Available online at www.sciencedirect.com

SciVerse ScienceDirect

Lingua 142 (2014) 27–41

Lingua

www.elsevier.com/locate/lingua

Acoustic reduction and the roles of abstractions and exemplars in speech processing

Mirjam Ernestus

Radboud University Nijmegen & Max-Planck Institute for Psycholinguistics, Netherlands

Received 16 May 2010; received in revised form 14 December 2012; accepted 15 December 2012
Available online 3 July 2013

Abstract

Acoustic reduction refers to the frequent phenomenon in conversational speech that words are produced with fewer or lenited segments compared to their citation forms. The few published studies on the production and comprehension of acoustic reduction have important implications for the debate on the relevance of abstractions and exemplars in speech processing. This article discusses these implications. It first briefly introduces the key assumptions of simple abstractionist and simple exemplar-based models. It then discusses the literature on acoustic reduction and draws the conclusion that both types of models need to be extended to explain all findings. The ultimate model should allow for the storage of different pronunciation variants, but also reserve an important role for phonetic implementation. Furthermore, the recognition of a highly reduced pronunciation variant requires top down information and leads to activation of the corresponding unreduced variant, the variant that reaches listeners' consciousness. These findings are best accounted for in hybrids models, assuming both abstract representations and exemplars. None of the hybrid models formulated so far can account for all data on reduced speech and we need further research for obtaining detailed insight into how speakers produce and listeners comprehend reduced speech.
© 2013 Elsevier B.V. All rights reserved.

Keywords: Acoustic reduction; Exemplar-based models; Abstractionist models; Speech comprehension; Speech production

1. Introduction

In spontaneous speech, words are often produced with fewer segments than in careful speech. For instance, the English word *ordinary* may be pronounced as *onry*, *computer* simply as *puter*, and *yesterday* as *yeshay*. This phenomenon of acoustic reduction has received very little attention in the linguistic and psycholinguistic literature so far. Nevertheless, like all other language phenomena, it is important for the formulation of linguistic and psycholinguistic theories, since these theories need to account for all linguistic behavior. This article discusses data obtained from corpus-based research and psycholinguistic experimentation on the production and comprehension of reduced speech and how they present challenges for the two types of theoretical frameworks that are nowadays popular in both linguistics and psycholinguistics: abstractionist models and exemplar-based models. It thus contributes to the debate on the relevance of exemplars and abstract representations and generalizations in language processing.

This article consists of 5 parts. The remainder of this section introduces the phenomenon of acoustic reduction in more detail (section 1.1) as well as the central ideas of simple abstractionist and exemplar-based models that are discussed in this article and the accounts these models offer for the phenomenon of acoustic reduction (sections 1.2 and 1.3). In the two following sections, I summarize research on the production (section 2) and comprehension (section 3) of acoustic

E-mail address: mirjam.ernestus@mpi.nl.

0024-3841/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.lingua.2012.12.006

reduction and the implications of the findings for the two theoretical frameworks. Both frameworks contain characteristics that appear necessary to account for the processing of reduced speech. I therefore then discuss some hybrid models that have been proposed in the literature and how they could account for the data on acoustic reduction (section 4). Finally, I briefly summarize the most important conclusions (section 5). Although this article appeared in 2013, it was finalized in August 2010 and therefore does not take into account publications that were unknown to me by that time.

1.1. Acoustic reduction

Acoustic reduction is an important characteristic of everyday speech that affects a high percentage of the word tokens in spontaneous conversations. For instance, Dalby (1984) showed that during news interviews on television some speakers of American English delete unstressed vowels in 25% of their word tokens, and Johnson (2004) reported that in the Buckeye Corpus of Spontaneous American English (Pitt et al., 2005), no less than 6% of the word tokens miss a complete syllable. Acoustic reduction is highly frequent also in other Germanic languages, such as Dutch (e.g., Ernestus, 2000) and German (e.g., Kohler, 1990), and it occurs in non-Germanic languages as well, such as French (e.g., Adda-Decker et al., 2005) and Finnish (Lennes et al., 2001). The number of studies on reduction in non-Germanic languages is, however, minimal.

Some reduction patterns that can be observed in spontaneous speech may be explained by the speaker's tendency to reduce articulatory effort when possible (Lindblom, 1990). These patterns may result from reduction in the sizes of the articulatory gestures and from time overlap of these gestures (Pagliuca and Mowrey, 1987; Browman and Goldstein, 1990). For instance, if speakers intend to say *perfect memory*, but close their lips for the production of the following *m* before the release of the word-final *t* (possibly in combination with reducing the size of the closing gesture for this *t*), the *t* may be inaudible. Hence, segments may be missing in the acoustic signal, but nevertheless may have been produced by the speaker (hence the term *acoustic* reduction).

Acoustic reduction leads to an enormously high number of possible pronunciation variants of especially long words. For instance, the possible variants for the Dutch word *natuurlijk* 'natural(ly)', especially when used as a discourse marker (comparable to English 'of course'), include (but are not restricted to) [natyrlək] (citation form), [natylək], [ntylək], [ntyk], [tyrlək] (which also occurs in written language, as *tuurlijk*), [tylək], [tylk], [tyk], [tyg], [dyk], and [dyg] (Ernestus, 2000). In fact, in spontaneous speech more than one pronunciation variant for a word is rather the rule than the exception. Hence, the phenomenon of acoustic reduction asks for a principled account of how to deal with pronunciation variation in linguistic and psycholinguistic theories.

1.2. Abstractionist models

The abstractionist framework is one of the most popular frameworks within linguistics. It has received explicit formulation especially by Chomsky and Halle (1968) and other linguists working within the generative framework (including Optimality Theory, e.g., Prince and Smolensky, 2004), and has formed a source of inspiration for many psycholinguistic models of speech production and comprehension (e.g., Levelt, 1989; Norris, 1994). In this article I will discuss the most basic and simple version of abstractionist models which has as its key assumption that the mental lexicon contains only one lexical representation for every word (or morpheme), consisting of a string of abstract symbols, for instance, phonemes. Thus, there is only one lexical representation for *yesterday*, even though it is sometimes produced as *yeshay* in conversational speech. Pronunciation variants are derived from this single abstract representation by means of general processes, which apply to several words. In the following, the term abstractionist models will always refer to models based on this simple key assumption. I will thus ignore the many variants of abstractionist models that allow for the storage of pronunciation variation and phonetic detail. In this way, it is easier to focus on the merits of the key assumption of a sparse lexicon supplemented with abstract generalizations.

Within abstractionist (psycholinguistic) models, the production of a word involves the activation of its lexical representation and the application of phonological rules (or constraints in the framework of Optimality Theory), adapting the word to its phonological context (e.g., final devoicing, place assimilation). The resulting abstract phonological representation is translated by phonetic mechanisms (again, rules or constraints) into articulatory/acoustic events. Together phonological and phonetic mechanisms take care that, for instance, word-final /n/s are produced as [m]s before /b/-initial words (place assimilation: e.g., *garden bench* is pronounced as *garde*[m b]*ench*) and that word-final syllables are lengthened when followed by prosodic phrase boundaries (prosodic lengthening).

Abstractionist models may account for the production of acoustic reduction by means of both the phonological and the phonetic component. Since acoustic reduction appears optional, also the mechanisms in these components have to be optional. Optional phonological rules may have their application frequency specified in their structural descriptions (see e.g., Cedergren and Sankoff, 1974), while in Optimality Theory, the optionality of phonological processes can be accounted for by assuming that speakers may choose among different constraint rankings and that the frequency of a

given pronunciation variant is the direct result of the number of rankings having that variant as the optimal output (see e.g., Antilla and Cho, 1998; Boersma, 1998). The variation resulting from phonetic mechanisms may be controlled for by parameters specifying, among others, the speaker's articulatory effort.

