Corpora and exemplars in phonology

Mirjam Ernestus *,[1]

Radboud University Nijmegen and Max Planck Institute for Psycholinguistics

R.Harald Baayen

Department of Linguistics, University of Alberta

1 Introduction

This chapter reviews the role of corpora in phonological research, as well as the role of exemplars in phonological theory. We begin with illustrating the importance of corpora for phonological research as a source of data. We then present an overview of speech corpora, and discuss the kinds of problems that arise when corpus data have to be transcribed and analyzed. The enormous variability in the speech signal that emerges from speech corpora, in combination with current experimental evidence, calls for more sophisticated theories of phonology than those developed in the early days of generative grammar. The importance of exemplars for accurate phonological generalization is discussed in detail, as well as the characteristics of and challenges to several types of abstractionist, exemplar, and hybrid models.

2 The importance of corpora for phonology

2.1 Getting the facts right

Why are corpora becoming increasingly important as a data source for phonologists? One answer is that corpora help us bridge the gap between the

* Corresponding author. P.O.Box 310, 6500 AH, Nijmegen, The Netherlands
Email addresses: Mirjam.Ernestus@mpi.nl (Mirjam Ernestus), baayen@ualberta.ca (R.Harald Baayen).

[1] This research was supported by an European Young Investigator award to the first author.

analyst's conception of the data and the actual data themselves. Phonologists have formulated generalizations, some of which, as we know now, thanks to corpus-based research, do not do full justice to the data. Language appears to be much more complex than is generally assumed and this complexity is important for theories of phonology as well as for theories of speech production and comprehension. By way of example, we discuss a number of corpus studies on assimilation, intonation, and language change.

Our first example concerns regressive voice assimilation in Dutch. There is broad consensus in the phonological literature that obstruents are obligatory voiced before /b/ and /d/ within prosodic words, including compounds (see, e.g., Booij, 1995; Wetzels and Mascar, 2001). Thus, the compound *we*/t/ + /b/*oek* 'law'+'book' is pronounced as *we*[db]*oek*. The exceptional combined presence of final devoicing, regressive voice assimilation, and progressive voice assimilation in Dutch has received considerable attention in the theoretical literature on the nature of the feature voice (privative or not) and the typology of voice (see, e.g. Lombardi, 1999; Zonneveld, 2007). However, the data are much less straightforward when we consider what speakers actually produce by investigating speech corpora. Ernestus et al. (2006) extracted all 908 word tokens that according to the literature should show regressive voice assimilation from the subcorpus of read speech in the Spoken Dutch Corpus (Oostdijk, 2000). Three phoneticians listened to the audio files and classified all of the obstruents as voiced or voiceless. Unexpectedly, only 43% of the clusters (instead of the predicted 100%) exhibited regressive voice assimilation. In 25% of the clusters progressive voice assimilation was observed, even though progressive assimilation is traditionally seen as impossible in these contexts. Thus, *wetboek* was sometimes also pronounced as *we*[tp]*oek*. Furthermore, no assimilation was observed for 20% of the data (*we*[tb]*oek*).

This is a first illustration showing that there can be a remarkable and disquietingly large gap between the received phonological wisdom and the actual data. This gap in turn questions the adequacy of phonological theories that build on the — supposedly — exceptional facts from Dutch. Of course, the corpus findings could be explained away by the assumption that the Dutch grammar only allows regressive voice assimilation, and that the observed cases of no assimilation and also those of progressive voice assimilation are due to performance factors. However, this would introduce an unsurmountable gap between phonological competence and phonetic reality, and effectively render phonological theories unfalsifiable.

As a second illustration, consider regressive place assimilation in English. The traditional wisdom holds that alveolar word-final stops (/n, t, d/) often assimilate to the place of assimilation of the following labial or velar consonant. As a consequence, *gree*/n b/*oat* is often pronounced in connected speech as *gree*[m b]*oat* (Gimson, 1970). A substantial amount of research in psycholin-

guistics has investigated the consequences of this assimilation process for the listener. Researchers have argued both in favor and against a role of perceptual compensation for assimilation and its role in language acquisition (e.g., Gaskell, 2003; Gow, 2001; Mitterer and Blomert, 2003).

Dilley and Pitt (2007) investigated regressive place assimilation in conversational English, using the Buckeye Corpus of Conversational Speech (Pitt et al., 2005). Regressive place assimilation was observed relatively infrequently, much less frequently than standard descriptions would lead one to believe: on average only for 9% of their data. In contrast, deletion of the alveolar stop (32%), glottalization (15%), or unassimilated pronunciations (44%) were present more often. Again, we see that the phonologists' generalizations underestimated the complexity of the data. A phenomenon that is relatively easy to observe with minimal training in phonetics, assimilation of place, made it into the standard literature, even though it is infrequent in everyday speech. Phonological processes that are much more common in the same phonological environment went unnoticed until Dilley and Pitt's careful survey of what people actually say.

An example from the domain of intonation comes from Dainora (2001). Dainora studied downstep in American English on the basis of the Boston University Radio News (Ostendorf et al., 1995). Downstep refers to the phenomenon that during a sequence of high tones, the last tones may show a somewhat lower fundamental frequency, which is annotated with an exclamation mark (!H* versus H*) in Tones and Break Indices (Pierrehumbert, 1987).

Do high and downstepped high tones represent two fundamentally different categories? If so, we would expect that the frequency distance between two successive high tones (H*H*) would be smaller than the distance between a high tone and a following downstepped high tone (H*!H*). On average, there is indeed such a difference. Dainora (2001), however, pointed out that the distribution of the two frequency distances appear to form one single normal distribution, with the distances between successive high tones forming the distribution's left half and the distances between high and downstepped high tones its right half. It is not the case that we have two disjunct normal distributions, one for the H*H* distances and one for the H*!H* distances. This suggests that we should not interpret !H* as a separate category in its own right, since it forms one natural continuum with H*. Instead, !H* is a marker of where the lower variants of H* occur.

Our final illustration concerns the study of rhoticity in New Zealand English by Hay and Sudbury (2005). In many dialects of English, postvocalic /r/ has been lost before consonants, and word-final /r/ is only pronounced before vowel-initial words (*car* versus *ca*[r] *alarm*). In addition to this linking /r/, these non-rhotic dialects may have intrusive /r/, which appears between vowel-

final and vowel-initial words, as in *ma r and pa*. The phonological literature offers several accounts of the loss of rhoticity and the rise of linking and intrusive /r/. One theory holds that in a first stage postvocalic /r/ was lost, except in linking positions. Linking /r/ was subsequently interpreted as a sandhi-process, which gave way to intrusive /r/ (Vennemann, 1972). Other researchers have argued that in non-rhotic dialects, linking /r/ spread to new words by reanalysis on the part of the listener, and that both linking /r/ and intrusive /r/ are underlyingly present (Harris, 1994). Hay and Sudbury (2005) investigated the loss of rhoticity and the rise of linking and intrusive /r/ on the basis of a diachronic corpus of New Zealand English (Gordon et al., 2007). They found that the first generation of New Zealanders was still partly rhotic, in contrast to what is generally assumed. More surprisingly, some of these New Zealanders also showed intrusive /r/, which shows that the complete loss of preconsonantal /r/ was not necessary for the rise of intrusive /r/ (in contrast to the first theory). Furthermore, the data show that intrusive /r/ and linking /r/ are clearly different phenomena, as intrusive /r/ is less frequent than linking /r/, and linking /r/ appears more often in high-frequency collocations and morphologically complex words, whereas intrusive /r/ is seldom found in these contexts.

All these studies clearly show that speech corpora are substantially broadening the empirical scope of phonological research. Corpora show that many well-established basic facts that constitute a kind of canon feeding both phonological theory and psycholinguistic theories involve substantial simplifications that do not do justice to the variability and the range of phenomena that characterize actual speech.

2.2 Discovering new facts

Corpora are also becoming increasingly important as a data source for phonologists because they reveal new facts of which we did not know that they were right there in our own languages. It is difficult to pay attention to the details of the acoustic signal, when we are listening to our own language, since in normal language use the focus of attention is on content instead of form. This is especially so when listening to casual speech. As a consequence, we know very little about the fine phonetic detail of words in fast, unscripted speech. Such details are relevant for phonological theory, however, as they constitute an intrinsic part of speakers' knowledge of their language.

Take for example the pronunciation of homophones, such as *time* and *thyme*. It is generally assumed that homophones have exactly the same pronunciation, and differ only in meaning. This view has informed the theory of speech production developed by Levelt and colleagues (Levelt, 1989; Levelt et al.,

1999). In this theory, *time* and *thyme* have separate conceptual and syntactic representations, but share the same word form representation. In this model, there is no way in which the difference in meaning between *time* and *thyme* can be reflected in speech. Yet this is exactly what Gahl (2008) observed. Gahl analyzed roughly 90,000 tokens of homophones in the Switchboard corpus of American English telephone conversations. She found that words with a high token frequency, such as *time*, tend to have shorter realizations than their low-frequency homonyms, such as *thyme*, even after having controlled for factors such as speech rate and orthographic regularity. More in general Bell et al. (2003), Aylett and Turk (2004), and Pluymaekers et al. (2005b) all document, on the basis of speech corpora, shorter durations of segments, syllables, and words if these linguistic units or their carriers are of a higher frequency of occurrence. Such differences in fine phonetic detail must therefore be accounted for in linguistic theories and in psycholinguistic models of speech production.

An important phenomenon that can only be well studied on the basis of speech corpora is reduced speech. Well-known by now is the phenomenon of t-deletion (e.g., Browman and Goldstein, 1990), which has been studied extensively in sociolinguistics (e.g., Guy, 1980; Neu, 1980). Recent research has shown, however, that reduction in everyday speech is much more pervasive than the classical example of t-deletion would suggest. In addition to /t/, many other segments are prone to deletion, and deletion is not restricted to single segments, but may affect complete syllables. For instance, English *ordinary* is often pronounced as [ɔnri], *because* as [k^{h}z], and *hilarious* as [hlɛrɛ] (Johnson, 2004). Johnson's counts, based on the Buckeye corpus, show that some form of reduction characterizes no less than 25% of the words in colloquial American English. An example from Dutch illustrates the wide range of possible pronunciations a word may have: *natuurlijk* 'of course' may be pronounced not only in its canonical form [natyrlək], but also as [nətyrlək], [natylək], [ntylək], [nətyk], [ntyk], [ndyk], [tylək], and [tyk], among others (Ernestus, 2000). Similar observations have been made by Kohler for German (see, e.g., Kohler, 1990).

