced pronunciation variants are common both for function and content words.

(1) Examples of reduced pronunciation variants in American English[2]

	Full form	A reduced form
he already	/hi ɔlɹɛɾi/	[iɹɾɪ]
chillin' in the	/tʃɪlɪ̃n ɪ̃n ðə/	[tʃ̃ɪ̃ə̃nə̃nː]
computer	/kʰəmpʰjuɾɚ/	[kʰpʰyɾɚ]
a little	/ə lɪɾl̩/	[ɔlə]
a little while	/ə lɪɾl̩ waɪl/	[ɔɾl̩wa]
you guys	/ju gaɪz/	[jyɣʌz]
get out or	/gɪɾ aʊɾ ɔɹ/	[gɛːaːɾɚː]
gonna go	/gə̃nə goʊ/	[ɣə̃nəɣ̩oʊ]
see it	/si ɪt/	[sij]
weekend	/wikɛ̃nd/	[wiçə̃]
yesterday	/jɛstɚɾeɪ/	[jɛʃɪ]
for a	/fɔɹ ə/	[fɨ]
do you have time	/du ju hæv tʰaɪm/	[tjutəm]
then I	/ðɛ̃n aɪ/	[ə̃aɪ]
you're just	/jɔɹ dʒʌst/	[ɥiz]
he was	/hi wʌz/	[ɥiz]
but I was like	/bʌɾ aɪ wəz laɪk/	[b̥ɹʌʒləɻ]
we were	/wi wɚ/	[wʲɚ]
out in the	/aʊt ɪn ðə/	[aʊ̯ːðə]
wouldn't you	/wʊdn̩t ju/	[ũːn̩ːy]

(2) Examples of reduced pronunciation variants in Dutch

	Full form	A reduced form	
aardrijkskunde	/ˈardrɛiksˈkəndə/	[arəskənə]	'geology'
allemaal	/ɑləˈmal/	[ɑməl]	'all'
computer	/komˈpjutər/	[pjutə]	'computer'
familie	/fɑˈmili/	[fmili]	'family'
gewoon	/xəwon/	[xon]	'normal'
inderdaad	/ɪndərˈdat/	[ɪdat]	'indeed'
in ieder geval	/ɪn ˈidər xəˈvɑl/	[ifɑl]	'in any case'
koninklijk	/ˈkonɪŋkləkˈ/	[koŋk]	'royal'
Nederland	/ˈnedərˈlɑnt/	[nelɑnt]	'Netherlands'
ongeveer	/ɔnxəˈver/	[ɔfer]	'approximately'
een gegeven moment	/ənxəˈxevəmoˈmɛnt/	[xefmən]	'a given moment'
overigens	/ovərɪxəns/	[ovəs]	'by the way'
precies	/prəˈsis/	[psis]	'precisely'
persoon	/pɛrˈson/	[pson]	'person'
politie	/poˈlitsi/	[plisi]	'police'
vakantie	/vaˈkɑntsi/	[fkɑsi]	'holidays'
verkopen	/vərˈkopə/	[fkop]	'sell'
volgend	/ˈvɔlxənt/	[flnt]	'next'
wedstrijd	/ˈwɛtstrɛit/	[wɛs]	'game'
zo zeer	/oˈzer/	[sː]	'that much'

The fact that reduced variants form an important portion of the words uttered in spontaneous speech is clear from several quantitative studies. Johnson (2004) studied 88,000 word tokens produced by 40 native speakers of American English in interviews. The tokens were first automatically provided with phonemic transcriptions reflecting the words' full forms, which were then manually corrected by trained phoneticians. Johnson found that 6% of the content word tokens and 4.5% of the function word tokens were produced with at least one syllable less than their full forms. In addition, 5–12% of the word tokens lacked at least one segment as compared to the words' full forms. These findings are in line with the syllable deletion rates (9%) reported by Dalby (1986), who manually transcribed television interview speech in American English, and with the corpus data from Greenberg (1999) and Shattuck-Hufnagel and Veilleux (2007). Schuppler, Ernestus, Scharenborg, and Boves (2011) reported similar results for Dutch. They automatically transcribed 10 informal conversations among friends, consisting of 153,200 word tokens produced by 20 male native speakers, and reported a syllable deletion rate of 19% and a segment alternation/deletion rate of 40%.

A number of sociolinguistic and phonetic studies have investigated which types of speakers use reduced pronunciation variants (e.g., Guy, 1992; Keune, Ernestus, Van Hout, & Baayen, 2005). These studies convincingly show that reduction is not just a feature of a small subgroup of the population, used, for example, by young women belonging to specific social groups. (In the US, for example, some listeners assume reduced speech is "Valley Girl talk".) All speakers use reduction, although there are differences among social groups. In general, men reduce more than women, young people reduce more than older people, and speakers from different regions may differ (e.g., native speakers of Dutch tend to reduce more if they come from the Netherlands rather than Flanders).

Reduced pronunciation variants are not restricted to English and Dutch. List (3) shows some examples from other languages, including non-Germanic languages.

(3) Examples of reduced pronunciation variants in other languages (taken from Canavan & Zipperlen, 1996; Engstrand & Krull, 2001; Kohler, 1990; Lennes, Alarotu, & Vainio, 2001; Torreira & Ernestus, 2011 and our students' and our work).