Comprehension within abstractionist models consists of the mapping of the acoustic input onto the corresponding string of abstract symbols in the mental lexicon. This mapping entails speaker-normalization: The comprehension process very early on abstracts away from all characteristics of the acoustic input that are speaker or situation specific (e.g., pitch and loudness of the voice, exact quality of the vowel) in order to produce a pre-lexical abstract representation that consists of the same abstract symbols as the lexical representations. This pre-lexical representation is then compared with all lexical representations, and the activation of a lexical representation is increased depending on how well it matches the pre-lexical representation. If the activation of a lexical representation reaches a certain threshold, the acoustic input is recognized as that word (see for more details and an overview of several models, McQueen, 2005).

Potential deviations between the pre-lexical and lexical representations are taken care of by the assumption that the lexical representations of some segments are underspecified for certain features (e.g., Lahiri and Reetz, 2002) or by the application of (probabilistic) generalizations (e.g., Gaskell, 2003). For instance, if the acoustic input gives rise to the pre-lexical representation *garde*[m b]*ench*, this representation can be mapped onto *garde*/n b/*ench* in the lexicon, because final /n/s are assumed to be underspecified for place of articulation such that they match with all nasal plosives, or because there is a generalization linking [mb] to /nb/. In order to account for the comprehension of reduced speech, abstractionist models need well-defined mechanisms mapping reduced pronunciation variants on the corresponding abstract lexical representations.

1.3. Exemplar-based models

Exemplar-based models differ crucially from abstractionist models in the assumption that the mental lexicon contains many exemplars of every word, together forming a word cloud. These exemplars represent the different tokens encountered by the language users, either in their own productions or in the speech of others. Importantly, exemplars contain detailed information about the acoustic or articulatory characteristics of the tokens. Thus, unlike abstractionist models, the stored representations do not consist of abstract symbols and they do not abstract away from speaker or situation specific characteristics. There are several variants of exemplar-based models, but in the following the term ''exemplar-based models'' refers to models with just these basic assumptions, in order to focus on the merits of exemplars.

The assumption of phonetically fully specified lexical representations is supported by a number of experimental studies on speech production and comprehension (e.g., Craik and Kirsner, 1974; Schacter and Church, 1992). For instance, Goldinger (1998) reported that participants tend to mimic previously heard pronunciations in their fine phonetic detail and Cole et al. (1994) found that participants are faster in determining whether two words in a sequence are identical if these two words are produced by the same voice (see also Mattys and Liss, 2008). These findings ask for the storage of indexical information (including information about the speaker's characteristics and speech rate) at least in short-term memory.

Exemplar-based models treat acoustic reduction in the same way as they treat pronunciation differences between speakers and situations: All pronunciation variants are stored in the mental lexicon. Thus, the lexicon would contain representations for *computer* and *puter*, and for *yesterday* and *yeshay*, and different representations of these forms for different voices and situations. At first sight, therefore, exemplar-based models can easily incorporate acoustic reduction and – unlike abstractionist models – they do not need additional assumptions.

In exemplar-based models, the production of a word involves the activation of all its exemplars. In some variants, only one of these exemplars is eventually selected for pronunciation (which one is determined by the frequencies of occurrence and recency of the exemplars, e.g., Bod, 2006). In other variants, all exemplars together form some kind of representation, with the contribution of each exemplar being proportional to its frequency and recency, and it is this echo of exemplars that forms the input to phonetic implementation (Goldinger, 1998). So far, the different descriptions of exemplar-based models have paid little attention to the presence and function of a phonetic component. It is therefore unclear how much of an actual pronunciation is assumed to be determined by the exemplars in the lexicon and how much by the articulatory implementation of (the echo of) these exemplars.

Comprehension within exemplar-based models consists of the direct mapping of the perceived acoustic event on the exemplars in the mental lexicon. An exemplar is activated depending on how well it matches the acoustic input and it passes its activation to its word node, which carries all types of information on the word, including its semantic and syntactic properties. Importantly, the comprehension process thus proceeds without a mediating pre-lexical representation and without speaker normalization. This assumption solves the problem of the extraction of such an

abstract pre-lexical representation from the acoustic input, a process that appears very difficult to formalize and to implement computationally (for an overview, see e.g., Johnson, 1997).

2. Production

As mentioned above, many simple reduction patterns can be explained by the assumption that speakers reduce the sizes of their articulatory gestures or produce them partly simultaneously. Thus, in the English phrase *perfect memory* the /t/ may be inaudible because the articulatory closure gesture is reduced in time and space and its release coincides with the closure of the following bilabial stop. Similarly, /ə/ may be inaudible in words like *potato* and *tomorrow*, as a result of articulatory overlap of the surrounding consonants (Davidson, 2006). These articulatory explanations originally come mainly from Articulatory Phonology (Browman and Goldstein, 1990, 1992), which has as its basic units abstract representations of articulatory gestures, instead of phonemes or phonological features. Importantly, however, these mechanisms can also easily be incorporated in a phonetic component, processing the (abstract) output of a phonological component.

Several studies support the assumption that acoustic reduction results above all from phonetic implementation (rather than from some phonological component), showing that especially speech rate and the predictability of the words and of the neighboring words affect reduction degree. Other studies suggest that the production of acoustic reduction is a more complex process, which can only be well accounted for with the lexical storage of at least some pronunciation variants. The following sections discuss these studies, focusing on how they contribute to our understanding of the relevance of abstract representations and exemplars in speech production.

2.1. Speech rate

One of the most robust predictors of reduction degree documented so far is speech rate. For instance, speakers more often delete word-internal /t/ and /d/ in American English (Raymond et al., 2006) and more often lenite intervocalic stops to fricatives or delete them completely in Florentine Italian, when speaking fast (for more examples see Kirchner, 1998). Prosodic factors, such as the position of stress and prosodic boundaries, may be considered as affecting local speech rate, and accordingly they have been shown to correlate with reduction degree as well. Segments, syllables, and words tend to be less reduced if they carry word stress or sentential accent, or are in the initial or final position of a prosodic unit (e.g., Dalby, 1984; Kohler, 1990; and Kirchner, 1998 for many more examples).

This positive correlation between local speech rate and degree of reduction may suggest that acoustic reduction results from time pressure: Speakers wish to speak fast but may only be able to do so by reducing the number of segments and syllables. A high speech rate, however, does not necessarily lead to speech reduction (van Son and Pols, 1990, 1992): some speakers are well capable to completely control their articulators and produce unreduced pronunciations at high speech rates as well. Furthermore, the question arises whether acoustic reduction indeed *results* from high speech rates or whether the positive correlation is rather due to the fact that both high speech rates and high reduction degrees are indicators of casual speech, the register in which reduction is most common.

If acoustic reduction results especially from time pressure, this can be well accounted for by a phonetic component, which is one of the components of abstractionist models. It can be accounted for within exemplar-based models, if these models incorporate a phonetic component as well. In addition, these models may assume that a speaker's choice of a pronunciation variant may depend on the time available for the articulation of the word. The duration of a variant may be specified in the exemplar directly or indirectly by the number and type of its articulatory gestures or acoustic characteristics.

2.2. Predictability

In addition to speech rate, the predictability of a word and its components appears an important predictor for reduction degree: as documented in several corpus studies, (phonological) units tend to be more reduced if they are more predictable. Note that a greater predictability does not necessarily imply that listeners can really predict the unit, but just that the chance that they guess correctly is higher.

One of the best investigated measures of a unit's predictability is its frequency of occurrence, which can be considered as its a priori likelihood. Several corpus studies have demonstrated that more frequent words tend to be more reduced and shorter than less frequent words (see e.g., Pluymaekers et al., 2005a; Bell et al., 2009). Similar results have been obtained for syllables (Aylett and Turk, 2006). These frequency effects may be accounted for with the assumption that, as a result of practice, speakers are better in anticipating upcoming articulatory gestures in high frequency units and may accordingly modify the current ones. This may lead to compressed gestures produced in overlap (e.g., Bybee, 2001:164).

Another well investigated measure of predictability is the unit's probability given the preceding and following units. Scheibman and Bybee (1999) reported that the English word sequence *don't* is most reduced if it is preceded by *I*, the word that most frequently precedes *don't*. Similarly, the sequence is more reduced before the words that most often follow *don't* (*know*, *think*, *mean*). More recent corpus studies have convincingly demonstrated that there is not just a difference between high frequency and low frequency combinations, but that the probability of a unit given the preceding and following units gradually predicts degree of reduction. The predictability of a unit given the preceding or following unit is typically captured by the frequency of the combination, the conditional probability of the unit given the other units, or by entropy measures (e.g., Bell et al., 2003, 2009; Pluymaekers et al., 2005b). Interestingly, reduction degree appears to be correlated especially with the predictability of the word given the following word rather than the preceding word.