These reduction processes might be argued to be phonetic variation and outside the domain of inquiry of phonology. However, what segments reduce and the extent to which they reduce seems to be subject to a variety of intrinsically phonological constraints. For instance, a high degree of reduction is observed only for words without sentential accent in utterance medial position (e.g. Pluymaekers et al., 2005a,b). Sometimes, reduction is made possible by prosodic restructuring (Ernestus, 2000). Furthermore, although some phonotactic constraints that govern unreduced speech are relaxed for reduced speech, reduced speech nevertheless remains subject to many phonological and phonotactic constraints.

In turn, reduction provides information about phonological structure in casual

speech. An interesting example is the reduction of *don't* in American English. On the basis of 135 tokens of *don't* from a corpus of conversational American English, Scheibman and Bybee (1999) showed that *don't* may be realized with schwa, but only after the word that most frequently precedes *don't*, that is, after *I*. The presence of *I* is more important than the identity of the word following *don't*, even though reduction is also more likely and greater if this following word is more frequent after *don't* (e.g., *know*, *think*, *mean*). These data suggest that there is a tighter cohesion within *I don't* than within, for instance, *don't know* or *don't think*. This is exactly the opposite of what would be expected given the syntactic structure of these phrases, which group together the two verb forms rather than the pronoun and the first verb. This corpus-based research thus supports earlier observations on possible syntax-phonology mismatches, which led to the development of Prosodic Phonology (e.g., Nespor and Vogel, 1986).

As a final example of how corpora can reveal new facts, we mention the study of endangered languages. Collecting data from native speakers of minority languages without a tradition of literacy is often difficult if not impossible. For endangered minority languages, speakers tend to be old, monolingual, and not used to carry out tasks that require metalinguistic skills. Fortunately, story telling avoids such experimental problems, and corpora of recorded stories or dialogues may provide valuable information for the phonologist. Russell (2008) studied a corpus of oral narratives in Plains Cree. He investigated two vowel sandhi processes. He measured the formants and durations of some 450 sequences of /a#o/ that may be produced as [o:], and showed that this sandhi process is gradient and probably results from gestural overlap. The more specialized, possibly morphosyntactically governed, coalescence of /a+i/ or /a:+i/ to [e:] (some 250 tokens), in contrast, appeared to be more categorical. Data such as these raise the theoretical question whether gradient sandhi processes are part of phonology or of phonetics.

2.3 Understanding phonology in its wider context

The role of discourse and pragmatics in the grammar of pronunciation is becoming a more and more important field of research. An example is the study by Fox-Tree and Clark (1997). These researchers investigated the pronunciation of the definite article *the* in a corpus of spontaneous conversations. Traditional wisdom holds that the vowel of *the* is pronounced as [ə] before consonant-initial words and as [i] before vowel-initial words. Fox-Tree and Clark showed that speakers also use the realization with [i] in non-fluent speech when they are dealing with a problem in production, ranging from problems with lexical retrieval to problems with articulation. By using [i], speakers may signal that they would like to keep the floor. The same discourse effect has

been observed by Local (2007) for the realization of English *so*. On the basis of a survey of tokens of *so* extracted from corpora of spontaneous speech, Local shows that this word is reduced less when speakers want to keep the floor. It is more reduced and trails off when *so* marks the end of a turn. Such subtle use of phonetic cues is part and parcel of the grammar of a native speaker of English.

Other types of pragmatic function may affect pronunciation as well. Plug (2005), for instance, discussed the Dutch word *eigenlijk* 'actually, in fact', and documented, on the basis of a corpus of spontaneous speech (Ernestus, 2000), that this word is more reduced when it signals that speakers provide information which contrasts with information that they provided previously in the discourse. If tokens of *eigenlijk* introduce information that contradicts the presuppositions attributed to the listener, they tend to be less reduced.

Corpora have also been used to study phonological variation across social groups. Keune et al. (2005), for instance, investigated degree of reduction in Dutch as a function of speakers' social class, gender, age, and nationality (Belgium versus the Netherlands) on the basis of the Spoken Dutch Corpus (Oostdijk, 2000). The data showed a difference between men and women (with women reducing less) and differences between social classes (but only in Belgium). Furthermore, while there was on average more reduction in the Netherlands than in Flanders, degrees of reduction varied strongly with individual words. Thus, whereas *natuurlijk* 'of course' reduces more often in the Netherlands, other words with the same morphological structure, such as *waarschijnlijk* 'probably', show very similar degrees of reduction across the two countries. These differences between men and women and between Flanders and The Netherlands suggest that reduction is not just driven by articulatory processes but is in part a cultural phenomenon. Phenomena such as these raise questions about how phonological theory should account for variation in the grammars of different groups of speakers in the larger speech community.

3 Using speech corpora

3.1 An overview of speech corpora

Speech corpora are a relatively recent data source compared to corpora of written language. Traditionally, phoneticians and phonologists based their analyses on incidental observations and carefully designed experiments. Experiments have the advantage that they offer complete control over the materials. Words, phonemes, or phrases can be placed in exactly the right contexts and can be elicited in soundproof environments, free from background noise. Experiments,

however, are not without disadvantages. The amounts of data gathered tend to be small and typically cannot be re-used for different purposes. Moreover, speech styles elicited in the context of experiments tend to be formal and not spontaneous, and materials are presented in isolation or in small, often artificial, contexts. To complement experimental research, the last decades have witnessed the development of several speech corpora designed specifically for spoken (American) English and Dutch. We discuss some of the most important ones, stressing the differences in speech type and sound quality.

An important early speech corpus, the TIMIT corpus of read speech (Fisher et al., 1986) [2], provides the data of what can be regarded as a large production experiment. TIMIT sampled read speech (6300 sentences) from 630 speakers from several dialect regions of the United States. Two sentences were constructed to elicit as many differences between dialects as possible. Further sentences were constructed to provide a good coverage of phone pairs. A third set of sentences was sampled from existing sources to add to the diversity of sentence types and phonetic contexts. This corpus was designed and has been used extensively for the development of Automatic Speech Recognition systems.

A few years later, the HCRC Map Task Corpus [3] was published (Anderson et al., 1992). It provides a set of 128 dialogues (18 hours of speech) that were experimentally elicited with the Map Task. In this task, the two speakers in a dialogue are provided with a map that the other cannot see. One speaker has a route marked on her map, and has to guide the other speaker such that she reproduces this route on her own map. The crucial manipulation in this experiment is that the two maps are not identical, which forces speakers to engage in extensive discussions in order to complete their task. This leads to (the repetition of) specific words (especially of the missing landmarks), corrections, questions, and so on with a high probability. For instance, by annotating a landmark picture as *vast meadow*, Anderson and colleagues targeted t-deletion. All dialogues in the HCRC Map Task Corpus are transcribed and annotated for a wide range of behaviors including gaze. Map Task corpora have also been built for many other languages, including Italian, Portuguese, Czech, Japanese and Dutch.

In contrast to TIMIT and the HCRC Map Task Corpus, the speech sampled in the Switchboard corpus (Godfrey et al., 1992) [4] was under no experimental control whatsoever. This corpus comprises some 2430 telephone conversations of on average 6 minutes involving speakers who did not know each other. In all, the corpus consists of 240 hours of recorded speech with about three million

[2] `http://www.ldc.upenn.edu/Catalog/readme_files/timit.readme.html`

[3] `http://www.hcrc.ed.ac.uk/maptask/`

[4] `http://www.ldc.upenn.edu/Catalog/readme_files/switchboard.readme.html`

word tokens, produced by 500 speakers, both males and females, from all major dialects of American English. The corpus is fully transcribed, and each transcript is accompanied by a time alignment file which provides estimates of the beginning and end of words. Detailed information about the speakers is also available, including age, sex, education, current residence and places of residence during the formative years.

More recently, a corpus of spontaneous conversations has become available with a high-quality acoustic signal, the Buckeye Speech corpus (Pitt et al., 2005) [5] . Data were collected in a quiet room with head-mounted microphones for 40 speakers (20 men, 20 women, cross-classified by age) from Columbus Ohio. Each speaker was interviewed for one hour, leading to a corpus of some 300,000 words. Conversations were orthographically transcribed and phonemic transcriptions were obtained with the help of automatic speech recognition software. Time stamps are available for each of the phones.

Ernestus (2000) compiled a corpus of 15 hours of conversational Dutch with 10 pairs of speakers. She selected the speakers for each pair on the criterion that they knew each other very well, in the hope that they would feel free to engage in spontaneous and lively discussion, even in a sound-proof booth, with a separate microphone for each speaker. A recording session consisted of two parts. During the first part, the speakers talked freely about all kinds of subjects. Conversations were so free that a substantial amount of gossip was elicited. During the second part of the session, the speakers had to engage in role playing, enacting scripts in which they knew each other very well. The corpus has been transcribed orthographically, and a broad phonemic transcription is available that has been obtained using automatic speech technology (Schuppler et al., tted). This corpus has been a crucial source of information for the study of reduction in spontaneous Dutch. Similar corpora have recently been compiled for French, Spanish and Czech [6] .

An example of a recent corpus that provides speech from a wide range of spoken registers is the Spoken Dutch Corpus (Oostdijk, 2000) [7] . This corpus (in all some 9 million words, 800 hours of speech) includes a 2.1 million word subcorpus of spontaneous face-to-face conversations, a 900,000 word subcorpus of read speech (recorded books from the library for the blind), and a two million word subcorpus of telephone conversations. The spontaneous face-to-face conversations were recorded at people's homes with a single microphone, in order to optimize the likelihood of obtaining spontaneous speech. The drawback, however, is that the quality of the recordings is not optimal due to the presence of substantial background noise. The subcorpus of read speech, however,

[5] `http://buckeyecorpus.osu.edu/php/corpusInfo.php`

[6] For information about these four corpora: `http://mirjamernestus.ruhosting.nl/Ernestus/Corpora.html`

[7] `http://lands.let.kun.nl/cgn/ehome.htm`

provides very high quality sound files.