French	*c'était*	/setɛ/	[stɛ]	'was'
Finnish	*niinku*	/niːŋku/	[nik]	'like'
German	*wagen*	/vaːgən/	[vaːŋ]	'car'
Japanese	*de aru*	/de aɾɯ/	[deaɾ̥ɯ̥]	'be'
Japanese	*nihongo*	/nihoŋgo/	[ĩj̃õːj̃]	'Japanese language'
Mandarin	*bu zhi dao*	/bu tʃɹ dao/	[bəɹao]	'don't know'
Mandarin	*ban zhang*	/ban tʃaŋ/	[bãnjã]	'section leader'
Swedish	*som alla*	/somalːa/	[smala]	'as all'
Korean	saenggakp'oda	/sɛŋgakp̚oda/	[sampoda]	'than expected'

The articles in the current issue examine spontaneous speech reduction in three West-Germanic languages (English, Dutch, and

[2] We use // to enclose phonetic transcription of the expected careful pronunciation, and [ ] for the variant as it was produced.

German) and one Romance language (French). Reduction has also been studied at least in Swedish (Engstrand & Krull, 2001), Finnish (Lennes et al., 2001), Greek (Nicolaidis, 2001), Mandarin (Cheng & Xu, 2009; Tseng, 2005), and Japanese (Maekawa & Kikuchi, 2005; Nakamura, Iwano, & Furui, 2007). These studies show that reduction occurs in many, typologically different languages, and that these languages may share some similar types of reduction (see also Barry & Andreeva, 2001; Simpson, 2001). For instance, Bürki, Fougeron, Gendrot, and Frauenfelder (in this issue) study schwa deletion in French, a phenomenon which is also well-known in English in certain phonological environments (Beckman, 1996). Likewise, the phenomenon of extreme reduction that Niebuhr and Kohler (in this issue) study in German is also familiar from spontaneous speech in other languages (e.g., Dutch, Ernestus, 2000), particularly in high-frequency function words. Other studies have shown, however, that even typologically related languages may substantially differ in their reduction patterns (Torreira & Ernestus, in press). Future research is necessary to establish how comparable the acoustic results and perceptual interpretation of reduction are across languages. This question is important in particular for psycholinguistic models of speech production and comprehension.

2.2. Phonetic gradience

The examples in (1–3) may give the impression that reduction is a categorical phenomenon, changing one phonological feature into another, or deleting complete segments. Whereas this may be true for some processes, such as word-medial schwa deletion in French (Bürki, Ernestus, & Frauenfelder, 2010), most reduction processes seem gradient rather than categorical. First, sounds may be very short and weakly articulated, but still be present. Second, reduced sections of speech often seem to contain clear cues to some phonological features (e.g., one perceives nasalization clearly, or perceives rhotacization somewhere in the word, or perceives non-back vocalic material), but one cannot definitively localize these features or identify them as segments. That is, one can identify some features of a relatively long stretch of speech, but one cannot confidently transcribe the speech into an ordered sequence of segments, even by making extensive use of IPA diacritic markers.

Fig. 1 exemplifies both of these properties of reduction (short segments and presence of features in the absence of identifiable segments). The left panel displays a token of *we were* with two audible /w/ segments and some r-coloring. The amplitude envelope of the first half of the sequence, along with changes in F2, reflects a relatively clear initial /w/. The slight drop in F2 at the arrow could reflect the second /w/, but the acoustic faintness of this second /w/ indicates gradient reduction of the segment. Further, in isolation the final vowel sounds high but not rhotacized, and there is no drop in F3. The right panel of Fig. 1 provides another such example, from a different speaker, with a very reduced token of *I haven't* (from *I haven't talked to her in, like, years!*). Absence of the final /t/ is not surprising, as the following word begins with a clearly articulated /t/. However, the sequence also shows great reduction of the vowel space, and absence of the vowel of an expected stressed syllable (*have*). The rather long medial consonant is perceptibly labial, and its length may cue the otherwise absent /hæ/ by allowing the listener to infer that some segments are absent or merged into this long consonant. Again, some features are perceptible (labiality, non-low vowels), but for many segments, it is difficult to state definitively that they are either present or absent. (The IPA transcription is provided for convenience, and is not a claim that particular segments were produced.)

Since reduction can be gradient, it spans a wide range, from subtle weakenings, such as the absence of a burst on an otherwise clear stop (e.g., Mitterer, in this issue), to the absence of multiple syllables, as in some examples in (1–3) above or the reduction investigated by Niebuhr and Kohler (in this issue).

2.3. Intelligibility of reduced pronunciation variants

What is perhaps most surprising about extreme reductions is that they are usually quite intelligible, if only to native listeners of approximately the speaker's dialect with normal hearing, and when heard in context. In most cases, native listeners hearing recordings of discourses containing pronunciation variants like examples (1)–(3) and Fig. 1 agree without trouble on what words the speaker said.