These predictability effects can be well accounted for with the assumption that more predictable words are easier to plan and to retrieve from the mental lexicon (e.g., Jescheniak and Levelt, 1994) and therefore do not require speakers to slow down: While planning highly predictable words, speakers can continue speaking as fast as they would like to. The higher speech rate with which highly predictable units can be produced would be responsible for their higher reduction degree (see e.g., Pluymaekers et al., 2005a; Bell et al., 2009). This account has the advantage that it also explains why the predictability of a word given the following word appears a better indicator for degree of reduction than the word's predictability given the preceding word: Speech rate is determined especially by the planning of the next word rather than that of the preceding or current word. In addition, this account can accommodate effects of a word's predictability given words at a greater distance in the conversation, such as the presence of another token of the same word, which tends to induce more reduced realizations (e.g., Fowler and Housum, 1987; Aylett and Turk, 2004). Finally, this account also predicts correctly that words are less reduced if they are followed by hesitations, which indicate planning problems (e.g., Jurafsky et al., 2001). Note that this account considers sentences just as strings of words without any hierarchical structure.

In contrast to this speaker-driven account of the observed predictability effects is the listener-driven explanation: Speakers would like to reduce as much as possible in their articulatory effort but only reduce those units that can easily be recognized by the listener, in order to guarantee smooth conversation. More predictable units are easier to understand and speakers would therefore reduce especially highly predictable units (e.g., Aylett and Turk, 2004; Boersma, 1998). This listener-driven hypothesis is in line with Lindblom (1990)'s Hyper- and Hypospeech theory (H&H Theory), which states that speakers adapt their speech style according to the communicative and situational demands. In addition, the listener-driven hypothesis easily accommodates van Son and Pols (2003a,b)'s findings that segments are reduced less the more they contribute to the disambiguation of the word. Furthermore, Kuperman et al. (2008) documented that in the four languages they investigated (Dutch, English, German, and Italian) speakers use especially combinations of speech sounds of medium durations, probably because the long ones are expensive in articulatory effort, while the short ones are difficult to identify. The relevance of the listeners' expectations is still an open question, however. Bard and colleagues (Bard et al., 2000) showed in a map-task that a speaker tends to produce the second mention of a word as more reduced not only when it is the second mention for the listener as well, but also when it is just the first mention for the listener. Furthermore, Ernestus and Baayen (2007) showed that a word's lexical frequency hardly predicts ease of recognition for reduced pronunciations, which makes it unlikely that speakers tend to reduce especially high frequency words because listeners would understand them more easily than low frequency words.

In conclusion, several studies have documented that more predictable units tend to be reduced to a greater extent. Importantly, especially the speaker's expectations appear relevant, which suggests that the predictability effects are part and parcel of the production process. This finding can easily be incorporated in abstractionist models assuming that phonetic implementation is sensitive to the planning of the upcoming words. Gestures would be reduced in their sizes and they would overlap more in time if the next words are ready for production. Exemplar-based models can also deal with the effects of planning if they adapt the same assumption that would also be necessary to account for effects of speech rate (see section 2.1), that is, phonetic implementation is partly responsible for reduction, or the speakers' choice of a pronunciation variant may be influenced by the time span that speakers would like to fill with that particular word.

2.3. Differences among speakers

The studies on the roles of speech rate and predictability in acoustic reduction may create the impression that reduction results completely automatically from the process of speech production, over which speakers have little control. If so, the social and economic position of the speaker should have little effect on acoustic reduction. Given that acoustic reduction is still a highly under-investigated topic, it is not surprising that so far only few studies have addressed the role of sociolinguistic factors in acoustic reduction. Importantly, however, those that did, show clear sociolinguistic effects.

First, several studies have reported an effect of gender. These studies confirm the general differences observed between male and female speech: women tend to use standard and prestigious forms more often than men (e.g., Labov, 1972), and they also tend to take the lead in language change (e.g., Chambers, 1995; Labov, 2001). For instance, women

delete glides (Phillips, 1994) and word-final /t/s and /d/s (Guy, 1991) less often than men in American English, and they less often reduce the suffix *lijk* /lək/, for instance to [k], in Dutch (Keune et al., 2005). In contrast, women more often use reduced pronunciation variants that are the norm or have prestige and that may be used in more formal speech as well (and are therefore of a very different nature than the reduced variants attested only in spontaneous conversations). Examples are the realization of the infinitive suffix /ən/ as [ə] in Dutch (van Hout and van de Velde, 2000), and the use of the glottal stop instead of [q], which is the classical, literal variant, in Arabic (for an overview, see e.g., Chambers, 1995).

Second, a speaker's age may affect degree of reduction. Language change often results from reduction patterns introduced by young people (e.g., the loss of word-final segments, the loss of case systems, etc., see e.g., Hay and Sudbury, 2005). Therefore, when a language change is in progress, we may expect that elderly speakers tend to reduce less than young people. In addition, most elderly people hear less well, which hinders the comprehension of reduced speech (Janse and Ernestus, 2011), and they may reduce less than they used to do, because they talk mostly to other elderly people, who need more, redundant, acoustic input, or because they wish to hear themselves well. Initial evidence that older people indeed reduce less than younger speakers comes, for instance, from Cedergren (1987), showing that in Panama Spanish [t] tends to be produced as [ʃ], especially by people born after 1950, from Guy (1991), who demonstrated that older speakers delete word-final /t/s and /d/s less often than younger speakers of American English, and from Strik et al. (2008), who reported that older people less often delete segments in spontaneous Dutch.

Finally, several studies have documented a role for the speaker's socio-economic status in the use of reduced forms. Speakers of a lower status use several types of less prestigious reduced pronunciation variants more often, and also tend to show smaller differences between their speech registers, as appears for instance from the pronunciation of the English progressive suffix *-ing* as [ən] (Labov, 2001:265) and from the pronunciation of /ϴ/ in *think* and *with* as [t] in New York City English (Labov, 1972). Interestingly, speakers of a middle social class do not always take an intermediate position between the high and low social classes, because of hyper-correction: They may recognize a certain pronunciation as characteristic of a lower social class and therefore try to avoid it completely, even though the pronunciation also occurs in the speech of the upper social class. Generally, what appears relevant, both for the speakers of a lower status and of a middle social status, is whether the variation has attained socio-stylistic meaningfulness (markers, stereotypes) or not (indicators – see Labov, 1972 for this tripartition).

In summary, sociolinguistic studies show that speakers have control over their reduction degree. This conclusion is supported by van Son and Pols (1990, 1992), who reported differences between speakers in the effect of speech rate on reduction (see section 2.1). Importantly, this implies that if reduction results from phonetic implementation, phonetic implementation is under control of the speaker as well.

2.4. Differences among words

Nearly all studies investigating acoustic reduction start from the assumption that words with the same phonological, prosodic, and predictability characteristics do not differ from each other in their reduction degree, which would imply that reduction can be completely accounted for by phonetic processes. Upon closer investigation, this assumption appears too simplistic.

First, there appears to be a fundamental difference in the reduction of function words and content words. These word types are affected differently by their predictability: Whereas lexical frequency and number of mentions are stable predictors for reduction degree of especially content words, the effect of the word's predictability given the preceding word appears to be restricted to highly frequent function words or discourse markers, both in American English (Bell et al., 2009) and Dutch (Pluymaekers et al., 2005b). In addition, function words differ from content words in that they tend to be more reduced, which cannot be completely ascribed to their higher predictability (Bell et al., 2009).