A corpus of a very different type is the ONZE corpus of New Zealand English (Gordon et al., 2007) [8] , one of the few diachronic speech corpora. It consists of three subcorpora: a collection of radio recordings of some 300 speakers born between 1851–1910, a collection of recordings of some 140 speakers born between 1890–1930, and a more recent and still growing collection of recordings of speakers born between 1930–1984. All sound files come with orthographic transcriptions.

Ideally, speech corpora would be paired with video, allowing researchers to investigate the roles of gesture, gaze direction, facial expression and so in spontaneous speech. An example of such a recent multimodal corpus is the IFA Dialog Video corpus, developed by van Son and Wesseling [9] . This corpus has recordings of maximally 15 minutes for some 50 speakers of Dutch, with orthographic transcriptions, automatically derived word and phoneme alignment, part-of-speech labeling, and annotations for gaze direction. An audiovisual corpus of read speech for English is reported by Hazen et al. (2004).

3.2 Transcriptions in speech corpora

A collection of just speech files does not constitute a speech corpus. Speech corpora make the audio data accessible by means of transcriptions and links between the transcriptions and the speech files. The most basic transcription is a straightforward orthographic transcription, which serve the function of providing a search heuristic for accessing the speech files. Some corpora also provide phonological or phonetic transcriptions. Obtaining reliable phonological or phonetic transcriptions, however, is a non-trivial enterprise.

One possible procedure is to base the transcriptions on acoustic measurements. This is an option if the features to be transcribed have obvious correlates in the acoustic signal. Most features, however, such as the voice of obstruents, are cued by different aspects of the acoustic signal (e.g., the duration of the obstruent, the duration of the preceding vowel, the presence of vocal fold vibration and so on). When the relative contributions of the different aspects to the overall percept are not well known, and when they may vary across speakers and registers, transcriptions based on (automatically obtained) acoustic measurements are infeasible.

Transcribing utterances by ear, however, is also not a trivial task, as it requires great concentration and even then remains error prone. Moreover, human tran-

[8] `http://www.ling.canterbury.ac.nz/onze/`

[9] http://www.fon.hum.uva.nl/IFA-SpokenLanguageCorpora/IFADVcorpus/

scribers tend to be influenced by their expectations, based on the words' pronunciations in clear speech, spelling, the phonotactics of the language, and so on (e.g., Cucchiarini, 1993). Vieregge (1987, 9) even argues that human transcriptions are influenced by the transcribers' expectations without exception, and are never objective reflections of reality. Along the same lines, Keating (1998) suggests that pronunciation variability is necessarily confounded with transcription variability in studies with human transcribers.

Expectations play an important role especially when the acoustic signal is less clear, for instance due to background noise. Speech may also be less clear because speakers reduced their articulatory effort and produced smaller and overlapping articulatory gestures. In such casual speech, the reduced forms may differ substantially from their unreduced counterparts. Yet transcribers will tend to hear the reduced forms as unreduced.

Since transcribing is such a difficult and subjective task, listeners often disagree about the correct transcription. Notoriously difficult is deciding on the presence versus absence of sonorant segments (such as schwa and liquids) and about segments' voice specifications. For instance, Ernestus (2000) reported that her three transcribers disagreed about the presence versus absence of the first vowel of the word *natuurlijk* 'of course' for 58% of the 274 tokens, while they disagreed on the voicing of intervocalic plosives for 15% of the more than 2000 cases. Similar figures have been reported by Ernestus et al. (2006), Coussé et al. (2004), and Pitt et al. (2005). Disagreement arises even when listeners do not provide detailed transcriptions but classify word forms roughly into predefined categories of "no to low reduction" or "high reduction" (Keune et al., 2005).

What to do with tokens for which transcribers disagree? One obvious solution is not to incorporate them into the analyses. If the number of problematic tokens is low, this is feasible. However, when there are many problematic cases, the number of available data points may decrease substantially, and as a consequence, the power of subsequent statistical analyses as well. Furthermore, the problematic data points may all belong to a small number of classes (e.g., high vowels, or segments preceeded by liquids, or segments in unstressed syllables) which may provide crucial information and hence should not be excluded from the analysis a priori. In fact, such data points may be of theoretical interest, for instance, they may be indicative of an ongoing sound change (Saraçlar and Khudanpur, 2004).

Another way of dealing with disagreements is to ask transcribers to listen to the problematic tokens again (and again) and see whether they are willing to change their classifications. This method does not necessarily lead to more accurate transcriptions, however, since the transcribers, when listening for the second time, know each other's classifications, and the classification which is

eventually accepted may not be the best one, but the one obtained from the most confident transcriber. Finally, note that even when listeners are in full agreement, this does not necessarily imply that they provide the correct transcriptions: The transcribers may all be led astray by the same expectations.

Both these procedures to handle disagreements face yet another problem, since a high degree of disagreement may indicate that the phenomenon under investigation is continuous rather than categorical. For instance, when studying reduction or voicing, segments can be partially deleted or partially voiced, and requesting raters to give absolute judgments may not do justice to the complexity of the data. Below, we will mention yet another way to deal with inconsistent transcriptions which is based on the use of statistics and avoids this problem in a principled way.

To what extent do automatic speech recognition (ASR) systems provide a solution for this problem? An obvious advantage is that the slow cumbersome, and subjective work by human transcribers is replaced by a computer algorithm that will always yield the same results. Unfortunately, ASR systems need to be trained on phonetically transcribed materials and as a consequence their accuracy depends heavily on the quality of these human made training transcriptions. Several experiments have shown that ASR transcriptions generally show a somewhat lower agreement with human transcribers than human transcribers among each other (e.g., Van Bael et al., 2006; Wester et al., 2001). ASR systems have difficulties especially with those classifications that are notoriously difficult also for human transcribers (presence versus absence of schwa, liquids, etc).

The field of ASR systems is still in full development. One interesting new direction is the replacement of phonemic transcriptions by continuous transcriptions of articulatory based features (e.g., King and Taylor, 2000; Ten Bosch et al., 2006). The choice of the set of features is largely inspired by both the theory of distinctive features (Chomsky and Halle, 1968) and the gestural theory of speech production (Browman and Goldstein, 1992). This type of ASR systems may prove especially useful for the study of fine phonetic detail.

3.3 Analyzing corpus data

Corpus data should be used responsibly. Corpora are not build for looking up some incidental examples, however interesting they may be. We all too easily find examples that fit the hypothesis driving our research, and we all too easily overlook examples that do not fit our theory. Moreover, it has been very well documented by now that speakers show probabilistic behavior leading to (varying degrees of) intraspeaker variation. Finding one or two tokens

of a word displaying the phenomenon of interest (e.g., schwa deletion) does not provide us with information about the way the speaker normally realizes the word. These two tokens may just represent exceptional pronunciations. Furthermore, we also have to investigate where the phenomenon under study could be expected but did not occur, since our theories should account for these facts as well. As in any other domain of scientific inquiry, we have to survey all potentially relevant data.

Corpus research thus necessarily implies the inspection of very large data sets, and for this statistical analysis is indispensable. In what follows, we give a brief introduction to a technique that is of particular relevance for the analysis of corpus data, linear mixed-effects modeling (Baayen, 2008; Jaeger, 2008). We illustrate this general modeling tool using a small, simplified, constructed data set that mirrors part of the structure of the data set of Hay and Sudbury (2005) on postvocalic /r/ in New Zealand English that we discussed above.

Table 1
Counts of non-rhotic and rhotic variants (in that order) for four subjects (S1, S2, S3, S4) for 15 word pairs (W1 ... W15) with varying log frequencies for the second word of the pair (`FreqWord`) and for the complete word pair (`FreqWordPair`).

words	FreqWord	FreqWordPair	S1	S2	S3	S4
Pair1	4.69	0.26	0 4	0 4	0 5	0 1
Pair2	4.25	0.26	0 4	3 1	1 2	0 3
Pair3	4.21	0.45	0 4	3 1	2 2	0 6
Pair4	4.56	0.34	0 4	2 1	1 2	0 6
Pair5	4.73	0.64	0 3	7 1	4 2	0 4
Pair6	3.04	0.25	0 2	2 0	4 1	1 4
Pair7	3.26	0.46	1 1	4 0	3 1	1 7
Pair8	1.46	0.17	3 3	4 0	1 0	1 3
Pair9	4.35	0.40	2 0	5 0	7 0	2 1
Pair10	4.24	0.26	0 4	1 3	0 3	0 3
Pair11	4.00	0.21	0 3	1 3	1 4	0 2
Pair12	4.99	0.03	0 4	1 2	0 1	0 2
Pair13	3.62	0.22	0 5	1 0	2 5	0 1
Pair14	2.78	0.30	0 5	5 2	2 3	0 4
Pair15	3.91	0.54	0 5	3 1	4 3	0 4

Consider Table 1, which lists for four speakers (`S1, S2, S3, S4`) and for fifteen word pairs (`Pair1` to `Pair15`) the log-transformed lexical frequency of the second word (`FreqWord`), the log-transformed frequency of the word

pair (`FreqWordPair`)[10], and the number of times /r/ was observed absent and present for each of the four subjects for each word pair. Given these observations, we ask ourselves the following questions.

(1) Does the probability of the presence of /r/ decrease with `FreqWord`?
(2) Does this probability increase with `FreqWordPair`?
(3) Does the probability of /r/ vary between speakers?
(4) Does the probability of /r/ vary between word pairs?