Notably, though, if even one of the conditions listed above is not true, for example if the listener does not speak the same dialect or the word is presented in isolation, listeners may not be able to identify the words accurately at all. Audiences of native listeners reliably misperceive many tokens of *we were* as *we're* out of context, but hear *we were* in context. They also identify *he already* pronounced as [iɹɾɪ] as "maybe something like 'dirty' but not really" out of context, and often fail to recognize any English words at all in *but I was like* pronounced as [b̥ɹʌʒlə̥ɪ] out of context. Research on Dutch has shown that in order to recognize highly reduced pronunciations, listeners need context that contains more than just the surrounding vowels and any intervening consonants,

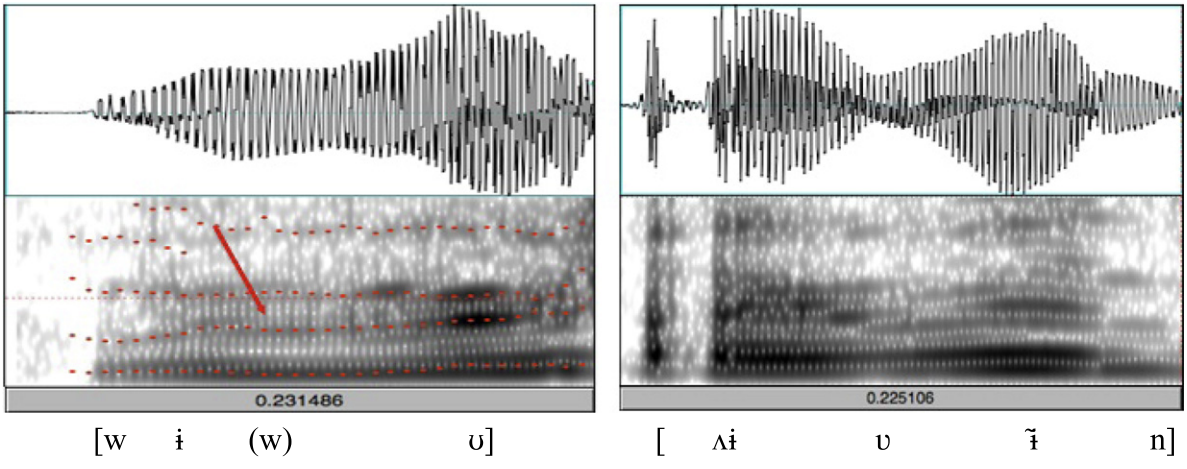

Fig. 1. Waveforms and spectrograms, with formant overlay, of *we were* (left panel) and *I haven't* (right panel) from a casual conversation. The arrow in the left panel indicates a slight dip in F2 that may reflect a second /w/.

which form cues to speech rate and co-articulation (Ernestus et al., 2002). This is especially true for older listeners, who are hard of hearing (Janse & Ernestus, in press).

Nevertheless, there are examples of reduced pronunciation variants for which even native listeners of the dialect, hearing the words in their full discourse context, cannot agree on the words' identity. The meaning of the utterance as a whole is completely clear, but natives cannot reach agreement on the specific words contained. For example, the second author of this article perceived [bɹʌʒləɪ] (see above) as *but I was like*, while another native speaker provided the orthographic transcription *but I was just like*, meaning that it would contain even more reduction than the researcher had supposed, with absence of three syllables instead of two. The presence of some unintelligible words in spontaneous speech is not restricted to English. During the workshop from which this volume stems, many presenters played examples of reduced speech from a variety of languages. In several cases, the native listeners in the audience could not be sure exactly what words were included in the speech, even on hearing it in context multiple times, although they understood the meaning of the overall utterance. Importantly, the words on which listeners cannot reach consensus usually carry little or no distinctive meaning (like *just* in the example above) and in natural conversations are seldom followed by questions from the listeners about their meaning. Apparently, as native listeners, we do not need to recover all a speaker's intended words, although we do successfully recover most of them. When listening to conversation, we probably do not even notice when we have just failed to recover several words in a reduced stretch of speech. This is a testimony to the role of redundancy in speech.

3. Relevance of research on reduced pronunciation variants

3.1. Theoretical relevance

Since speakers produce and listeners comprehend reduced pronunciation variants, the processing of these variants must be accounted for by psycholinguistic models of speech comprehension and production. The existing models were developed mostly from evidence about the processing of words produced carefully in isolation or in simple sentences in the laboratory. It is not self-evident that these models can also account for the processing of reduced pronunciation variants (for a more detailed overview, see Ernestus, in press; Warner, 2011).

The existing psycholinguistic models form a continuum with purely abstractionist models on one end. These assume that pronunciations are stored in the mental lexicon in the form of abstract representations (e.g., as sequences of phonemes) and that the number of lexically represented pronunciations is limited (in the most extreme case to the word's full pronunciation). In order to account for the production of reduced pronunciation variants, these models have to assume mechanisms, for instance phonetic implementation rules, that convert full pronunciations into reduced variants during speaking. Comprehension models have to reconstruct the word's full pronunciation on the basis of, for instance, subtle acoustic cues in the context or in the variant itself (e.g., the precise pronunciation of the English consonant cluster /sp/ may contain cues to an absent schwa, as in the pronunciation [spɔrt] for *support;* Davidson, 2006), or on the basis of phonotactic contraints (e.g., since the cluster /fk/ does not occur in carefully pronounced words in Dutch, it cues a missing vowel, as in the pronunciation [fkopə] for /vərkopə/ 'to sell'). These models need detailed information about the production and comprehension of reduced variants, such that they can further specify the production and reconstruction mechanisms.