Furthermore, differences in reduction degree may be found between words that differ from each other only in their meaning. For instance, Local (2003) reports that the word combination *I think* is more reduced when it occurs in word-final position and conveys a pragmatic meaning (e.g., in the sentence *they should be here by the time you come out next weekend I think*) than when it is followed by the complementizer *that* and has above all a lexical meaning (e.g., *I think that people have not yet woken up*). Similarly, Plug (2005) reports that the reduction degree of the Dutch word *eigenlijk* 'actually' depends on whether the word signals contrast with what the speaker has suggested before or with what the listener may think.

Other studies report differences between words that we cannot (yet) explain by differences in their formal characteristics. Keune and colleagues (2005) reported differences among Dutch words ending in the suffix *lijk* /lək/: While six words show only a medium degree of reduction, eight words also show high reduction (i.e., the pronunciation of /lək/ as [k]). Moreover, some of these words are more reduced in Flanders, the Dutch speaking part of Belgium, while others are more reduced in the Netherlands. These differences suggest that we are dealing with idiosyncratic reduction, which is easier to explain if we assume the presence of several pronunciation variants for each word in the mental lexicon.

Support for the storage of multiple variants comes from research showing that words differ in the relative frequencies of their pronunciation variants and that these relative frequencies may affect the production process. Bürki et al. (2010) asked native speakers of French to produce words with or without their schwa in the initial syllable (e.g., *fenêtre* or *fnêtre* 'window') upon seeing a prompt on a computer screen (a picture of the object or an abstract symbol). They found that speakers produce a variant more quickly the more frequent it is relative to the other variant. This suggests that French speakers have lexical representations for the unreduced as well as the reduced pronunciation variants which are specified for their relative frequencies. Further research is necessary to investigate whether this conclusion on schwa deletion in French generalizes to other reduction patterns and other languages or whether it is restricted to special cases of ''lexicalized'' reduction.

2.5. Conclusions for production models

As discussed above, acoustic reduction appears to be a complex process affected by several characteristics of the speech stream, of the words in this stream, and by the speaker's characteristics. The findings reported so far allow us to shed some light on which assumptions abstractionist models and exemplar-based models of speech production need to make in order to account for the production of reduced speech.

The roles of speech rate, planning ease, and the articulatory and acoustic characteristics of the segments suggest that at least some reduced pronunciation variants result from speakers' reduction in the sizes of the articulatory gestures and from having them overlap in time. When speakers wish to speak fast and are not delayed by planning problems, they may reduce articulatory effort, which allows them to easily keep up their speed, but also results in acoustically reduced pronunciations. Hence, all models can account well for many reduction patterns if they assume that reduction results at least partly from phonetic implementation.

This phonetic component may also account for prosodic influences on degree of reduction. It may translate prosodic structure into adjustments of the local speech rate (i.e. the presence of prosodic boundaries and accents may result in lower local speech rates), and a lower rate would lead to stronger articulatory gestures and less articulatory overlap, which decrease the probability of reduction. Further research is necessary to investigate the possibility that prosody codetermines a word's pronunciation also directly, independently from speech rate.

Importantly, degree of reduction appears under speakers' control, since a high speech rate does not necessarily lead to more reduction and since reduction degree can be affected by the speaker's socio-economic position. If reduction results at least partly from phonetic implementation, this implementation should therefore be under the speaker's control as well, instead of being a completely automatic process. This finding fits in with modern views of the phonetic component as being language-specific and therefore part of the speakers' knowledge of their languages (e.g., Keating, 1990).

The question now arises how much of the observed reduction and which reduction patterns exactly can be explained by just phonetic implementation. The assumption that articulatory gestures may be reduced in size and may overlap in time can easily explain the acoustic absence of single segments, such as the /t/ in *perfect memory*. However, may these mechanisms also account for the absence of several segments in a sequence, such as the pronunciation *yeshay* for English *yesterday* and [xon] for Dutch /xəwon/ 'normally'? We then have to assume that several articulatory gestures in a sequence may be reduced to size zero. This assumption is very similar to phonological deletion rules or constraints, as assumed in abstractionist production models based on segments or features.

Additional evidence that reduction degree is not only determined by phonetic implementation comes from the observed differences in reduction degree among phonologically highly similar words. A word's degree of reduction may, for instance, be co-determined by its exact pragmatic function in the discourse. Abstractionist models therefore have to assume that also the various grammatical components preceding phonetic implementation play an important role in reduction. These components have to contain rules or constraint constellations that may apply to only a few words (e.g., discourse markers) and that may be sensitive, among others, to the function of these words in the running discourse. This is almost similar to assuming that at least several words are stored together with all their pronunciation variants, and thus that one and the same word may be represented by several underlying forms, specified for the conditions in which they appear.

So far, only one study has provided direct, on-line evidence for the use of different lexical representations for the different pronunciation variants of the same word in production (Bürki et al., 2010). Additional evidence comes from Hinskens (2011), who showed in a corpus study that vowel reduction in Dutch is sensitive especially to the structural characteristics of the vowels and their syllables as well as to the frequency characteristics of a word. As these frequency characteristics are word-specific, this result suggests that degree of reduction is lexically stored. If these results generalize to other reduction patterns, also in other languages, they form convincing evidence that the lexical storage of more than one pronunciation variant for a word is rather the rule than the exception. It would question which of the results discussed above also have to be accounted for by the lexical storage of pronunciation variants, instead of rules, constraint constellations, or even phonetic implementation.

Purely exemplar-based models of speech production can do without the assumption that some reduction patterns result from phonetic implementation. Such models may assume that all possible pronunciation variants are lexically stored and that phonetic factors, such as speech rate and planning ease, codetermine which exemplar is selected as production target. Produced pronunciation variants then always closely resemble the exemplar selected for production. Nevertheless, also these models have to incorporate a role for articulatory mechanisms if they wish to account for which pronunciation variants may emerge, including which types of reductions may be expected in language change. They need to assume that a speaker selects a given exemplar (or echo of exemplars) as production target but that the resulting acoustic output may differ from this input due to overlapping and reduced articulatory gestures. This raises the question of which part of the observed pronunciation variation reflects characteristics of the selected exemplar and which part results from exact implementation of this exemplar. This question is urgent, since the role of phonetic implementation has been neglected so far in the discussion of exemplar-based models.

In conclusion, both abstractionist and exemplar-based models may well account for the facts that so far have been gathered about the production of acoustically reduced pronunciation variants. Both frameworks, however, need to make additional assumptions. Exemplar-based models have to take into account that the exact articulatory implementation of the phonological (output) forms plays an important role in determining the characteristics of the acoustic outputs. Abstractionist models have to accept that similar words may differ in their degree of reduction and, moreover, that the production process appears to be sensitive to the frequencies of different pronunciation variants, which supports the storage of several pronunciation variants for at least some words.

3. Comprehension

So far, only little attention in the literature has been devoted to the comprehension of acoustically reduced pronunciation variants. As a result, we still know very little about which mechanisms are involved, how they interact, and whether their relevance depends on reduction degree.

At first sight, accounting for the comprehension of reduced pronunciation variants may seem unproblematic, since these variants could simply be mapped onto the best matching unreduced lexical representation. We only have to assume that the mapping process allows for weakened and missing segments. Thus, [jɛʃei] may be mapped onto *yesterday*, since *yesterday* is the only English word starting with /jɛ/ followed by a sibilant and /ej/ later in the word. However, this view appears too simplistic. Several reduced pronunciation variants are very similar to the unreduced pronunciations of other words. For instance, *mist* with an acoustically absent /t/ is very similar to *miss*. How do listeners know that a speaker intended *mist* instead of *miss* if the acoustic signal matches *miss* best? Furthermore, this simple pattern matching cannot explain how listeners know which word a speaker intended if the acoustic signal matches several unreduced forms equally well. For instance, Dutch reduced [ɛik] used in the sense of 'actually' is not only embedded in *eigenlijk* /ɛixələk/ 'actually', but also in /ɛindələk/ 'finally', /ɛisələk/ 'terrific' /ɛisvərkopər/ 'ice cream seller', /ɛizərstɛrk/ 'strong as iron', etc. In conclusion, an abstractionist model in combination with just simple pattern matching cannot explain the comprehension of reduced forms.