To answer these questions, we fit a regression model to the data with as predictors `FreqWord`, `FreqWordPair`, `Speaker`, and `Word Pair`. Our dependent variable requires special care. Each observation in our dataset has one of two values: present (success) or absent (failure). What we are interested in is the *probability* of an /r/ given specific values for our predictors. One possibility is to analyze the percentages of successes. Percentages (and the corresponding proportions or probabilities), however, have mathematical properties that make them unsuited for regression analysis (see, e.g., Harrell, 2001; Jaeger, 2008, for detailed discussion). The most important one is that percentages are bounded between 0 and 100 (and proportions and probabilities between 0 and 1). A commonly used solution is to model the logarithm of the odds ratio of the successes and failures L_{ij} for Speaker i and Word pair j:

$$L_{ij} = \log \frac{p}{1-p}, p = \log \frac{\#\text{successes}}{\#\text{failures} + \#\text{successes}} \tag{1}$$

The log odds ranges from minus infinity to plus infinity. When there are more successes than failures, the log odds is positive, when the number of successes is the same as the number of failures, it is zero, and when the number of successes is smaller than the number of failures, it is negative. Given a regression model for the log odds, the predictions of the model on the probability (rather than the log odds) scale can be obtained using the relation

$$P_{ij} = \frac{1}{1 + e^{-L_{ij}}}. \tag{2}$$

In what follows, we model the log odds ratio L_{ij} for Speaker i and Word Pair j as a function of baseline odds ratio β_0 (the intercept), adjustments b_i and b_j to this baseline for Speaker i and Word Pair j, together with coefficients β_1 and β_2, which represent the effects of the frequency of the second word and the frequency of the word pair. These two coefficients represent slopes, the increase in rhoticity for a unit increase in frequency.

$$L_{ij} = (\beta_0 + b_i + b_j) + \beta_1 \text{FreqWord}_j + \beta_2 \text{FreqWordPair}_j + \epsilon_{ij}. \tag{3}$$

[10] For frequencies, log transformations are required in order to reduce the enormous skew which is normally present in the distributions of frequencies.

When we fit this linear mixed-effects model to the data in Table 1, we find that the slope for the frequency of the word pair is 6.6, indicating that the likelihood of rhoticity increases as this frequency increases. The slope for frequency of the second word is estimated at -1.4, which means that rhoticity is less likely as this frequency increases. The model also provides detailed information about how the likelihood of rhoticity varies from speaker to speaker, and from word pair to word pair. For instance, S4 is the least rhotic speaker of the four and Speaker S2 the most rhotic. Of the word pairs, `Pair 9` is realized most often with [r], for `Pair 1` the reverse holds. Tests of significance confirm that the effects of the two frequencies are significant, and that there is significant variability between speakers and between word pairs.

Of course, the real data studied by Hay and Sudbury are much more complex, and required inclusion of predictors such as speaker's sex (men turned out to produce /r/ more often than women) and the nature of the preceding and following vowels (front vowels disfavored [r]). Such variables can be added straightforwardly to the statistical models.

Our constructed example does not do justice to the non-randomness and non-independence in natural discourse. Pickering and Garrod (2004), for instance, call attention to various priming effects in dialogue. How a given word is actually realized often depends on how that word, or similar words, were realized in the preceding discourse. This non-independence requires special care in statistical analysis (Rietveld et al., 2004). In mixed-effects models, it is often possible to bring such dependencies under control with the help of longitudinal variables (De Vaan et al., 2007; Balling and Baayen, 2008). For instance, the number of times a given word appeared with a given realization in the preceding discourse can be added as a predictor to the model.

Above, we discussed the problem that transcribing speech is a difficult and subjective task that often leads to disagreement among transcribers. Hay and Sudbury (2005) had the same analyst transcribe the same materials twice with a couple of months intervening. They included in their analysis only those cases where on both occasions the same judgment was made, and thus accepted data loss. A solution explored by Ernestus et al. (2006) makes use of mixed-effects modeling and considers as dependent variable the individual classifications produced by the raters, but adds the identity of the rater as an additional factor to the model. The idea is to predict what individual listener-raters think they heard instead of aggregating over listener-raters to compute a verdict of what was actually said. The regression model determines the role of the different predictors (e.g., lexical frequency, phonological properties of the word) as well as the influence of the different listener-raters for the classifications. In other words, it is left to the regression model to handle disagreements between listener-raters.

3.4 Generalizing data to different speakers

We are now in the position to address the issue of how corpus-based statistical analyses relate to the theory of grammar. One question is phrased by Newmeyer (2003, p. 696) as follows.

> The Switchboard Corpus explicitly encompasses conversations from a wide variety of speech communities. But how could usage facts from a speech community to which one does not belong have any relevance whatsoever to the nature of one's grammar? There is no way that one can draw conclusions about the grammar of an individual from usage facts about communities, particularly communities from which the individual receives no speech input.

Recall that the Switchboard Corpus sampled speakers from all major varieties of American English. At first sight, it does indeed seem highly implausible that data from a set of speakers of variety A would help us to understand the grammar of an individual from variety B. However, mixed effects modeling offers us the means for carefully teasing apart what is common to all speakers and what is specific to a particular dialect. Let's return to our hypothetical data on /r/ sandhi in New Zealand English. Suppose we have not just 4 speakers, but 40 speakers from dialect A, 30 speakers from dialect B, and 50 speakers from dialect C. (Dialects D, E, F, ... are not sampled.) The model that we would now fit to the data would include dialect as a second random-effect predictor modifying the intercept (b_k).

$$L_{ij} = (\beta_0 + b_i + b_j + b_k) + \beta_1 \text{FreqWord}_j + \beta_2 \text{FreqWordPair}_j + \epsilon_{ijk}, \quad (4)$$

The adjustment b_k to the intercept for Dialect k informs us about the extent to which Dialect k differs from the language as a whole. Similarly, the adjustments b_i and b_j to the intercept for speaker i word pair j give us further information about the individual differences in the rate of occurrence of postvocalic [r] for the speakers and the word pairs. The coefficients β_0, β_1, and β_2 estimated by such a model tell us what is common to all dialects and to all the different word pairs and speakers within these dialects. Crucially, information of speaker X from dialect A contributes to our estimates of these β-coefficients, and therefore to our understanding of the grammar of speaker Y from dialect B. In other words, our mixed-effects model helps us to separate out the role of Dialect, the role of the individual Speaker, and the role of the shared grammar.

There are many other dimensions of variation that we will need to consider in our corpus-based models. One such dimension is register, contrasting, for instance, read speech with scripted speech, telephone conversations, and face-to-face conversations. Other dimensions are time, social class, education. There

are currently no speech corpora that properly sample across all these dimensions. As a consequence, conclusions based on corpus data are by necessity conditional on the input data.

4 Abstractionist and exemplar-based models

Corpus-based research has made more than obvious that pronunciation variation is inherent to natural language. We have also seen that statistical models help clarify which patterns are characteristic of a language (variant) and which are of a more idiosyncratic nature. Moreover, such models indicate which factors (sociolinguistic, phonological, morphological, etc.) help explain this variation. All this information helps the researcher to develop better linguistic and psycholinguistic models.

Broadly speaking, present-day linguistic and psycholinguistic models can be classified along a continuum with at one endpoint purely abstractionist models and at the other endpoint purely exemplar-based models. These two types of models differ in their views of the nature of linguistic generalizations and the amount of detailed knowledge that is assumed to be available in the mental lexicon.

4.1 The nature of linguistic generalizations

Early generative phonology and direct successors, including Optimality Theory (e.g., Chomsky and Halle, 1968; McCarthy and Prince, 1993), are typical examples of purely abstractionist models. They assume that generalizations over the language, such as Final Devoicing and the position of word stress, are stored independently from the words in the mental lexicon in the form of abstract representations. These abstract generalizations can be applied directly to new words, such as loan words, without reference to the words from which these generalizations were previously deduced during learning. For instance, the English verb *save* is pronounced in Dutch, a language with Final Devoicing, as [sef]. According to abstractionist theories, this is due to the application of a rule of Final Devoicing that exists independently of the data. In machine learning, learning strategies that build on abstract generalization are called eager or greedy learning strategies (Daelemans and Van den Bosch, 2005).

Purely exemplar-based models, on the other hand, do not posit generalizations in the form of abstract rules that are stored independently from the individual words. Generalizations are extracted from the exemplars only when they are needed (see e.g., Semon, 1923, the first to discuss exemplar-based mod-

els). The English verb *save* is pronounced as [sef] in Dutch, because on-line checking of its nearest phonological neighbors in the Dutch lexicon ([lef] 'live', [nef] 'nephew', [xef] 'give', ...) reveals overwhelming and in fact exceptionless support for the /f/. Exemplar-based models are thus characterized by lazy learning: generalization is delayed until a query is made to the system. The reason for this delay is, as we shall see below, that generalization accuracy is optimal when all exemplars ever encountered are available for consideration. Forgetting rare, low-frequency forms is harmful.

The "on-line checking" in exemplar-based models involves the simultaneous evaluation of all relevant exemplars in memory. This imposes a large computational burden. Two different approaches have been explored. Skousen (2002) has developed algorithms for his computationally highly demanding theory of analogical modeling of language that anticipate the advent of quantum computing. Even for computationally less demanding algorithms, measures have to be taken to speed up processing. In machine learning, it is common to use tree-based memory structures that may afford compression rates of 50% or more, and hence allow shorter searches and faster retrieval of the nearest neighbors (see, e.g., Daelemans and Van den Bosch, 2005, p.47). To increase the speed of evaluation at run-time even more, generalizations can be built into the tree-based memory, but, as we shall see below, this tends to go hand in hand with a decrease in the quality of the generalizations of the model (Daelemans and Van den Bosch, 2005, p.67–73). In short, the hybrid solution trades quality for speed. We will return to this hybrid approach below.

In what follows, the focus of our discussion will be on models assuming exemplars at some linguistic level, as purely abstractionist models are presented in detail in the other chapters of this handbook. Furthermore, due to limitations of space, only the main properties of different types of models are discussed. We also challenge the traditional conception of phonology as a subdiscipline of pure linguistics. Many phonologists working within abstractionist frameworks view their task as developing a theory of just the declarative knowledge one must know as a speaker of a language. We see many problems with such a conception of the field. First, it is unclear what data fall under the 'jurisdiction' of the phonologist. In the preceding section, we have reviewed a wide range of phenomena that illustrate subtle aspects of the knowledge that speakers have about the sound structure of their language. Some of these phenomena can be explained with the theoretical apparatus of traditional phonology, others, however, will require this field to broaden its scope. Second, science in the 21st century is increasingly becoming an interdisciplinary endeavor. The likelihood that phonology will make significant advances while dismissing recent achievements in other fields, be it computational linguistics, psycholinguistics and neurolinguistics, or phonetics, as irrelevant, is in our view unnecessarily small.