The other end of the continuum is occupied by pure exemplar models that assume that all tokens of a word ever produced or perceived by the language user are stored in the mental lexicon with all their acoustic details (e.g., Goldinger, 1998). The production of a word involves the activation of one particular exemplar or a group of exemplars representing the word, while comprehension involves the mapping of the acoustic signal onto one or several of its exemplars. These exemplars may be reduced, which allows modeling of how speakers produce and listeners understand reduced pronunciation variants. Further, it has been proposed that the exemplars could contain information about the context in which the tokens occurred, which would explain why reduced variants cannot be recognized well out of their linguistic contexts (Hawkins, 2003; Hawkins & Smith, 2001). So far, it is not completely clear how exemplar models should account for the empirical finding that the full form appears to play a special role in speech production and comprehension (e.g., Bürki et al., 2010; Ranbom & Connine, 2007; Tucker, 2007). Moreover, many studies have shown that speech planning is an important predictor for degree of reduction, which implies that speech planning influences the choice of the exemplar or that exemplar models have to assume that production mechanisms adapt the selected pronunciation, as in abstractionist models.

The middle of the theoretical continuum hosts so-called "hybrid models", which assume that a word's pronunciation is stored in the mental lexicon both in the form of abstract representations and exemplars. In these models it is especially important to determine whether variation resulting from reduction is stored in abstract representations, and whether this type of variation has a different status than, for instance, variation resulting from characteristics of the speaker's voice.

In summary, wherever a theory is located on the theoretical continuum, it needs further specification of the production and comprehension mechanisms involved in the processing of reduced pronunciation variants. This specification cannot easily be obtained by means of research on the processing of just the words' full forms. From a theoretical point of view, studies on the human processing of reduced speech can therefore make great contributions.

3.2. Societal relevance

During recent decades, speech technology has made great progress in the recognition (that is, in the automatic generation of orthographic transcriptions) of read speech. Automatic Speech Recognition (ASR) systems now produce transcriptions for careful read speech with only very few errors. The situation is completely different, in contrast, for reduced spontaneous speech. For instance, while ASR systems have an error rate of less than 10% for TIMIT (e.g., Nakamura, Iwano, & Furui, 2008), a corpus of read speech, their error rates are as high as 30–50% for Switchboard (e.g., Novotney & Callison-Burch, 2010), a corpus consisting of spontaneous telephone conversations.

Nearly all ASR systems assume a lexicon which contains only one pronunciation variant for every word, which is typically the word's full pronunciation. This pronunciation is stored in the form of a sequence of phones. During recognition, the ASR system selects the sequence that best fits the acoustic signal, using acoustic phone models. Reduced pronunciation variants are problematic in ASR systems of this type since they do not map well onto the phone representations of their full forms. For instance, Dutch [tyk] does not match well with the full pronunciation of the word *natuurlijk* /natyrlək/ 'of course'. One obvious solution to this problem is to incorporate reduced pronunciation variants in the system's lexicon. For this, detailed phonetic

research would be necessary to describe the possible pronunciation variants of all words in the lexicon.

Just the incorporation of reduced variants is not sufficient, however. This increases the confusability in the system for unreduced words and consequently the error rate for these words. For instance, if we incorporate /tyk/ as a possible pronunciation of Dutch *natuurlijk*, the system is more likely to produce an error for the Dutch low frequency word *tule* /tylə/ 'tulle', since the system then has to choose between the words /tyk/ and /tylə/. The incorporation of pronunciation variants in the ASR lexicon can only decrease word error rate if the lexical representation for a pronunciation variant is not only specified for its frequency of occurrence, but also for the context in which it is likely to occur, which is a topic for future research. Research on how humans process reduced pronunciation variants may also provide new information that could improve ASR systems.

Reduced pronunciation variants also form serious problems for foreign language listeners. For instance, if an English speaker says [wɨ (w) ʊ ɸɨsɨ siːjːɛʃɐɪ b ɐɪ fɛt̪ ɻɨli bæd] for *we were supposed to see it yesterday, but I felt really bad*, foreign language listeners, even those who use English on a regular basis, typically have no clue what the speaker means by [ɸɨsɨ siːjːɛʃɐɪ] (and may not understand other parts, either). During their English courses, they did not learn that *supposed to* may be reduced to [ɸɨsɨ] and *yesterday* to [jɛʃɐɪ], or that *it* can be realized just by lengthening of surrounding segments. Importantly, it is also very difficult for foreign language users to find out by themselves which words the different pronunciation variants represent. They cannot look these variants up in the dictionary (e.g., *yeshay* is not incorporated in any regular dictionary of English) and they cannot ask for help from native speakers, since native speakers are not aware of most variants in their language.

Future research on reduced pronunciation variants will show which reduced pronunciation variants occur in a language and to what extent the occurrence of specific variants is principled and predictable. This information can be incorporated in instruction methods for foreign language learners, which will increase the naturalness of the speech produced by these users and their ability to understand casual speech. Furthermore, research on native listeners' comprehension of reduced pronunciation variants will show which cues they use above all to understand these variants, and this information may help foreign listeners to learn to understand these variants as well.