Similarly, a simple exemplar-based model cannot account well for the comprehension of conversational speech. Lexical representations for reduced pronunciation variants may explain the comprehension of such variants, but they also imply a very high number of lexical representations and consequently severe lexical competition. This competition would be especially harmful for low frequency forms that are phonologically similar to more frequent forms. For instance, the Dutch word *tule* /tylə/ 'tulle' would have the additional competitor [tyk], a reduced form of the high frequency word *natuurlijk* 'of course', and would therefore be very difficult to recognize, which is not in line with our daily experience with comprehending conversational speech. The question therefore is how exemplar-based models should deal with this competition.

In this section, I will discuss what we have learned from the few studies on comprehension that have been described in the literature, focusing on the implications for our view of the mental lexicon. I will first summarize studies on the role of top down information and then discuss several other factors that appear relevant in the comprehension of reduced speech. I will conclude by discussing the implications for psycholinguistic models of speech comprehension.

3.1. Top down information

The very first study on the comprehension of highly reduced pronunciation variants, such as [ɔnri] for *ordinary* and [jɛʃei] for *yesterday*, was based on the incidental observation that speakers are typically not aware of such variants in their own speech or in the speech of others, and that they do not recognize these variants when presented out of context. Ernestus et al. (2002) confirmed this incidental observation with an experimental study. They selected from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000) word tokens with either high, medium or low reduction. Participants listened to these tokens, either spliced out of their contexts (e.g., [mok], a reduced form of /moxələk/ 'possible'), together with the

neighboring vowels and any intervening consonants (e.g., [ɛlmokna]), or in their prosodic phrases (e.g., *zo snel* [mok] *naar eh* 'as fast as possible to uhm'). Participants recognized the tokens with low or medium reduction in more than 85% of the trials, independently of how much context they heard. The recognition of the highly reduced tokens, in contrast, was much more problematic, as these tokens were only well recognized in their phonological phrases (92% correct identifications). Their identification scores dropped to 70% when they were presented with just their surrounding vowels plus intervening consonants (e.g., [ɛlmokna]), and to 52% when they were presented in isolation (e.g., [mok]). Hence, listeners need more than just local phonetic context to understand highly reduced pronunciation variants. Interestingly, further analysis showed that the identification scores for the tokens presented out of context were correlated with the phonetic distances between these tokens and the corresponding unreduced forms: The higher the relative number of altered or missing segments, the worse the identification scores.

A possible explanation for why listeners are not aware of highly reduced pronunciation variants, even though these variants are very frequent in everyday speech, is that listeners unconsciously reconstruct reduced variants to their unreduced counterparts on the basis of context. This hypothesis was tested in a series of phoneme monitoring experiments (Kemps et al., 2004), in which Dutch listeners were presented with stretches of spontaneous speech, containing the suffix /lək/. This suffix was either unreduced ([lək]) or highly reduced to [k]. In the first experiment, participants were asked to press a button as soon as they heard an [l]. When the suffix was presented in isolation, participants were very well able to perform this task and they only pressed the button upon hearing an unreduced pronunciation variant. When the suffix was presented in its full sentential context, however, most participants reported an [l] also when hearing the reduced suffix realization [k], without [l]. Apparently, upon hearing a reduced form of the suffix, participants reconstructed the corresponding unreduced variant, on which they based their responses. Interestingly, this reconstruction is time consuming, as it took participants on average 240 ms longer to press the button after hearing the reduced form [k] than after hearing the unreduced form [lək]. Two follow up experiments showed that the reconstruction was based both on the pronunciation of the unreduced form and on the orthographic transcription of the suffix (*lijk*), which contains an *l*.

These two series of experiments may raise the question whether listeners base their comprehension of highly reduced pronunciation variants on the acoustic properties of these forms at all. Maybe listeners just formulate predictions about the identity of a reduced form on the basis of its contexts, and they believe to hear the most likely word. This, however, appears not to be the case.

Van de Ven et al. (2012) presented Dutch listeners with sentences from spontaneous conversations, with a reduced target word substituted by a beep. Participants guessed correctly which word, out of four options, was left out in only 50% of the trials. Apparently, listeners also need the acoustic characteristics of the pronunciation variant itself for correct word identification.

Given the importance of context, we may expect also some role for other types of top down information in the comprehension of reduced pronunciation variants. Indeed, Mitterer and Ernestus (2006) and Janse et al. (2007) have documented a role for lexical information in the comprehension of Dutch words with weakened or absent final /t/. Mitterer and Ernestus presented Dutch listeners with words ending in acoustically weak /t/s and asked them whether a /t/ was present. Participants tended to report the presence of /t/ more often when the /t/ turned the word into an existing word of Dutch (in the case of *orkes* +*t* 'orchestra') than if it did not (in the case of *moeras*; only *moeras* 'swamp' is an existing word in Dutch, *moerast* is not). Janse, Nooteboom, and Quené documented reconstruction of /t/ based on lexical information even in the complete absence of any acoustic trace of /t/. Participants were asked to press a button as soon as they heard a predefined word. They pressed the button in many trials where the word was produced without its final /t/ (e.g., *orkes* instead of *orkest*), but only if this stimulus without /t/ did not form an existing word by itself, as in *orkest* (in contrast to *moeras*). Thus, there were few false alarms for stimuli such as *kas*, which is a word by itself ('green house') as well as the reduced form of the /t/-final word (*kast* 'cupboard').

Abstractionist models can easily account for most of these findings. The low identification scores for highly reduced pronunciation variants presented in isolation fit in with the assumption that only unreduced pronunciation variants are stored in the mental lexicon. The more an acoustic form deviates from the corresponding unreduced pronunciation, the more difficult it is to map that acoustic form onto the unreduced lexical representation. Furthermore, listeners are not aware of reduced pronunciation variants, also because they comprehend reduced pronunciation variants by mapping them on the corresponding unreduced variants. More challenging for abstractionist models is the finding that reconstruction is facilitated by context (I will come back to this in section 3.3).

The findings appear more challenging for simple exemplar-based models than for abstractionist models. Given the assumption that all pronunciation variants are lexically stored, it is surprising that listeners cannot easily recognize highly reduced variants out of context. Moreover, listeners' reconstruction of reduced forms and the positive correlation between the identification scores for highly reduced variants presented out of context and these forms' phonetic distances to the corresponding unreduced variants suggest that the unreduced pronunciation has a privileged status, which is also unexpected in a simple exemplar-based framework.

3.2. Frequencies of use, auditory mechanisms, and learning

In addition to top-down information, many other factors play a role in the comprehension of reduced speech. So far, additional factors have mostly been investigated on the basis of mild reductions, especially lenited or deleted /t/s, schwas, and flaps. Also such mildly reduced forms are less easily recognized than their unreduced counterparts, at least in isolation or in simple, constructed sentences (e.g., Ernestus and Baayen, 2007; Janse et al., 2007; Tucker and Warner, 2007). This is the case even though the reduced pronunciation variant may be more frequent than the unreduced variant.

Ranbom and Connine (2007) reported that there are frequency effects in the comprehension of reduced pronunciation variants. They investigated the comprehension of American English words containing an /nt/ sequence that can be pronounced as a flap (e.g., *gentle*). In a lexical decision experiment, participants' response latencies were shorter for the [nt] realizations than for the flap realizations, which replicates the finding that reduced pronunciation variants are less easily recognized than their unreduced counterparts. Crucially, however, the difference in response latencies between the two variants was smaller for the words that are relatively often produced with the flap. Ranbom and Connine replicated these results in an experiment in which the words were presented at the end of simple sentences and participants performed lexical decision on the orthographic representations of these same words (cross-modal repetition priming). Ranbom and Connine conclude that both pronunciation variants of an /nt/ word, the [nt]-variant and the flap variant, are stored in the mental lexicon, but that the unreduced [nt]-variant has a special status, possibly, among others, because it corresponds to the word's orthographic representation. Note that Bürki et al. (2010) reached the same conclusion on the basis of production data for schwa deletion in French. Hence, there is converging evidence that speakers and listeners are sensitive to the frequencies of the pronunciation variants resulting from simple alternations (e.g., flapping of /nt/ sequence or schwa deletion). These frequencies much therefore be stored, which suggests that these pronunciation variants have their own lexical representations.