4.1.1 The importance of many exemplars

Purely abstractionist models assume that a relatively small sample of exemplars is sufficient for developing robust generalizations. In this approach, once a generalization has been established, further incoming evidence has no role to play, and is disregarded. By contrast, exemplar-based models assume that generalizations are most precise when based on as large an instance base as possible. Importantly, several studies have shown that generalizations based on all available evidence are indeed better predictors of speakers' behavior (see, e.g., Daelemans et al., 1999). By taking more examples into account, more specific generalizations become possible, enabling exemplar-based models not only to replicate the general regularities captured by traditional grammars, but also to formulate more local, detailed regularities. Such more restricted regularities are important because they allow us to predict for which words speakers are uncertain, and to predict forms that speakers produce even though these forms are not expected under an abstractionist account. Thus Skousen's Analogical Modeling of Language not only correctly predicts that the English indefinite article tends to be *a* before consonants and *an* before vowels, but also simulates speakers' behavior in tending to chose *a* for some vowel-initial nouns which are special due to the characteristics of the phonemes following the initial vowels (Skousen, 1989).

Similarly, we have shown that the traditional description of regular past-tense formation in Dutch is too simplistic (Ernestus and Baayen, 2004). It is true that most verbal stems ending in a voiceless obstruent (before the application of Final Devoicing) are affixed with [tə] and all other stems with [də], but for some verbs speakers produce non-standard forms quite often (choosing [də] instead of [tə], or vice versa). The final obstruents of these verbs have voice-specifications that are unexpected given the other words ending in the same (type) of rhyme. For instance, the verb *dub* 'waver' is special in Dutch since it ends in a voiced bilabial stop, whereas the sequence short vowel - voiceless bilabial plosive is much more frequent (e.g., in *klap*, *stop*, *nip*, *step*, *hap*). In line with this local generalization, speakers often choose *te*, instead of *de* as the past-tense allomorph. Importantly, when speakers produce standard past-tense forms for these exceptional verbs, they need more time to select the correct past-tense allomorph than when producing standard past-tense forms for non-exceptional verbs. Past-tense formation in Dutch does not only obey the general high-level generalization formulated in traditional phonological models, but also more local generalizations within the words' sets of phonologically similar words.

As a final example we mention the work by Plag and colleagues on stress assignment in English compounds (Plag et al., 2007, 2008). Their comprehensive surveys revealed that traditional factors (such as argument structure and the semantics of the head noun) were only moderately successful in predicting the

position of stress. They obtained much better predictive accuracy by considering the distribution of stress positions in the modifier and head constituent families (the sets of compounds sharing head or modifier). For instance, street names involving *street* as their right-hand member pattern alike in having leftward stress (e.g., *Oxford street*, *main street*), whereas street names ending in *avenue* have rightword stress (e.g., *Fifth avenue*, *Maddison avenue*). Similar biases for left or right stress, although often less pronounced, are found across the lexicon for other constituent families. Their conclusions harmonize well with work on the interfixes in Dutch and German compounds (Krott et al., 2001) and on the semantic interpretation of compounds (Gagné, 2001).

Several models assuming abstract generalizations have incorporated the idea that generalizations should be based on many exemplars. Two of these have been computationally implemented: Stochastic Optimality Theory (Boersma, 1998; Boersma and Hayes, 2001), and Minimal Generalization Learning (Albright and Hayes, 2003). Stochastic Optimality Theory implements, unlike most other abstractionist theories, a continuous learning process in which stochastic constraints are continuously updated. The Minimal Generalization Learner constructs a large set of weighted rules that are learned during training. Once learning is completed, the rules are applied on-line during 'testing'.

As shown by Keuleers (2008), the Minimal Generalization Learner and TIMBL, are computationally equivalent, with TIMBL executing similarity-based reasoning at runtime, and the Minimal Generalization Learner executing previously learned weighted rules at runtime. This shows that at the computational level, abstractionist and exemplar-based models can be equivalent. In such cases, evaluation should be guided by how much insight and guidance the models provide given current theories across theoretical linguistics, computational linguistics, psycholinguistics, and cognitive science.

4.1.2 The productivity of generalizations

Purely abstract models assume that all generalizations are fully productive. They are assumed to apply across the board to any input that meets their input requirements. However, several studies have argued that a generalization's productivity depends on the number of exemplars in the lexicon supporting the generalization (e.g., Bybee, 2001). Regularities are in general more productive if they are supported by more exemplars. Thus, word-specific pronunciation variation, which is characterized by only little lexical support (e.g., only from the lexical item itself), tends to be unstable and to disappear in favor of variation shared with other, phonologically similar, words. Only a high frequency of occurrence can protect isolated words against regularization (e.g., Bybee, 2001).

Furthermore, generalizations based on words which are more similar are more productive than generalizations based on words that are less similar. A lesser degree of similarity has to be compensated for by a greater number of exemplars (and vice versa). Thus, a single exemplar can only affect a neighboring word if the two neighbors are already highly similar (Frisch et al., 2001).

In contrast to models assuming abstract generalizations, exemplar-based models are able to account for the effect of the number of exemplars and the similarity among the exemplars on degree of productivity. In these models generalizations are formulated by on-line checking of all exemplars. Each exemplar may contribute to the generalization based on its similarity. More exemplars and exemplars showing higher similarities may lead to stronger and therefore more productive generalizations. Skousen (1989), for instance, has incorporated these mechanisms in his Analogical Modeling of Language, by distinguishing sets of exemplars which differ in their influence based on their set size, their (phonological) distance to the target word, and also the consistency among the exemplars with respect to the outcome of the generalization (e.g., voiced versus voiceless for syllable-final obstruents in Dutch).

Note that it is important to carefully distinguish between *generalization* and *abstraction* (Daelemans & van den Bosch, 2005). Exemplar-based models and abstractionist models share the goal of *generalization*, of being able to predict the behavior of unseen cases, and to understand how this prediction follows from past experience. The crucial difference is how this goal is achieved. In purely abstractionist approaches, individual tokens (at a given level) are used to formulate abstract rules. Once the rules have been formulated, the individual tokens considered in formulating the rules are redundant, and discarded as theoretically unimportant. By contrast, exemplar-based approaches are driven by the conviction that every token counts, and that in order to achieve maximum prediction accuracy, it is essential to carefully consider the contribution of each exemplar. Thus, perhaps the most crucial difference between abstractionist and exemplar-based models is their very different evaluation of the role of human memory in language.

4.2 The content of the mental lexicon

Abstractionist models typically work with sparse lexicons, with as only exception in generative grammar the work of Jackendoff (1975). Once the linguistic generalizations of the language have been deduced from the input, the input words are no longer needed to support the generalizations. If they are morphologically complex and completely regular in all respects, they can even be removed from the lexicon, as they can always be recreated via the morpho-phonological generalizations. The lexicon can be as sparse as to contain only

lemmas (morphologically simplex forms, such as *tree* and *school*) and morphologically complex words that are semantically, morphologically, syntactically, or phonologically irregular (e.g., *children* and *juicy*). Regular morphological derivations and inflections are always derived by means of morphophonological generalizations (see, e.g., Kiparsky, 1982; Pinker, 1991).

This approach, advocated especially by generative grammar, implies that the form stored in the (mental) lexicon need not be phonotactically well-formed and identical to a form that occurs in the actual linguistic output. Take for instance regular plural nouns in Dutch, which consist of the noun stem and the suffix [ə] or [s]. The affixation with [ə] may lead to voice alternation of the stem-final obstruent, for instance, singular [hɔnt] *hond* 'dog' versus plural [hɔndə] *honden* 'dogs'. The [t] of [hɔnt] is predictable, since Dutch words cannot end in voiced obstruents (Final Devoicing), whereas the [d] of [hɔndə] is not (compare the plural [hɔndə] with the plural [lɔntə] 'matches'). Therefore, generative grammar is forced to assume that the stored form is /hɔnd/, from which both the singular (Final Devoicing) and the plural ([ə]-affixation) can easily be computed. This underlying form is however phonotactically illegal as a surface form (see, e.g. Booij, 1981; Wetzels and Mascar, 2001).

Exemplar models differ from abstractionist models in that the lexicon is viewed as a database containing huge numbers of exemplars (see, e.g., Bybee, 1985, 2001; Johnson, 2004). As it is difficult, if not impossible, to determine the relevance of abstract generalizations and exemplars in the lexicon, it is not surprising that many researchers have brought evidence from language processing into the debate. In what follows, we discuss evidence for exemplars at different linguistic levels: for regular morphologically complex words, for pronunciation variants of one and the same word, and for exemplars of individual acoustic/articulatory events.

4.2.1 Storage of regular morphologically complex words

An important finding from the psycholinguistic literature is that the processing of completely regular morphologically complex words is known to be affected by these words' frequencies of occurrence. For instance, Stemberger and MacWhinney (1988) demonstrated that speakers produce fewer errors for high frequency than for low frequency regular past-tense forms. Similarly, numerous studies have demonstrated that readers' and listeners' recognition times of regularly inflected and derived words in a wide variety of languages is affected by these forms' frequencies of occurrence (e.g., Baayen et al., 1997; Sereno and Jongman, 1997; Bertram et al., 1999; Baayen et al., 2008; Kuperman et al., 2008; Baayen et al., 2007). These form-specific frequency effects show that language users have detailed knowledge at their disposal about how likely specific forms are. Such detailed knowledge is totally unexpected from

the purely abstractionist perspective, especially when abstractionist models are projected straightforwardly onto language processing (see, e.g., Pinker, 1991), but harmonizes well with exemplar-based models.

Additional evidence for the storage of regular morphologically complex words comes from language change. Bybee (2001) discusses the historical lengthening of short vowels (accompanied by a change in vowel quality) in Dutch open syllables. This change resulted in morphologically conditioned pronunciation variation in several noun stems. Later, the change became unproductive. If the alternation had been completely governed by an abstract generalization stored independently of the relevant nouns, the loss of the generalization should have resulted in the disappearance of all the vowel alternations governed by that generalization. This, however, is not the case: Modern Dutch still shows the alternation for some words (e.g., *sch*[ɪ]*p* - *sch*[e]*pen* 'ship' - 'ships'), words which otherwise have a fully regular plural inflection. This can only be explained if it is assumed that the different forms in a word's paradigm become entrenched in lexical memory, irrespective of whether they are regular or not (see also, e.g., Tiersma, 1982).