4. Research methods

4.1. *The use of speech corpora* 语料库研究法

Six of the research articles in this special issue are to some extent based on corpora of spontaneous speech. Corpus-based studies typically investigate the characteristics of reduced pronunciation variants on the basis of phonemic or more detailed phonetic transcriptions of speech in the corpus. Such transcriptions are, however, difficult to obtain. Transcribers often disagree with each other (disagreements on about 15% of the tokens are no exception, see, e.g., Coussé, Gillis, Kloots, & Swerts, 2004; Ernestus, 2000; Pitt, Johnson, Hume, Kiesling, & Raymond, 2005) and one and the same transcriber may make different decisions for the same stretch of speech at different times. The low reliability of phonetic transcriptions results from the gradient nature of reduction, from the influence of listeners' expectations on what they think they hear, and from the inherently categorical nature of phonetic transcription (e.g., Cucchiarini, 1993; Keating, 1998; Vieregge, 1987). Plug's research method in the present volume minimized the role of the transcriber's expectations based on knowledge of the language. He studied a corpus in his native language and first transcribed all relevant utterances himself. He then provided a non-native speaker with the canonical transcriptions of the utterances and asked this transcriber to adjust these transcriptions. He agreed with his non-native co-transcriber on the number of absent or non-segmentally realised canonical segments for 71% of the utterances. Reinspection of the utterances on which the transcribers disagreed increased the percentage of agreements to 83%. For the remaining 17% of problematic utterances (8 tokens), Plug (in this issue) used the transcriptions provided by the non-native speaker, since this transcriber was less likely to be affected by expectations based on knowledge of the language or of the topic of the study.

Since the manual transcription of spontaneous speech is highly time consuming and error prone, several studies use phonemic transcriptions generated by Automatic Speech Recognition systems (see, e.g., Adda-Decker & Snoeren, in this issue; Bürki et al., in this issue; Meunier & Espesser, in this issue; all in this volume). These ASR systems typically have as their inputs the orthographic transcription of the utterance, a pronunciation dictionary in which the orthographic transcription of each word is related to one or more phonemic transcriptions representing the word's pronunciation (variants), and phone models, which relate the phonemes in the phonemic transcriptions to the acoustic signal. The ASR model determines the possible phonemic transcriptions of an utterance given the orthographic transcription, and then chooses the transcription that fits the acoustic signal best, using the phone models. These automatically generated transcriptions have the advantage of being consistent and they typically deviate from transcriptions generated by humans as much as humans tend to differ from each other, or only slightly more (see, e.g., Schuppler et al., 2011). They therefore appear to be a viable substitute for manual transcriptions.

ASR transcription systems, however, also have several limitations. First, they can only recognize those pronunciation variants that are incorporated in their lexicons. They are therefore especially useful for the classification of tokens as known variants (as, e.g., in Bürki et al., in this issue), but not for the detection of new variants. Second, ASR systems can only recognize phonemes with a certain minimal duration, such as 24 ms (Meunier & Espesser, in this issue) or 30 ms (Adda-Decker & Snoeren, in this issue). As a consequence, short segments, which may be abundant in spontaneous speech, tend to be incorrectly classified as absent. Third, ASR transcription systems produce errors, like humans do. In general, ASR transcription systems have difficulties transcribing those sounds that are also notoriously difficult for human transcribers (e.g., presence of schwa between liquids, voicing in obstruents). This is unsurprising, because the machines base themselves on the same acoustic signal as humans do. Moreover, the ASR's phone models have to be trained on phonemic transcriptions created by humans, which contain errors especially for those segments that are difficult to transcribe for humans. Finally, ASR systems are likely to contain errors when acoustic features are clear, but their association to segments is not, as in Fig. 1 (right), where labialization and nasalization are clearly present, but it is not clear what segments those might be associated with. Human transcribers may have the same problem unless a large inventory of narrow transcription symbols, including many diacritics, is used, since the sounds in such reductions often do not sound like any regular segment of the language.

In order to overcome these restrictions, while still profiting from the advantages of ASR transcription systems (being fast and consistent), Bürki et al. (in this issue) had a phonetician check all transcriptions made by their ASR system. Moreover, 50% of these transcriptions were also checked by a second phonetician, who showed high agreement with the first transcriber. This double

checking suggests that their transcription data are highly reliable and do not suffer from the restrictions typical for ASR transcriptions.

4.2. Experimental approaches 实验研究法

Corpus studies cannot answer all questions about the production of reduced pronunciation variants, and they provide little information about speech comprehension. Research on reduction therefore also requires experimental approaches, and these are also well represented in this volume (as in the articles by Mitterer, Niebuhr & Kohler, by Tucker, by Pitt, Dilley, & Tat, and by Bürki and colleagues).

The articles in this volume reporting perception experiments make use of various experimental paradigms. Bürki and colleagues directly asked participants whether a stretch of speech contained a schwa. Mitterer investigated listeners' interpretation of a single speech segment as a function of the surrounding segments by means of a four interval oddity task (in which listeners hear four variants of a word and have to indicate which of them differs). Niebuhr and Kohler, who also studied the interpretation of segments, asked participants to perform a dictation task and to match questions with responses that contained reduced stretches of speech. Finally, Tucker (in this issue) as well as Pitt and colleagues conducted lexical decision tasks in order to determine on the basis of reaction times whether recognition is influenced by the type of reduction, the word's frequency of occurrence, and the frequency of the reduction type.

