Chapter 19. Corpora, databases, and internet resources

Corpus phonology with speech resources

Jennifer Cole and Mark Hasegawa-Johnson

University of Illinois at Urbana-Champaign

1 Introduction

This section introduces the methods of corpus phonology using speech databases, for the investigation of phonological variation, for understanding the phonetic underpinnings of phonological phenomena, and for research on the category structures of spoken language. We address the practical challenges in identifying an existing corpus appropriate for phonological research, the corpus resources mentioned in **Loehr and van Guilder, and Post and Nolan, this volume**. The challenges of annotation that arise with the creation of a new corpus are also highlighted, with further discussion of annotation in relation to community standards taken up in **Loehr andvanGuilder_chap19**. Finally, corpus research involves processing large speech databases, and this section ends with a discussion of the computational and statistical tools that are widely used in corpus studies, with related applications in the development of speech technologies.

Corpus analysis for phonological research involves investigation of the phonetic, phonological and lexical properties of speech for the purpose of understanding the patterns of variation in the phonetic expression of words, and the distributional patterns of sound elements in relation to the linguistic context. In some respects corpus methods complement laboratory-based experimental methods in phonology, and for some fields of inquiry corpus materials are essential. The central role of speech databases and lexical corpora for the study of frequency and similarity in phonology corpus data is clearly demonstrated in **Frisch_chap19**.

2 Phonetic considerations in phonological research

Phonology is concerned with characterizing the sound patterns of language, typically presented in terms of a *system* of contrastive sound elements (e.g., syllables, segments, features) and the *distribution* of those sounds in the makeup of phonological words and phrases. This focus on the sound system and the characteristic sound patterns of words is what distinguishes the study of phonology from the study of phonetics, as these two fields are traditionally construed. Yet the phonologist's perspective on sound systems is typically rooted in knowledge about the phonetic properties of the sound elements that make up a language, and reflects direct observation of the phonetic form of spoken words and phrases.

Considering the phonetic substance of phonological forms presents a challenge and an opportunity. The challenge arises from the inherent variability in the phonetic realization of a word, which can make it difficult to identify a unique description of its core phonetic properties. For instance, the English words *rapid* and *rabid* in careful pronunciation are phonetically distinct in their medial consonants, [p] and [b], but in casual pronunciation this distinction can be reduced, with an absence of voicing during closure for [b] and a shortening of the voice onset time for [p], rendering the two medial consonants phonetically very similar. This reduction of a phonological contrast poses a question about the nature of phonological encoding in long-term memory (i.e., the lexical form, see **chapter 8 on lexical representation** for discussion). If the phonetic detail related to reduced forms is not encoded in lexical representations, then the question shifts to address processing (see also **chapter 14** and **Ernestus, this volume** on this topic): what is the process by which a speaker/hearer establishes a mapping between the phonetic forms that are experienced and their encoding in the mental lexicon?

The variability of phonetic form is also a source of insight for phonology. Very often we can observe patterns of fine-grained phonetic variation that mirror phonological alternations or distributional restrictions. For example, the graded coarticulation of a vowel under the influence of a vowel in the upcoming syllable in English mirrors the phonological pattern of assimilation found, e.g., in local processes of umlaut or vowel harmony in other languages (Beddor et al. 2002; Cole et al. 2010).

Observing patterns of 'low-level' (ie., sub-phonemic), gradient phonetic variation sheds light on how the phonetic context of a sound element can shape phonological patterns that restrict the occurrence of that element, and there is growing interest in uncovering the bases of phonological sound patterns in properties of phonetics and speech processing (e.g., Archangeli and Pulleyblank 1994; Blevins 2004; Hayes, Kirchner and Steriade 2004).

3 Motivating corpus analysis for phonology

3.1 Variation and phonetic form

In order to explore the variable phonetic substance of phonological elements, and the influence of phonetics in shaping sound patterns, the phonologist must go beyond the analysis of citation forms and examine words in connected speech in corpora that represent variation due to different speech styles (Hirschberg 2000; Yuan et al. 2005; see also **Ernestus_chap5**) and speaker variation (Foulkes in press; see also **chap4 on speaker related variation**). All this variety is part of the everyday experience of language for the speaker/hearer, and comprises the phonetic basis over which phonological patterns are learned.