The storage of large numbers of regular derivational and inflectional forms makes it unnecessary to posit, as in generative grammar, underlying representations that would differ from the words' actual pronunciations. If all forms of a paradigm are stored in a redundant lexicon, there is no need to assume that the stem's underlying representation contains all unpredictable properties. If both Dutch /hɔnt/ 'dog' and /hɔndə/ 'dogs' are stored in the mental lexicon, there is no need to assume that the morpheme for 'dog' is represented as /hɔnd/ with the unpredictable final /d/. Neither speakers nor listeners need to compute the plural [hɔndə] from the underlying lexical representation of *hond*, since either the plural is stored in the mental lexicon together with /hɔnt/, or the voice specification of the obstruent can straightforwardly be inferred from its nearest phonological neighbors (/vɔndə/ 'found', /mɔndə/ 'mouths', /mɑndə/ 'baskets' ...).

4.2.2 Storage of pronunciation variants

The wide pronunciation variation observed in speech corpora cannot be accounted for by the storage of just the canonical pronunciations of the words or word forms in the lexicon. The words stored have to be accompanied by information about their possible pronunciations. Abstractionist models assume phonological rules (or interactions of phonological constraints) which derive the possible pronunciations (during speech production) and deduce the stored representations from the observed realizations (during speech comprehension). For instance, a phonological rule of flapping specifies in which segmental (and probably social) contexts American English /t/ may be realized as a flap (e.g.,

in the word *butter*). Similarly a rule (possibly the same) specifies that a flap in American English maps on /t/ in lexical representations. This rule of flapping applies to hundreds of words, and therefore represents a true generalization over American English.

This account of pronunciation variation faces an important challenge. Many types of pronunciation variations are restricted to just a few words, instead of all words satisfying the structural description of the generalization, as in the case of flapping /t/. For instance, in Dutch, word-final /t/ can be absent in utterance-final position only in the word *niet* 'not', and word-final velar fricatives may be absent only in *toch* 'nevertheless' and *nog* 'still' (Ernestus, 2000). In general, we see that words are more reduced the higher their frequency of occurrence, which may lead to word-idiosyncratic pronunciation variation. In abstractionist models, word-specific pronunciations imply either word-specific rules or constraints, or the storage of several pronunciations for the same word (see, e.g. Booij, 1995). A question that arises in this context is how many different words have to show the same pronunciation variation for a generalization to come into existence.

Such questions are irrelevant for models that simply assume that each word is stored in the mental lexicon together with all its possible pronunciations. The representations of all these possible pronunciations might be abstract in nature (e.g., strings of phonemes), in which case the model is close to purely abstract models. Alternatively, these representations may be abstract labels for clouds of exemplars each representing one individual acoustic/articulatory event (see section 4.2.3). In this case, the model is more similar to a purely exemplar-based model. In both types of models, the Dutch word *niet* is stored with the pronunciations [nit] and [ni], which "explains" why this word may occur with and without [t] in all sentence positions. Importantly, these models account for word-specific pronunciation variation as well as for the productivity of alternations displayed by a wide range of words, such as /t/ flapping in American English.

Several studies have produced experimental evidence for the storage of at least some pronunciation variants. Racine and Grosjean (2002) showed that native speakers of French can well estimate how often a particular word is produced with and without schwa in spontaneous speech: The correlation between subjects' estimates of the relative frequencies and the relative frequencies observed in a speech corpus was $r = 0.46$. Apparently, speakers know the likelihoods of both pronunciation variants. In a purely abstractionist approach, it might be argued that this probability information is stored with the unreduced form and affects the likelihood of the application of a schwa deletion rule. This account implies that there must be some memory trace for the reduced form, albeit not instantiated in the form of a separate lexical representation, but in the form of a word-specific probability of schwa deletion. However, from a com-

putational perspective, this word-specific probability is difficult to distinguish from a separate representation in an exemplar-based model.

From an exemplar-based perspective, these facts would be captured by positing that the two variants are represented by two exemplars (or two clouds of exemplars) that may have different long-term probabilities of becoming active in speech comprehension or production. Connine and colleagues (for an overview see Connine and Pinnow, 2006) showed that the frequencies of pronunciation variants play a role in word recognition. Their study of the nasal flap as a pronunciation variant of /nt/ in American English showed that listeners recognize words pronounced with a nasal flap faster if these words are more often produced with a nasal flap instead of [nt] (Ranbom and Connine, 2007). This illustrates once again that language users are sensitive to the probabilities of pronunciation variants.

The assumption that all pronunciation variants of a word are lexically stored is not unproblematic. In Ernestus et al. (2002), we showed that listeners recognize reduced word forms presented in isolation with a higher accuracy the more similar they are to the corresponding unreduced forms. Thus, we found a strong positive correlation between the number of missing sounds and the number of misidentifications ($r = 0.81$). This strongly suggests that listeners recognize reduced pronunciations, spliced out of their contexts, by means of the lexical representations of the unreduced counterparts. This finding can only be explained within exemplar-based theory if we make the assumption that lexical representations are specified for the context in which they occur (see e.g., Hawkins, 2003). Reduced pronunciations would then be specified as "not occurring in isolation". This specification would also explain why the number of misidentifications was much lower when the reduced pronunciations were presented in their natural contexts instead of in isolation.

4.2.3 Storage of acoustic and articulatory tokens

The most extreme variant of exemplar-based models assumes that the mental lexicon contains all acoustic and articulatory tokens of all words that the language user has ever encountered (e.g., Johnson, 2004). The lexicon thus would contain millions of tokens of every word form, many of which hardly differ in their phonetic characteristics. The lexicon would therefore be very similar to a speech corpus itself. Tokens sharing meaning would then be organized in clouds of words (cognitive categories) and would be interconnected as in all other versions of exemplar-based theories. We will refer to this specific type of exemplar-based models as episodic models.

Episodic models differ in another crucial characteristic from the exemplar-based models described so far. They assume that all tokens are stored with

all their fine phonetic detail. In contrast, models allowing just one or a small number of lexical representations for every word, each reflecting a different pronunciation type, typically assume that lexical representations are built up from abstract symbols such as phonemes, allophones, or phonological features. Listeners are assumed to abstract away from the details of the speech signal that cannot be captured by these abstract categories. The tacit assumption is that these details would not be relevant for higher-level generalizations. The models discussed in the previous sections are thus closer to the endpoint of abstractionist models of the continuum than episodic models, which occupy the other extreme endpoint.

Lexical representations consisting of abstract symbols, such as phonemes, are problematic because the conversion of real speech into such abstract symbols, which includes the process of speaker normalization, has proven difficult to capture. For instance, the categorization of a sound as a certain phoneme (or allophone) is determined by many factors, including segmental context, the speaker's gender, and the listener's expectations (for an overview, see e.g., Johnson, 1997). Episodic models obviate the need for problematic processes such as speaker normalization by assuming that every word token is stored together with all its fine phonetic detail, including the characteristics of the speaker (e.g., high versus low voice, Northern versus Southern accent).

The assumption that human beings store all their experiences in full detail, as claimed by episodic models, is not new. It has been developed in the categorization literature, which also contains discussions of purely abstractionist (see e.g., Homa et al., 1979) and exemplar-based (see e.g., Nosofsky, 1986) models. Exemplar-based models have been highly popular ever since the article by Medin and Schaffer (1978), but have recently been seriously criticized by Minda and Smith (2002).

The popularity of episodic models within (psycho)linguistics does not only stem from the possibility to do without speaker normalization, but also from experimental evidence showing that listeners store token specific fine phonetic detail, including detail carrying indexical information (i.e., information about speaker identity and speech rate). For instance, Craik and Kirsner (1974) showed that words are recognized faster and more accurately when they are produced by the same voice. Likewise, Cole et al. (1974) found that participants are faster in determining whether two words in a sequence are identical, if these two words are recorded from the same speaker. Furthermore, Schacter and Church (1992) demonstrated that when presented with stems participants tend to form complex words which they have heard before, especially if these complex words were produced by the same voice as the stems. For production, Goldinger (1998) reported that participants tend to mimic previously heard pronunciations in their fine phonetic detail.

One of the few episodic models that has been described in (some) detail and that can capture this experimental evidence is MINERVA, developed by Hintzman (1986), and applied to speech by Goldinger (1998). In this model, word recognition involves the activation of all phonetically similar tokens in the lexicon, proportional to their similarity to the speech input. An aggregate of all activated exemplars constitutes an echo sent to the working memory, on the basis of which the speech input is recognized. The echo contains more idiosyncratic information of the exemplars in the lexicon if there are fewer of them present, while a higher number of exemplars results in a more general echo. Repetition of (the echo of) a low frequency word may therefore result in a token that is phonetically highly similar to one of the previously encountered tokens. Furthermore, the strength of an echo is proportional to the activation in the lexicon created by the input and a stronger echo facilitates the recognition process (and thus leads to shorter recognition times). Goldinger tested MINERVA by predicting the results of a shadowing experiment. In order to skip the first phase of the recognition process and to focus on the episodic part of the model, he converted the phonetic characteristics of the input signal and of the exemplars in the lexicon into simple vectors of numbers: Each token consisted of 100 name elements, 50 voice elements, and 50 context elements. The predictions made by MINERVA approximated the human data very closely. Thus, participants shadowed the fine phonetic detail of a stimulus more closely if they had heard only few tokens of that word and they were faster in shadowing high frequency (compared to low frequency) words.

Another influential episodic model is XMOD developed by Johnson (1997) for auditory word recognition. It differs from MINERVA especially in that it is an extension of the Lexical Access from Spectra (LAFS) model developed by Klatt (1979), which assumes that the incoming speech signal is transformed into a sequence of spectra (instead of vectors of abstract numbers). Johnson's XMOD assumes that during the recognition process, exemplars respond to the input in proportion to their similarity to this input. Their activation feeds activation of the abstract word nodes, which in turn enable recognition. Importantly, XMOD assumes that smaller units of linguistic structure, such as syllables and segments, emerge in the recognition process. Like word categories, these units are defined simply as sets of exemplars.