The choice of experimental paradigm is determined not only by the questions the researchers wish to investigate, but also by the type of speech process under investigation. The comprehension of speech processes that occur in words uttered in isolation can be investigated well with paradigms presenting words in isolation or in short phrases. This is for instance the case for place assimilation, one of the phenomena investigated by Mitterer with the four interval oddity task, in which he presented words in isolation. It is also the case for reduction of /t/, also investigated by Mitterer with the four interval oddity task and by Pitt and colleagues with the lexical decision task. Other phenomena only occur in connected speech and in order to obtain ecologically valid data, researchers have to use paradigms presenting participants with longer stretches of speech, as was done by Niebuhr and Kohler.

In order to produce data that are ecologically valid, researchers may strive to present listeners with (speech very similar to) spontaneous speech. The use of this type of speech, however, has a serious drawback since stretches of spontaneous speech vary greatly. Their use results in less well-controlled experiments, which is problematic for experimental paradigms for which absolute control over the stimuli is desirable (e.g., the four interval oddity task, see Warner (to appear), for a more extensive discussion). As a consequence, researchers use the entire continuum from carefully pronounced words to real spontaneous speech.

This is well exemplified by the perception experiments described in this volume. The methods employed by Mitterer and by Niebuhr and Kohler represent one end of a continuum. These authors gained insight from spontaneous speech corpora or general recordings of spontaneous speech in order to decide what phenomena to test in their perception study. They then recorded or created acoustic stimuli specifically for the experiment, relatively independently of what was found in the corpora. This resulted in highly controlled stimuli. Tucker took a more intermediate approach to how to combine actual recordings of spontaneous speech with experimental design. Like Mitterer and like Niehbur and Kohler, Tucker recorded words specifically for the use as stimuli, but his stimuli represent more variation than the stimuli in the experiments conducted by Mitterer and by Niehbur and Kohler. Moreover, he directly related his stimuli to words produced in spontaneous speech through a series of comparisons. Pitt and colleagues used stimuli in their perception experiment that were produced by naive speakers. They tested stimuli obtained in a production experiment in which speakers produced relatively natural and variable sentences with target words in an informal, but not spontaneous, speech style. Finally, Bürki and colleagues used stimuli representing the other end of the continuum, drawing their perception task stimuli directly from a speech corpus.

In order to produce ecologically valid data, perception experiments would ideally present listeners with a high number of lexical items. However, this is not always possible. For instance, Niebuhr and Kohler investigated which characteristics of the speech signal lead listeners to interpret a stretch of speech as a two-word phrase (*eine Rote* 'one red one') or as a highly reduced version of a three-word phrase (*eigentlich eine Rote* 'actually a red one'). Such confusable pairs do occur in spontaneous speech, but pairs that are truly homophonous under reduction, where both members of the pair are semantically plausible, are not common. Their experiments are consequently restricted to a small number of lexical items (one in fact, manipulated to create a range of stimuli). In contrast, Tucker, Pitt and colleagues, and Bürki and colleagues all use numbers of items on the order typically used in psycholinguistic experiments that are not about reduction (for example, approximately 70 items distributed among several conditions). This is possible, since the constraints on their lexical items are no more limiting than those in psycholinguistic experiments not on reduction (e.g., in Tucker's experiment, bisyllabic words with initial stress containing an intervocalic /g/). Obviously, a larger number of lexical items increases the generalizability of the results to other lexical items in the language.

5. Brief introduction to the eight research articles

This special issue starts with four articles focusing on the acoustic characteristics of reduced pronunciation variants. Three of them studied French and thus provide insights into the differences and similarities between French and well studied Germanic languages. All four studies are based on speech corpora (rather than on production or perception experiments).

Adda-Decker and Snoeren studied reduction in French in order to improve ASR systems. They analyzed the durations of vowels and consonants in different speech styles, as determined by an ASR system. They observed more durational variation in more spontaneous speech registers. Especially in casual speech, many segments were assigned the shortest possible duration, while other segments still received relatively long durations. The segments assigned very short durations are likely not to have been realized by the speakers at all, since the system assigns some duration to each segment, even if it is in fact absent. Future research on the characteristics of reduced pronunciation variants could focus on sequences of segments assigned very short durations by the ASR system.

Meunier and Espesser investigated the durations and spectral properties of vowels in many different types of words, as measured by an ASR system. The word tokens were extracted from a corpus of spontaneous conversations held by 16 speakers from the South-East of France. Replicating results on English and Dutch (Van Bergem, 1993), Meunier and Espesser found more durational and spectral reduction in word-internal than in non-word-internal syllables and in function than in content words. Moreover, they documented a strong correlation between durational and spectral reduction. This contribution by Meunier and Espesser is valuable also because they discuss the disadvantages of studies based on spontaneous speech corpora. In such studies,

the researcher cannot control for all properties of the words under investigation.

Bürki and colleagues focused on one particular vowel in French, word-internal schwa. They first studied the presence and duration of schwa in 3098 word tokens in radio broad-casted speech. Like Meunier and Espesser, they showed that French reduction patterns may be very similar to those of Germanic languages: schwa showed not only categorical reduction, which seems to be specific to French, but also gradient reduction. Bürki and colleagues then investigated which factors influence whether a listener perceived a schwa as present or absent. Their listeners appeared to base their judgments not only on the duration of the schwa present, but also on their expectations of its duration given the speech rate and word length, and on the possibility of assigning the vowel interval to the surrounding consonants.