Nevertheless, also the recognition of mildly reduced pronunciation variants may proceed via the corresponding unreduced variant. This appears to be the case for words ending in [s] followed by an acoustically weak or absent /t/. As mentioned above, listeners tend to perceive word-final /t/ after [s] even if there are at best only very weak acoustic cues for /t/ (Mitterer and Ernestus, 2006; Janse et al., 2007). Mitterer et al. (2008) show that this is, among others, because the human ear cannot well discriminate between the different variants of /t/ after [s], whereas this appears much easier after [n]. The most convincing evidence for insensitivity of the ear to some reduction patterns comes from the similarity between the discrimination abilities of Dutch listeners, who often produce and hear reduced variants of /t/ after [s], and monolingual Japanese listeners, for whom both the word-final /st/ and the word-final /nt/ sequence are phonotactically illegal.

Finally, comprehension of mildly reduced pronunciations via their stored unreduced counterparts may be facilitated by learning. In their comparison of Dutch and Japanese listeners in the discrimination of different variants of /t/, Mitterer and colleagues (2008) observed that Dutch listeners outperformed Japanese listeners for some reduced realizations of /t/, but only after [s], the context in which /t/ is typically reduced in Dutch (as in most other languages). Dutch listeners have more experience with reduced variants of /t/ after [s] than the Japanese, and apparently this facilitates discrimination, and possibly also comprehension.

In summary, some of the mild reductions that occur in conversational speech are just hardly audible for naive listeners, and therefore hardly affect the comprehension process. The comprehension process of more noticeable mild reductions involves the mapping of these reductions unto the corresponding unreduced lexical representations, a process that may be facilitated by learning, or unto lexical representations of the pronunciation variants themselves.

3.3. Conclusions for comprehension models

The comprehension of reduced speech is a complex process of which we have just started to discover the general mechanisms. The few experimental results make clear that none of the existing models of speech comprehension can easily account for reduced speech without additional assumptions.

As mentioned above, abstractionist models can easily account for all findings suggesting a privileged status for the unreduced pronunciation variant (see section 3.2). In addition, they can account for learning, which would imply the formation of rules or constraints, or the re-positioning of constraints in the hierarchy thus repairing unwellformed input (e.g., Boersma, 1998; Boersma and Hayes, 2001).

Abstractionist models need to make additional assumptions, however, to explain the observation that listeners cannot recognize reduced pronunciation variants out of their contexts. If we assume the same type of constraint interaction for word comprehension as for word production (Boersma, 1998), this observation could be accounted for as follows. We could assume that there are series of constraints which define the maximally permissible (phonetic) difference between an acoustic input and an optimal phonological output on which it can be mapped. If for a given input there is no phonological (output) candidate that satisfies the constraints, this input cannot be identified. Note that we need such constraints also to explain why listeners can classify acoustic inputs as non-existing words (instead of acoustic variants of existing words).

These constraints (or some of them) could be dominated by a series of constraints favoring outputs that are in line with the available top down information. A highly deviant form is then matched with the corresponding word in the lexicon if and only if it receives top down information.

Whereas the effect of context can well be accounted for within abstractionist models, this appears more difficult for the observed frequency effects of some pronunciation variants. Apparently, abstractionist models have to loosen their constraint of only one lexical representation for (nearly) every word, and to allow for several lexical representations for at least some (series of) words. Further research is necessary to establish which pronunciation variants are likely to be incorporated in the mental lexicon.

In contrast to abstractionist models, exemplar-based models can easily account for the frequency effects of different pronunciation variants, as their key assumption is that all pronunciation variants are lexically stored. In addition, like abstractionist models, exemplar-based models can explain learning. Connections may be established between different exemplars of the same word and between similar exemplars of different words (e.g., exemplars with lenited final /t/s). These connections then represent knowledge about possible pronunciation variants, which may be generalized to new words (lexical analogy, see, e.g., Ernestus and Baayen, 2006) and be used in speech comprehension.

Exemplar-based models are challenged by the privileged status of the unreduced form. First, only this form is well recognized out of context, and listeners' identification scores for highly reduced pronunciations presented out of context are correlated with these forms' phonetic distances to the corresponding unreduced pronunciations. We may account for these findings by assuming that reduced pronunciation variants are stored together with information about the conditions under which they occur (i.e., in which context). If variants are presented under different conditions, their lexical representations are not (or less) activated and recognition is based on the unreduced form. A greater phonetic difference between the acoustic input and the unreduced pronunciation variant then inhibits identification. A different account may state that the lexical representations of reduced pronunciation can only be accessed if they are supported by top down information (i.e., listeners' expectations, among others based on context). If no top-down information is available, recognition is based on the unreduced variant and the phonetic distance between the acoustic input and this unreduced variant is relevant. Second, the unreduced pronunciation variant appears special in that it is mostly this variant that listeners believe to hear. Exemplar-based models can account for this finding only by making yet an additional assumption. For instance, upon hearing a reduced variant, listeners activate the lexical representation of this variant and this activation spreads to all other representations in the same word cloud, including the unreduced representation, which gets activated sufficiently to be accessed by the listeners more quickly than reduced variants.

There are probably other ways to account for the challenging findings for abstractionist and for exemplar-based models. What is clear, however, is that both frameworks have to be extended to incorporate these findings, which will make these models more complex.

4. Hybrid models

Given the above, it may be concluded that both the production and comprehension of reduced speech present challenges for abstractionist and exemplar-based models. The most important challenges for abstractionist models are the frequency effects of the pronunciation variants in speech production and comprehension, while exemplar-based models need additional assumptions about the role of phonetic implementation and the privileged status of unreduced pronunciation variants.

Lately, exemplar-based models are also challenged by experimental evidence suggesting that the indexical information in a word token (i.e., information about the speaker's voice and about speech rate) only affects word comprehension if for some reason processing is slow. Apparently, listeners use exemplars under certain processing conditions and abstract lexical representations under other conditions. For instance, McLennan and Luce (2005) ran a series of lexical decision and shadowing experiments in which each target word occurred twice. The effect of the first token on participants' reactions to the second token (identity priming) appeared greatest if the two tokens were similar in speech rate or voice and if simultaneously processing was slowed down, either by very word-like nonwords in the experiment (lexical decision) or by a long forced time span between the stimulus and the response (shadowing). Similarly, Mattys and Liss (2008) reported that listeners are faster in deciding whether a word has occurred before in the same experiment if both tokens are produced by the same speaker and this speaker suffers from dysarthria (i.e., difficulty in articulating words due to disease of the central nervous system), which tends to slow speech comprehension. Indexical effects are therefore not always part and parcel of the comprehension process, in contrast to the prediction of simple exemplar-based models.

Since both pure abstractionist and pure exemplar-based models have advantages but also face challenges, several researchers have proposed hybrid models, combining the advantages of both frameworks. All proposed hybrid models assume both abstract generalizations and exemplars, but they differ in the importance assigned to these in the speech production and comprehension processes. Importantly, none of them have yet been fully implemented computationally

and future research therefore has to show to which extent they can account for the full range of data, including the production and comprehension of reduced speech.

Pierrehumbert (2002) was one of the first to propose a hybrid model and, contrary to most other researchers, she explicitly discussed both speech production and comprehension. Her model assumes abstract generalizations (e.g., prosodic final lengthening) and exemplar clouds associated with phonological units, including phonemes, phoneme sequences, and words. In speech production, speakers would use both exemplars and abstract generalizations, but comprehension would mainly involve the exemplars. In other words, Pierrehumbert's model is hybrid on the production side, but mostly exemplar-based on the comprehension side. As a consequence, it cannot easily account for the finding that indexical information plays a role especially if speech processing is slow. In addition, it faces the same challenges as simple exemplar-based models with respect to the comprehension of reduced speech (i.e., the privileged status of the unreduced pronunciation variants).