Interestingly, evidence is accumulating that when listeners make use of indexical information in previous mentions of a word, they do so only under slow processing conditions. McLennan and Luce (2005) showed this in a series of long-term repetition priming experiments, that is, lexical decision and shadowing experiments in which each target word occurred twice. Participants reacted faster on the second occurrence of a word, as expected. Importantly, the effect of identity priming was greatest if the second occurrence was similar to the first occurrence in speech rate or voice, and simultaneously also processing was slowed down, either by the nature of the nonwords in the experiment

(lexical decision) or by the forced time span between the stimulus and the response (shadowing). Mattys and Liss (2008) reported similar results for an experiment in which participants first listened to two series of words and had to indicate for the words in the second series whether they had heard them before. Participants were faster in identifying target words as "old" if the two occurrences were produced by the same speaker and this speaker suffered from dysarthria, which slowed down the average speed in the experiment.

4.3 Hybrid models

All models discussed so far have either abstract representations or exemplars at a given linguistic level. In addition, various models have been developed which assume both abstract generalizations and exemplars at the same linguistic level. We will refer to them as hybrid models. These hybrid models explicitly assume both a redundant lexicon and abstract generalizations. Several types of hybrid models have been formulated recently, but none of them have been fully implemented computationally.

One of the oldest hybrid models is the one proposed by Pierrehumbert (2002). She posits both abstract phonological representations and abstract phonological rules (e.g., prosodic final lengthening) as well as exemplar clouds associated with phonological units as exhibited in words (phonemes, phoneme sequences, and the words themselves). According to this model, speakers use all of this information during phonological encoding. Perception, in contrast, proceeds without the intervention of an abstract level, since fine phonetic detail in the speech signal, which would be abstracted away at an intermediate abstract phonological level, is known to affect the comprehension process.

McLennan et al. (2003) presented a hybrid model based on the Adaptive Resonance Theory (ART) of Grossberg and Stone (1986). This model assumes that an acoustic input activates chunks of lexical (words) and sublexical (allophones, features) representations. Some of the chunks are abstract (i.e., representations for words, allophones, phonological features) and others are captured by exemplars (e.g., speaker information). Chunks resonate with the input, and this resonance constitutes the listener's eventual percept. Importantly, more frequent chunks establish resonance with the input more easily and more quickly. Hence, by making the plausible assumption that more abstract representations are more frequent, McLennan and colleagues easily account for the finding that indexical information affects speech processing only when speech processing is slowed down.

McLennan and Luce (2005) already mention the possibility that the abstract representations and the exemplars are stored in different parts of the brain.

Goldinger (2007) discusses the Complementary Learning System (CLS) in which this is a central assumption. This model, which has been extensively developed into a computational model by O'Reilly and Rudy (2001) and Norman and O'Reilly (2003), assumes that an acoustic input first passes the cortical complex, where abstract processing takes place: The word is, among others, divided into its segments and acquires its meaning. It then passes, with all fine phonetic detail still present, via the entorhinal cortex to the hippocampal complex, where it is matched with acoustically similar traces and is stored itself as well. The hippocampal complex is a fast-learning network, which, again via the entorhinal cortex, affects the more stable cortical complex. This cortical complex is specialized to slowly learn statistical regularities. The CLS can account for why indexical properties play a role in speech perception especially when recognition is delayed. Listeners then react only after the acoustic input has arrived at the hippocampal complex, which processes indexical properties. Like MINERVA, the CLS does not yet have as its input realistic data: The model's input still consists of vectors with abstract numbers and letters.

The approach of Polysp (Polysystemic Speech Perception) developed by Hawkins and Smith (Hawkins and Smith, 2001; Hawkins, 2003) differs from the other models in two crucial respects. First, it stresses the assumption that a memory trace not only consists of acoustic information, but also contains its multimedial context, for instance, visual information about the speaker's articulatory gestures, information about the room the speaker was in, and information about the relationship between the speaker and the listener. Second, Polysp assumes that the analysis of an acoustic input into its linguistic units (phonemes, etc) is incidental. Circumstances dictate whether this analysis takes place at all, and if it takes place, whether the analysis precedes, coincides, or follows word recognition. Linguistic analysis may prevail especially in adults with extensive experience with identifying formal linguistic structure, in formal listening situations. This approach can thus account for the finding that, at least under some circumstances, indexical information affects word recognition only when processing is slow. Polysp has not been computationally implemented but Hakwins provides some suggestions, including incorporation in the ART model developed by Grossberg and colleagues (e.g., Grossberg and Stone, 1986). Note that this model is located on the continuum closer to the endpoint of exemplar-based models than any of the other models discussed above that assume both abstraction and exemplars.

4.3.1 Hybrid aspects of compressed lexicons

Current hybrid models build on the assumption that large numbers of individual exemplars are stored. Therefore, they run into the same problem that purely exemplar-based models have to face, namely, how to avoid an instance base with so many exemplars that it becomes impossible to query the instance

base in real time. In actual computational memory based models, some form of data compression is implemented. The role of data compression is worthy of further theoretical discussion.

Data compression has a long history in computer science. Efficient data structures for storing words were already discussed by Knuth in the early seventies (Knuth, 1973). Unsurprisingly, TIMBL, which is often applied to huge data sets, has implemented various compression algorithms. One of these, the "information gain tree" (IG-TREE), is especially interesting in the context of phonological generalizations with hybrid models.

An information gain tree is a kind of decision tree. Suppose we build such a decision tree in the context of predicting whether a final obstruent in Dutch is voiced or voiceless. Each successive decision in the tree considers a feature (e.g., the manner of articulation of the obstruent) and splits the data according to this feature, assigning to each of its daughter nodes the most likely outcome (voiced or voiceless) given the set of exemplars governed by that node. Note that in this tree data structure, similar exemplars share similar paths down the decision tree. In an IG-TREE, the successive decisions are ordered in such a way that as we move from its root down to its leaf nodes, the decisions become less and less important (and less successful) in separating the voiced from the voiceless realizations.

Now consider how such an IG-TREE performs under different time constraints. Under severe time constraints, only a few decision nodes can be considered. As a consequence, the choice for voiced or voiceless has to be based on the most likely outcome associated with decision nodes high up in the tree. As a consequence, this compressed memory will show rule-like behavior: the top nodes in the tree encode the highest-level generalizations. When time constraints are relaxed, more and more lower-level decisions will come into play, with at the lowest levels the individual exemplars. An exemplar memory compressed in this way has exactly the processing properties observed in the experiments of McLennan and Luce (2005) and Mattys and Liss (2008). This explanation is, however, completely different from that of the other hybrid models, which assume that abstract generalizations and exemplars are subserved by very different modules of the grammar (see also Ullman, 2004). Models with data compression show that computationally the abstract generalizations can be part and parcel of the organization of exemplars in memory. We note here that, as mentioned above, hybrid architectures in machine learning may speed on-line processing time, but may lead to somewhat degraded qualitative performance (Daelemans and Van den Bosch, 2005).

5 Concluding remarks

To conclude, advances in information technology, computer science, and psycholinguistics have created new possibilities for the study of phonology. Corpus-based research and computational modeling offer exciting new tools for understanding the knowledge that speakers and listeners have of the sound structure of their language.

References

Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90:119–161.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., et al. (1992). The IICRC Map Task corpus. *Language and Speech*, 34:351–366.

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

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 36:94–117.

Baayen, R. H., Levelt, W., Schreuder, R., and Ernestus, M. (2008). Paradigmatic structure in speech production. In Elliott, M., Kirby, J., Sawada, O., Staraki, E., and Yoon, S., editors, *Proceedings Chicago Linguistics Society 43, Volume 1: The Main Session*, pages 1–29, Chicago.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon*, 2:419–463.

Balling, L. and Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, 23:1159–1190.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, 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.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., and Hyönä, J. (1999). Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language. *Cognition*, 74:B13–B25.

Boersma, P. (1998). *Functional Phonology*. Holland Academic Graphics, The Hague.

Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32:45–86.

Booij, G. E. (1981). *Generatieve Fonologie van het Nederlands*. Het Spectrum, Utrecht.

Booij, G. E. (1995). *The phonology of Dutch*. Clarendon Press, Oxford.

Browman, C. and Goldstein, L. (1990). Tiers in articulatory phonology with some implications for casual speech. In Kingston, J. and Beckman, M. E., editors, *Between the grammar and physics of speech (Papers in Laboratory Phonology I)*, pages 341–376. Cambridge University Press, Cambridge.

Browman, C. and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49:155–80.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Benjamins, Amsterdam.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press, Cambridge.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper and Row, New York.

Cole, R., Coltheart, M., and Allard, F. (1974). Memory of a speaker's voice: Reaction time to same-or different-voiced letters. *The Quarterly Journal of Experimental Psychology*, 26:1–7.

Connine, C. M. and Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The linguistic Review*, 23:235–245.

Coussé, E., Gillis, S., Kloots, H., and Swerts, M. (2004). The Influence of the Labeller s Regional Background on Phonetic Transcriptions: Implications for the Evaluation of Spoken Language Resources. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, 4:1447–1450.

Craik, F. and Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, 26:274–284.

Cucchiarini, C. (1993). *Phonetic Transcription: A Methodological and Empirical Study*. Catholic University Nijmegen.

Daelemans, W. and Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge University Press, Cambridge.

Daelemans, W., Van den Bosch, A., and Zavrel, J. (1999). Forgetting exceptions is harmful in language learning. *Machine learning, Special issue on natural language learning*, 34:11–41.

Dainora, A. (2001). Eliminating Downstep in Prosodic Labeling of American English. *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*.

De Vaan, L., Schreuder, R., and Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2:1–23.

Dilley, L. and Pitt, M. (2007). A study of regressive place assimilation in

spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America*, 122:2340–2353.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface*. LOT, Utrecht.

Ernestus, M. and Baayen, R. H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, 42:873–903.

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

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120:1040–1051.