Plug investigated self-initiated self-repair in Dutch. He measured the speech rate and segmental reduction in the reparandum (the error that is corrected) and in the repair stretch, distinguishing between editing terms (e.g., *I mean*, *actually*), repeated items, and the words providing the new information in the repair stretch. Although the new information in repair stretches was sufficiently important for speakers to interrupt the speech flow, Plug showed that these stretches were pronounced at higher speech rates and with more segmental reduction than the reparanda. This was the case for all parts of the repair stretch, as well as for stretches repairing a lexical, syntactic, or pronunciation error and stretches repairing inappropriate words. Plug discusses the possible explanation that speakers reduce repair stretches in order to continue as quickly as possible with the discourse. Apparently, the information carried by a word influences its reduction degree, as observed in many other studies, in interaction with pragmatics.

The remaining four articles in this volume examine how listeners perceive reduced pronunciation variants, often combining perceptual data with acoustic study in order to determine the relation between what listeners perceive and what they are hearing. These articles use a variety of methods of creating or obtaining stimuli, and a variety of tasks to measure listeners' perception. They represent work on English, German, and Dutch.

Mitterer studied two reduction phenomena in Dutch (assimilation of the place of articulation of a nasal to a following consonant, as in *tui*[m]*ban*k for *tuinbank* 'garden bench'; and deletion of the burst of a cluster /t/ as in *kus*(*t*) 'coast'). Although he studied these using Dutch stimuli, these phenomena are both crosslinguistically common. Mitterer used a four-oddball task (or four interval oddity task), testing native listeners' ability to distinguish a reduced token from a clear token of the same string of phonemes. Mitterer found differences in how segmental context influences listeners' perception of the two reduction phenomena. Combining this with previous data, he concludes that not all types of reduction are processed using the same mechanisms. This article provides an example of studying perception of reduced speech in very similar ways to how careful speech perception is usually studied, and with great control. To achieve this, it uses very specific types of reduction, a small number of items, and resynthesis of stimuli.

Pitt, Dilley, and Tat (in this issue) investigated production and perception of the American English phoneme /t/ as a flap, glottal stop, canonical [t], or nothing (deletion). They first determined which variants were most common in which environments, by means of a production experiment, and then compared these frequencies to results they obtained from a perception experiment, testing the same variants. They used a lexical decision task, thus recruiting standard psycholinguistic methods for the study of speech reduction, which is possible since they focused on a reduction phenomenon that can be observed within a single word. They found that the variants that are produced more commonly are also recognized more easily, but that words containing the canonical variant [t] are recognized more easily than would be expected, even when [t] is not a common variant in production.

Tucker also investigated the comprehension of reduced pronunciation variants in isolation, focusing on intervocalic flaps and /g/ in American English. He elicited pronunciation variants of a high number of words from a trained phonetician, and showed that both the unreduced and reduced variants are very similar to tokens that occur in spontaneous speech, in the duration of the consonant, its intensity relative to the intensities of the neighboring vowels, and the presence of formant structure. Like Pitt, Dilley, and Tat, he tested listeners' comprehension in a lexical decision experiment. If the effect of reduction was tested by means of a two-level factor (unreduced versus reduced), the unreduced variants appeared to have a processing advantage. If reduction was captured by a continuous variable based on the intensity of the consonant, listeners appeared to recognize especially well variants of a medium reduction degree. This result demonstrates the relevance of investigating more detailed reduction variables than simple two-level factors.

Finally, Niebuhr and Kohler studied the perception of an extreme reduction leading to a pronunciation of German *eigentlich eine Rote* 'actually a red one' homophonous with the full pronunciation of *eine Rote* 'a red one'. Because of the difficulty of studying such extreme reductions in a controlled way, this article used a three-fold methodological approach. All these methods used pairs of stimuli with the same total duration, but with different durations of the initial vowel and the palatalized portion of *eine*. Two methods were modifications of standard phonetic tasks used for studying perception of careful speech, while the third task (a matching task, asking listeners whether a stimulus is a good answer to the question *How many do you want?,* which can only logically be followed by the answer *eine Rote*) represents a larger methodological innovation for studying reduced speech. The authors found that listeners are more likely to interpret the stimuli as highly reduced variants of *eigentlich eine Rote* when the palatal portion of the stimulus is longer.

6. Conclusion

In this article, we have introduced the phenomenon of reduced pronunciation variants and presented an overview of the eight research articles in this volume. The workshop from which this volume stems brought together researchers using a wide variety of methods to study many reduction phenomena in many languages. This allows the field to reach broad conclusions about what kind of acoustic information humans produce in reduced speech, and how listeners use that information to retrieve the speaker's message, if not all of the speaker's intended segments or words.

References

Adda-Decker, M., & Snoeren, N. D. Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics,* in this issue.

Barry, W., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association, 31*, 51–66.

Beckman, M. E. (1996). When is a syllable not a syllable?. In T. Otake, & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 95–123). Berlin: Mouton de Gruyter.

Van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, *12*, 1–23.

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, *62*, 421–437.

Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics,* in this issue.
Canavan, A., & Zipperlen, G. (1996). *CALLHOME Japanese speech*. Philadelphia: Linguistic Data Consortium.
Cheng, C., & Xu, Y. (2009). Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. In: *Proceedings of Interspeech 2009* (pp. 456–459). Brighton.
Coussé, E., Gillis, S., Kloots, H., & Swerts, M. (2004). The influence of the labeller's regional background on phonetic transcriptions: Implications for the evaluation of spoken language resources. In M. Lino, M. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the fourth international conference on language resources and evaluation* (pp. 59–62). Paris: European Language Resource Association.
Cucchiarini, C. (1993). *Phonetic transcription: A methodological and empirical study*. Dissertation. Nijmegen: Catholic University Nijmegen.
Dalby, J. M. (1986). *Phonetic structure of fast speech in American English*. Dissertation. Bloomington: Indiana University Linguistic Club.
Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap?. *Phonetica*, *63*, 79–112.
Engstrand, O., & Krull, D. (2001). Segment and syllable reduction: preliminary observations. *Working Papers Lund University, Departement of Linguistics*, *49*, 26–29.
Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology–phonetics interface*. Dissertation. Utrecht: LOT.
Ernestus, M. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, in press.
Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.
Greenberg, S. (1999). Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176.
Guy, G. R. (1991). Contextual conditioning in variable lexical phonoloy. *Language Variation and Change*, *3*, 223–239.
Guy, G. R. (1992). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, *3*, 1–32.
Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405.
Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics-Rivista di Linguistica*, *13*, 99–188.
Janse, E., & Ernestus, M. The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, in press.
Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama, & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis* (pp. 29–54). Tokyo: The National International Institute for Japanese Language.
Keating, P. A. (1998). Word-level phonetic variation in large speech corpora. In A. Alexiadou, N. Fuhrop, U. Kleinhenz, & P. Law (Eds.), *ZAS papers in linguistics 11*. Berlin: Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung.
Keune, K., Ernestus, M., Van Hout, R., & Baayen, R. H. (2005). Social, geographical, and register variation in Dutch: From written "mogelijk" to spoken "mok". *Corpus Linguistics and Linguistic Theory*, *1*, 183–223.
Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 69–92). Dordrecht: Kluwer Academic Publishers.
Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
Lennes, M., Alarotu, N., & Vainio, M. (2001). Is the phonetic quality of unaccented words unpredictable? An example from spontaneous Finnish. *Journal of the International Phonetic Association*, *31*, 127–138.
Maekawa, K., & Kikuchi, H. (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In J. van de Weijer, K. Nanjo, & T. Nishihara (Eds.), *Voicing in Japanese* (pp. 205–228). Berlin: Mouton de Gruyter.
Meunier, C., & Espesser, R. Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics,* in this issue.
Mitterer, H. Recognizing reduced forms: Different processing mechanisms for similar reductions. *Journal of Phonetics,* in this issue.
Nakamura, M., Iwano, K., & Furui, S. (2007). The effect of spectral space reduction in spontaneous speech on recognition performances. In *Proceedings of the 32nd international conference on acoustics, speech, and signal processing* (pp. 473–476). Honolulu.
Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech and Language*, *22*, 171–184.
Niebuhr, O., & Kohler, K. J. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics,* in this issue.
Nicolaidis, K. (2001). An electropalatographic study of Greek spontaneous speech. *Journal of the International Phonetic Association*, *31*, 67–85.
Novotney, S., & Callison-Burch, C. (2010). Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *Human Language Technologies: The 2010 annual conference of the North American chapter of the ACL* (pp. 207–215). Los Angeles.
Pitt, M. A., Dilley, L., & Tat, M. Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, in this issue.
Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*, 89–95.
Plug, L. Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of Phonetics,* in this issue.
Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.
Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*, 96–109.
Shattuck-Hufnagel, S., & Veilleux, N. (2007). Robustness of acoustic landmarks in spontaneously-spoken American English. In *Proceedings of the 16th international congress of phonetic sciences ICPhS-07* (pp. 925–928). Saarbrucken.
Simpson, A. P. (2001). Does articulatory reduction miss more patterns than it accounts for?. *Journal of the International Phonetic Association*, *31*, 29–41.
Torreira, F., & Ernestus, M. (2011). Vowel elision in casual French: The case of vowel /e/ in the word c'était. *Journal of Phonetics*, *39*, 50–58.
Torreira, F., & Ernestus, M. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, in press.
Tseng, S. C. (2005). Syllable contractions in a Mandarin Conversational Dialogue Corpus. *International Journal of Corpus Linguistics*, *10*, 63–83.
Tucker, B. V., 2007. *Spoken word recognition of the reduced American English flap*. Dissertation. Tucson: University of Arizona.
Tucker, B. V. The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics,* in this issue.
Vieregge, W. H. (1987). Basic aspects of phonetic segmental transcription. In A. Almeida, & A. Braun (Eds.), *Probleme der Phonetischen Transkription* (pp. 5–55). Stuttgart: Franz Steiner Verlag Wiesbaden GmbH.
Warner, N. (2011). Reduction. Invited chapter. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Malden, MA & Oxford: Wiley-Blackwell.
Warner, N. Methods for studying spontaneous speech. Invited chapter in A. Cohn, C. Fougeron, & M. Huffman (Eds.), *Handbook of laboratory phonology*, to appear.

Mirjam Ernestus *
Center for Language Studies, Radboud University Nijmegen & Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
E-mail address: m.ernestus@let.ru.nl

Natasha Warner
Department of Linguistics, University of Arizona, Box 210028, Tucson, AZ 85721-0028, USA

* Corresponding author. Tel.: +31 24 3612970; fax: +31 24 3521213.