Two other hybrid models, both developed for speech comprehension, can account for the effect of processing speed on the relevance of indexical information. The first model (McLennan et al., 2003) is based on the Adaptive Resonance Theory (ART), developed by Grossberg and Stone (1986). The second model is Goldinger's Complementary Learning System (CLS, 2007). Both models assume abstract representations for lexical and sublexical units while indexical information is stored in the form of exemplars. Furthermore, both models assume that processing based on abstract representations always precedes matching of the acoustic signal with acoustically similar exemplars, either because abstract representations are activated more frequently and therefore establish resonance with the acoustic input more easily and more quickly (ART), or because the acoustic signal first passes that part of the brain that is involved in abstract processing (CLS). The question is now where these models store information about pronunciation variation resulting from acoustic reduction (e.g., schwa deletion) rather than from differences between speakers. If the models would store this information in the form of abstract representations, nearly all words will be represented by several abstract representations. This blurs the difference between exemplars and abstract representations, which is against the nature of these models as they consider abstract representations and exemplars as clearly different constructs. Moreover, it would be unclear what determines whether a given pronunciation variant is stored as an abstract representation or as an exemplar. If, on the other hand, the models would represent pronunciation variation resulting from reduction only as exemplars, they predict that lexical representations for this variation play a role in speech comprehension especially when processing is slow (since also speaker characteristics only play a role when processing is slow), and consequently that listeners are sensitive to the frequencies of these pronunciation variants especially during slow processing. This is a hypothesis worth to be tested.

Polysp (Polysystemic Speech Perception), developed by Hawkins and Smith (Hawkins and Smith, 2001; Hawkins, 2003), can also account for the late effect of indexical information on speech processing, but differs substantially from ART and CLS. Most importantly, Polysp assumes that the analysis of an acoustic input into its linguistic units (phonemes, etc.) is not necessary and that circumstances dictate whether this analysis takes place at all, and whether it precedes, coincides, or follows the matching of the acoustic signal with exemplars. Listeners may thus recognize speech especially via the abstract representations (rather than the exemplars), and the unreduced pronunciation variant may consequently have a privileged status, but this is situation dependent. Polysp also differs from most other hybrid models in the assumption that a memory trace not only consists of acoustic information, but also contains its multi-medial context (e.g., visual information about the speakers' articulatory gestures, their mood, the situation). Hence, Polysp already incorporates the assumption that lexical representations for highly reduced pronunciation variants are stored together with the conditions under which they occur, which would explain why these variants are only well recognized in their contexts. In conclusion, Polysp appears very promising, but computational modeling has to show whether it can live up to the expectations it raises. Furthermore, the model needs explicit formulation of the production process, including the role of exemplars and the implementation of articulatory gestures in determining the precise acoustic characteristics of a word token.

In conclusion, the literature contains several proposals for hybrid models. However, most of them are developed only for speech comprehension, and it is unclear how they would account for speech production. More importantly, none of the models have been computationally fully implemented so far, and it is consequently unclear whether and how they would account simultaneously for all findings, especially since the factors playing roles in the processing of conversational speech appear to be of various natures.

5. Conclusions

The phenomenon of acoustic reduction is under-investigated, given its high frequency of occurrence in everyday conversations, but especially given its theoretical relevance. The few published studies on acoustic reduction have shed new lights on the speech production and comprehension process, including the relevance of abstract generalizations and

exemplars. Exemplars appear relevant since both the production and the comprehension process show sensitivity to the frequencies of reduced pronunciation variants, which suggests that these variants have their own lexical representations. Simultaneously, the unreduced variants have special status: They are well recognized out of context and the comprehension process of reduced variants appears to involve the activation of the corresponding unreduced variants. Finally, most of the reduction patterns appear to result from reduction in the sizes of articulatory gestures and from articulatory overlap, which are under the speaker's direct control. This suggests an important role for the exact implementation of articulatory gestures. These combined results ask for hybrid models, in which both abstract representations of the unreduced pronunciation variant, exemplars, and phonetic implementation play a role.

Many characteristics of the production and comprehension process of reduced speech are still unknown. Likewise, no hybrid model of speech production and comprehension has yet been computationally fully implemented. Obviously, more research on both acoustic reduction and hybrid models is necessary to obtain detailed insight in the human language capacity.

Acknowledgements

I would like to thank Frans Hinskens, Marc van Oostendorp and Ben Hermans for their valuable comments on earlier versions of this paper.

References

Adda-Decker, M., Boula de Mareüil, P., Adda, G., Lamel, L., 2005. Investigating syllabic structures and their variation in spontaneous French. Speech Communication 46, 119–139.

Antilla, A., Cho, Y.Y., 1998. Variation and change in Optimality Theory. Lingua 104, 31–56.

Aylett, M., Turk, A., 2004. The smooth redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Language and Speech 47, 31–56.

Aylett, M., Turk, A., 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. Journal of the Acoustical Society of America 119, 3048–3058.

Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., Newlands, A., 2000. Controlling the intelligibility of referring expressions in dialogue. Journal of Memory and Language 42, 1–22.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., Gidea, D., 2003. Effects of disfluencies, predictability and utterance position on word form variation in English conversation. Journal of the Acoustical Society of America 113, 1001–1024.

Bell, A., Brenier, J., Gregory, M., Girand, C., Jurafsky, D., 2009. Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language 60, 92–111.

Bod, R., 2006. Exemplar-based syntax: how to get productivity from examples. The Linguistic Review 23, 291–320.

Boersma, P., 1998. Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual Drives. Holland Academic Graphics, The Hague.

Boersma, P., Hayes, B., 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic Inquiry 32, 45–86.

Browman, C.P., Goldstein, L., 1990. Tiers in articulatory phonology, with some implications for casual speech. In: Kingston, J., Beckman, M.E. (Eds.), Between the Grammar and Physics of Speech [Papers in Laboratory Phonology 1]. Cambridge University Press, Cambridge, pp. 341–376.

Browman, C.P., Goldstein, L., 1992. Articulatory phonology: an overview. Phonetica 49, 155–180.

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.

Bybee, J., 2001. Phonology and Language Use. Cambridge University Press, Cambridge.

Cedergren, H.J., 1987. The spread of language change: verifying inferences of linguistic diffusion. In: Lowenberg, P.H. (Ed.), Language Spread and Language Policy: Issues, Implications and Case Studies. Georgetown University Press, Washington, DC, pp. 45–60.

Cedergren, H.J., Sankoff, D., 1974. Variable rules: performance as a statistical reflection of competence. Language 50, 333–355.

Chambers, J.K., 1995. Sociolinguistic Theory: Linguistic Variation in its Social Significance. Blackwell Publishers, Oxford.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper and Row, New York.

Cole, R.A., Coltheart, M., Allard, F., 1994. Memory of a speaker's voice: reaction time to same-or different-voiced letters. The Quarterly Journal of Experimental Psychology 26, 1–7.

Craik, F.I.M., Kirsner, K., 1974. The effect of speaker's voice on word recognition. The Quarterly Journal of Experimental Psychology 26, 274–284.

Dalby, J.M., 1984. Phonetic Structure of Fast Speech in American English. Indiana University, (PhD dissertation).

Davidson, L., 2006. Schwa elision in fast speech: segmental deletion or gestural overlap? Phonetica 63, 79–112.

Ernestus, M., 2000. Voice Assimilation and Segment Reduction in Casual Dutch, a Corpus-based Study of the Phonology–Phonetics Interface. LOT, Utrecht.

Ernestus, M., Baayen, R.H., 2006. The functionality of incomplete neutralization in Dutch: the case of past-tense formation. In: Goldstein, L.M., Whalen, D.H., Best, C.T. (Eds.), Varieties of Phonological Competence [Laboratory Phonology 8]. Mouton de Gruyter, Berlin, pp. 27–49.

Ernestus, M., Baayen, R.H., 2007. The comprehension of acoustically reduced morphologically complex words: the roles of deletion, duration and frequency of occurrence. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken. pp. 773–776.

Ernestus, M., Baayen, R.H., Schreuder, R., 2002. The recognition of reduced word forms. Brain and Language 81, 162–173.