Fisher, W., Doddington, G., and Goudie-Marshall, K. (1986). The DARPA speech recognition research database: speciheation and status. *Proceedings of the DARPA Speech Recognition Workshop,(February. 1986)*, 12:100–110.

Fox-Tree, J. and Clark, H. (1997). Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition*, 62:151–167.

Frisch, S., Large, N., Zawaydeh, B., and Pisoni, D. (2001). Emergent phonotactic generalizations in English and Arabic. *Frequency and the emergence of linguistic structure*, pages 159–179.

Gagné, C. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27:236–254.

Gahl, S. (2008). "time" and "thyme" are not homophones: Lemma frequency and word durations in a corpus of spontaneous speech. *Language*, 84:474–496.

Gaskell, M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31:447–463.

Gimson, A. (1970). An Introduction to the Pronunciation of English. *London: Edward Arnold*.

Godfrey, J., Holliman, E., and McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 517–520.

Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 49–54, Saarbrücken.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105:251–279.

Gordon, E., Maclagan, M., and Hay, J. (2007). The ONZE Corpus. In Beal, J., Corrigan, K., and Moisl, H., editors, *Models and methods in handling of unconventional digital corpora*, volume 2, pages 82–104. Palgrave.

Gow, D. (2001). Assimilation and Anticipation in Continuous Spoken Word Recognition. *Journal of Memory and Language*, 45:133–159.

Grossberg, S. and Stone, G. (1986). Neural Dynamics of Word Recognition

and Recall: Attentional Priming, Learning, and Resonance. *Psychological Review*, 93:46–74.

Guy, G. R. (1980). Variation in the group and the individual: the case of final stop deletion. In Labov, W., editor, *Locating language in time and space*, pages 1–36. Academic Press, New York.

Harrell, F. (2001). *Regression modeling strategies*. Springer, Berlin.

Harris, J. (1994). *English sound structure*. Blackwell, Oxford.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31:373–405.

Hawkins, S. and 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. and Sudbury, A. (2005). How rhoticity became /r/-sandhi. *Language*, 81:799–823.

Hazen, T. J., Saenko, K., La, C.-H., and Glass, J. R. (2004). A segment-based audio-visual speech recognizer: Data collection, development, and initial experiments. In *Proceedings of the International Conference on Multimodal Interfaces*, pages 235–242, Pennsylvania.

Hintzman, D. (1986). Schema abstraction: a multiple-trace memory model. *Psychological Review*, 93:411–428.

Homa, D., Rhoads, D., and Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, 5:11–23.

Jackendoff, R. S. (1975). Morphological and semantic regularities in the lexicon. *Language*, 51:639–671.

Jaeger, F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59:434–446.

Johnson, K. (1997). Speech perception without speaker normalization. In Johnson, K. and Mullennix, J., editors, *Talker variability in speech processing*, pages 145–166. Academic Press, San Diego.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.

Keating, P. A. (1998). Word-level phonetic variation in large speech corpora. In Alexiadou, A., Fuhrop, N., Kleinhenz, U., and Law, P., editors, *ZAS Papers in Linguistics 11*, pages 35–50. Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung, Berlin.

Keuleers, E. (2008). *Memory-based learning of inflectional morphology*. University of Antwerp, Antwerp.

Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. (2005). Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok'. *Corpus Linguistics and Linguistic Theory*, 1:183–223.

King, S. and Taylor, P. (2000). Detection of phonological features in con-

tinuous speech using neural networks. *Computer Speech and Language*, 14:333–353.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In Van der Hulst, H. and Smith, N., editors, *The structure of phonological representations*, pages 131–176. Foris, Dordrecht.

Klatt, D. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7:1–26.

Knuth, D. E. (1973). *The Art of Computer Programming. Vol. 3: Sorting and Searching*. Addison-Wesley, Reading, Mass.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: phonological effects and phonetic explanations. In Hardcastle, W. J. and Marchal, A., editors, *Speech production and speech modelling*, pages 21–33. Kluwer, Dordrecht.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, 39:51–93.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2008). Reading of multimorphemic Dutch compounds: towards a multiple route model of lexical processing. *Journal of Experimental Psychology: Learning, Memomy and Cognition*, in press.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. The MIT Press, Cambridge, Mass.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38.

Local, J. (2007). Phonetic detail and the organization of talk-in-interaction. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 1–10. Universität des Saarlandes, Saarbruecken.

Lombardi, L. (1999). Positional Faithfulness and Voicing Assimilation in Optimality Theory. *Natural Language & Linguistic Theory*, 17:267–302.

Mattys, S. L. and 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.

McCarthy, J. and Prince, A. (1993). Generalized alignment. In Booij, G. E. and Van Marle, J., editors, *Yearbook of Morphology 1993*, pages 79–153. Kluwer, Dordrecht.

McLennan, C., Luce, P., and Charles-Luce, J. (2003). Representation of Lexical Form. *Learning, Memory*, 29:539–553.

McLennan, C. T. and 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.

Medin, D. and Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, 85:207–238.

Minda, J. and Smith, J. (2002). Comparing Prototype-Based and Exemplar-Based Accounts of Category Learning and Attentional Allocation. *Learning, Memory*, 28:275–292.

Mitterer, H. and Blomert, L. (2003). Coping with phonological assimilation

in speech perception: Evidence for early compensation. *Perception & Psychopysics*, 65:956–969.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Foris Publications, Dordrecht.

Neu, H. (1980). Ranking of constraints on /t,d/ deletion in American English: a statistical analysis. In Labov, W., editor, *Locating language in time and space*, pages 37–54. Academic Press, New York.

Newmeyer, F. (2003). Grammar is grammar and usage is usage. *Language*, pages 682–707.

Norman, K. and O'Reilly, R. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review*, 110:611–46.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter*, 5:4–8.

O'Reilly, R. and Rudy, J. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, 108:311–345.

Ostendorf, M., Price, P., and Shattuck-Hufnagel, S. (1995). The Boston University Radio News Corpus. *Boston University Technical Report, ECS-95-001*, University of Boston.

Pickering, M. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

Pierrehumbert, J. (1987). The Phonetics and Phonology of English Intonation. *Unpublished Ph. D. dissertation, MIT.(Reproduced by Indiana University Linguistics Club.)*.

Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Phonology and Phonetics: Papers in Laboratory Phonology VII*, pages 101–140. Mouton de Gruyter, Berlin.

Pinker, S. (1991). Rules of language. *Science*, 153:530–535.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, 45:89–95.

Plag, I., Kunter, G., and Lappe, S. (2007). Testing hypotheses about compound stress assignment in english: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory*, 3:199–232.

Plag, I., Kunter, G., Lappe, S., and Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in english. *Language*, 84:760–794.

Plug, L. (2005). From words to actions: The phonetics of 'eigenlijk' in two communicative contexts. *Phonetica*, 62:131–145.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–

159.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005b). Frequency and acoustic length: the case of derivational affixes in Dutch. *Journal of the Acoustical Society of America*, 118:2561–2569.

Racine, I. and Grosjean, F. (2002). La production du e caduc facultatif est-elle previsible? un debut de reponse. *Journal of French and Language Studies*, 12:307–326.

Ranbom, L. and Connine, C. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57:273–298.

Rietveld, T., hout, R., and Ernestus, M. (2004). Pitfalls in Corpus Research. *Computers and the Humanities*, 38:343–362.

Russell, K. (2008). Sandhi in Plains Cree. *Journal of Phonetics*.

Saraçlar, M. and Khudanpur, S. (2004). Pronunciation change in conversational speech and its implications for automatic speech recognition. *Computer Speech & Language*, 18:375–395.

Schacter, D. and Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18:915–930.

Scheibman, J. and Bybee, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in english. *Linguistics*, 37:575–596.

Schuppler, B., Ernestus, M., Scharenborg, O., and Boves, L. (submitted). An automatic method to analyze acoustic reduction in a corpus of conversational dutch.

Semon, R. (1923). *Mnemische Empfindungen*. (B. Duffy, Trans.). Allen and Unwin, London. (Original work published 1909).

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

Skousen, R. (1989). *Analogical Modeling of Language*. Kluwer, Dordrecht.

Skousen, R. (2002). Analogical modeling and quantum computing. In Skousen, R., Lonsdale, D., and Parkinson, D., editors, *Analogical modeling: An exemplar-based approach to language*, pages 319–346. John Benjamins, Amsterdam.

Stemberger, J. P. and MacWhinney, B. (1988). Are lexical forms stored in the lexicon? In Hammond, M. and Noonan, M., editors, *Theoretical Morphology: Approaches in Modern Linguistics*, pages 101–116. Academic Press, London.

Ten Bosch, L., Baayen, R., and Ernestus, M. (2006). On speech variation and word type differentiation by articulatory feature representations. In *Proceedings of the Ninth International Conference on Spoken Language Processing*, pages 2230–2233, Pittsburgh, Pennsylvania.

Tiersma, P. M. (1982). Local and General Markedness. *Language*, 58:832–849.

Ullman, M. (2004). Contributionss of memory circuits to language: the declarative/procedural model. *Cognition*, 92:231–270.

Van Bael, C., Boves, L., van den Heuvel, H., and Strik, H. (2006). Automatic phonetic transcription of large speech corpora: A comparative study.

In *Proceedings of the Ninth International Conference on Spoken Language Processing*, pages 1085–1088, Pittsburgh, Pennsylvania.

Vennemann, T. (1972). Rule inversion. *Lingua*, 29:209–242.

Vieregge, W. (1987). Basic aspects of phonetic segmental transcription. *Probleme der phonetischen Transkription*, pages 5–48.

Wester, M., Kessens, J., Cucchiarini, C., and Strik, H. (2001). Obtaining phonetic transcriptions: A comparison between expert listeners and a continuous speech recognizer. *Language and Speech*, 44:377–403.

Wetzels, W. and Mascar, J. (2001). The typology of voicing and devoicing. *Language*, 77:207–244.

Zonneveld, W. (2007). Issues in Dutch devoicing: Positional faithfulness, positional markedness,and local conjunction. In Van der Torre, E.-J. and Van de Weijer, J., editors, *Voicing in Dutch*, pages 1–40. Benjamins, Amsterdam.