Fowler, C.A., Housum, J., 1987. Talkers' signaling of new and old words in speech and listeners' perception and use of the distinction. Journal of Memory and Language 26, 489–504.
Gaskell, G., 2003. Modelling regressive and progressive effects of assimilation in speech perception. Journal of Phonetics 31, 447–463.
Goldinger, S.D., 1998. Echoes of echoes? An episodic theory of lexical access. Psychological Review 105, 251–279.
Goldinger, S.D., 2007. A complementary-systems approach to abstract and episodic speech perception. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken. pp. 49–54.
Grossberg, S., Stone, G., 1986. Neural dynamics of word recognition and recall: attentional priming, learning, and resonance. Psychological Review 93, 46–74.
Guy, G.R., 1991. 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.
Hay, J.B., Sudbury, A., 2005. How rhoticity became /r/-sandhi. Language 81, 799–823.
Hinskens, F.L.M.P., 2011. Lexicon, phonology and phonetics. Or rule-based and usage-based approaches to phonological variation. In: Siemund, P. (Ed.), Linguistic Universals and Language Variation. Mouton de Gruyter, Berlin, New York, pp. 416–456.
van Hout, R., van de Velde, H., 2000. N-deletion in reading style. In: de Hoop, H., Van der Wouden, T. (Eds.), Linguistics in the Netherlands. John Benjamins, Amsterdam, pp. 209–219.
Janse, E., Ernestus, M., 2011. 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 39, 330–343.
Janse, E., Nooteboom, S.G., Quené, H., 2007. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. Language and Cognitive Processes 22, 161–200.
Jescheniak, J.D., Levelt, W.J.M., 1994. Word frequency effects in speech production: retrieval of syntactic information and of phonological form. Journal of Experimental Psychology: Learning, Memory, and Cognition 20, 824–843.
Johnson, K., 1997. Speech perception without speaker normalization: an exemplar model. In: Johnson, K., Mullennix, J.W. (Eds.), Talker Variability in Speech Processing. Academic Press, San Diego, pp. 145–165.
Johnson, K., 2004. Massive reduction in conversational American English. In: Yoneyama, K., Maekawa, K. (Eds.), Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium. The National International Institute for Japanese Language, Tokyo, pp. 29–54.
Jurafsky, D., Bell, A., Gregory, M., Raymond, W.D., 2001. Probabilistic relations between words: evidence from reduction in lexical production. Typological studies in language 45, 229–254.
Keating, P., 1990. Phonetic representations in a generative grammar. Journal of Phonetics 18, 321–334.
Kemps, R., Ernestus, M., Schreuder, R., Baayen, R.H., 2004. Processing reduced word forms: the suffix restoration effect. Brain and Language 90, 117–127.
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.
Kirchner, R.M., 1998. An Effort-based Approach to Consonant Deletion. University of California, (PhD dissertation).
Kohler, K.J., 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle, W.J., Marchal, A. (Eds.), Speech Production, Speech Modelling. Kluwer Academic Publishers, Dordrecht, pp. 21–33.
Kuperman, V., Ernestus, M., Baayen, R.H., 2008. Frequency distributions of uniphones, diphones and triphones in spontaneous speech. Journal of the Acoustical Society of America 124, 3897–3908.
Labov, W.L., 1972. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.
Labov, W.L., 2001. Principles of Linguistic Change: Social Factors. Blackwell Publishers, Oxford.
Lahiri, A., Reetz, H., 2002. Underspecified recognition. In: Gussenhoven, C., Warner, N., Rietveld, T. (Eds.), Phonology & Phonetics [Laboratory Phonology 7]. Mouton, Berlin, pp. 637–676.
Lennes, M., Alaroty, 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.
Levelt, W.J.M., 1989. Speaking. From Intention to Articulation. The MIT Press, Cambridge, MA.
Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H Theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modelling. Kluwer Academic Publishers, Dordrecht, pp. 403–439.
Local, J., 2003. Variable domains and variable relevance: interpreting phonetic exponents. Journal of Phonetics 31, 321–339.
Mattys, S.L., Liss, J.M., 2008. On building models of spoken-word recognition: when there is as much to learn from natural ''oddities'' as artificial normality. Perception & Psychophysics 70, 1235–1242.
McLennan, C.T., Luce, P.A., 2005. Examining the time course of indexical specificity effects in spoken word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition 31, 306–321.
McLennan, C.T., Luce, P.A., Charles-Luce, J., 2003. Representation of lexical form. Journal of Experimental Psychology: Learning, Memory, and Cognition 29, 539–553.
McQueen, J.M., 2005. Speech perception. In: Lamberts, K., Goldstone, R. (Eds.), The Handbook of Cognition. Sage Publications, London, pp. 255–275.
Mitterer, H., Ernestus, M., 2006. Listeners recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. Journal of Phonetics 34, 73–103.
Mitterer, H., Yoneyama, K., Ernestus, M., 2008. How we hear what is hardly there: mechanisms underlying compensation for /t/-reduction in speech comprehension. Journal of Memory and Language 59, 133–152.
Norris, D., 1994. Shortlist: a connectionist model of continuous speech recognition. Cognition 52, 189–234.
Pagliuca, W., Mowrey, R., 1987. Articulatory evolution. In: Ramat, A.G., Carruba, O., Bernini, G. (Eds.), Papers from the Seventh International Conference on Historical Linguistics [Current issues in Linguistic Theory 48]. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 459–472.
Phillips, B., 1994. Southern English glide deletion revisited. American Speech 69, 15–127.

Pierrehumbert, J., 2002. Word-specific phonetics. In: Gussenhoven, C., Warner, N., Rietveld, T. (Eds.), Phonology & Phonetics [Laboratory Phonology 7]. Mouton, Berlin, pp. 101–140.
Pitt, M., 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, 90–95.
Plug, L., 2005. From words to actions: the phonetics of eigenlijk in two communicative contexts. Phonetica 62, 131–145.
Pluymaekers, M., Ernestus, M., Baayen, R.H., 2005a. Lexical frequency and acoustic reduction in spoken Dutch. Journal of the Acoustical Society of America 118, 2561–2569.
Pluymaekers, M., Ernestus, M., Baayen, R.H., 2005b. Articulatory planning is continuous and sensitive to informational redundancy. Phonetica 62, 146–159.
Prince, A., Smolensky, P., 2004. Optimality Theory: Constraint Interaction in Generative Grammar. Blackwell Publishers, Oxford.
Ranbom, L.J., Connine, C.M., 2007. Lexical representation of phonological variation in spoken word recognition. Journal of Memory and Language 57, 273–298.
Raymond, W.D., Dautricourt, R., Hume, E., 2006. Word-internal /t,d/ deletion in spontaneous speech: modeling the effects of extra-linguistic, lexical, and phonological factors. Language Variation and Change 18, 55–97.
Schacter, D.L., Church, B.A., 1992. Auditory priming: implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition 18, 915–930.
Scheibman, J., Bybee, J., 1999. The effect of usage on degrees of constituency: the reduction of don't in English. Linguistics 37, 575–596.
Strik, H., van Doremalen, J., Cucchiarini, C., 2008. Pronunciation reduction: how it relates to speech style, gender, and age. In: Proceedings of Interspeech 2008, Brisbane. pp. 1477–1480.
Tucker, B.V., Warner, N., 2007. Inhibition of processing due to reduction of the American English flap. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken. pp. 1949–1952.
van de Ven, M., Ernestus, M., Schreuder, R., 2012. Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. Laboratory Phonology 3, 455–481.
van Son, R.J.J.H., Pols, L.C.W., 1990. Formant frequencies of Dutch vowels in a text, read at normal and fast rate. Journal of the Acoustical Society of America 88, 1683–1693.
van Son, R.J.J.H., Pols, L.C.W., 1992. Formant movements of Dutch vowels in a text, read at normal and fast rate. Journal of the Acoustical Society of America 92, 121–127.
van Son, R.J.J.H., Pols, L.C.W., 2003a. Information structure and efficiency in speech production. In: Proceedings of Eurospeech 2003, Geneva. pp. 769–772.
van Son, R.J.J.H., Pols, L.C.W., 2003b. An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness. In: Proceedings of 15th International Congress of Phonetic Sciences, Barcelona, pp. 2141–2144.