When speech is used for communicative goals, as in the everyday use of language, it is produced with prosodic patterns that convey the information structure and pragmatic context of an utterance, and prosodic context is also known to affect the phonetic realization of words (e.g., Wightman et al. 1992; van Bergem 1993; Kochanski et al.

2005; Calhoun 2006; Turk and Shattuck-Hufnagel 2007; Cole et al. 2007, Yoon 2007; see also **Turk_chap14 and **Frota_chap9**). In addition, though prosodic features are present in all forms of speech regardless of style, the expressive content of spontaneous speech gives rise to a particularly rich variety of prosodic patterns, different from read speech (e.g., Nakatani et al. 1995; Schaefer et al. 2005).

3.2 Phonology in relation to linguistic structure and usage

Spontaneous speech produced in communicative contexts offers the best opportunity to observe a wide range of phonetic variability; yet although researchers have devised methods for eliciting spontaneous speech in a laboratory setting (see **Warner; and Post and Nolan; this volume**), by engaging subjects in controlled communicative tasks (e.g., Anderson et al. 1991; Hirschberg and Nakatani 1996; Schaefer et al. 2005; Brown-Schmidt and Tanenhaus 2008; Khan 2008), the resulting data is (by design) less varied than speech produced in casual conversation. For direct observation of spontaneous, conversational speech researchers turn to speech databases for corpus analysis.

Phonology in relation to linguistic structure and usage

A speech corpus not only provides a basis for investigating variability in phonetic form, but it also provides a rich resource for studying the relationship between phonological form and other levels of linguistic structure. For instance, it has long been known that the sound patterns of a language may be sensitive to syntactic context (Kisseberth and Abasheikh 1974; Chen 1987) and may reflect discourse organization (Grosz and Hirschberg 1992). Clearly, evidence for any interaction between phonology and "higher" levels of linguistic structure must come from observation of whole phrases, multi-phrase utterances, and entire discourses. Similarly, a variety of syntactic, pragmatic and discourse contexts are required to understand the phonology of intonation and prosody, and corpus materials have been widely used in such work (**Post and Nolan_chap21**).

Usage frequency is another factor known to influence the phonetic form of words. Greenberg (2000), Bybee (2001), and Bell et al. (2003), among others, have shown that words that occur frequently in speech have a higher incidence of consonant lenition and vowel reduction compared to low-frequency words. Usage statistics are calculated based on large corpora, which also provide plenty of data that illustrate the effects of usage on phonetic form. Bybee (2001) has also shown that patterns of phonetic reduction that arise in high-frequency words can be phonologized, resulting in stable synchronic sound patterns. By examining phonetic variation in relation to usage frequency, it is possible to identify patterns that may be precursors to future sound change.

4 Choosing a corpus

There are several considerations in choosing a speech corpus for phonological research. The first concerns the goal of the research and the availability of an existing corpus. Aresearcher interested in the effect of the given/new distinction on phonetic form may want to see how repeated mention affects the phonetic properties of words. This requires a corpus where speakers talk on the same topic for an extended period, incorporating multiple utterances, or multiple conversation turns in the case of dialogue, because repeated mention of a word is more likely in an extended discourse. A suitable corpus might be one consisting of extended interviews such as the Buckeye Corpus (Pitt et al. 2007); dialogues that are focused on a topic that sustains interest over time, as in the Switchboard corpus (Godfrey and Holliman 1997) or CallHome corpus (Canavan et al. 1997); or dialogues over lengthy tasks that require repeated mention of objects, places or other things that are present in the task domain, as in the HCRC Map Task Corpus (HCRC Map Task Corpus 1993; *see also* Anderson et al. 1991). On the other hand, if the research goal is to investigate how speakers accommodate to the phonological and phonetic patterns of another person's speech, it would be essential to choose a corpus in which speakers with different speech patterns are engaged in interactive dialogue, such as the Fisher corpus (Cieri et al., 2004, 2005), which consists of telephone recordings from over 11,000 conversations between English speakers, representing a wide range of age groups and regional dialects, including non-US and foreign-accented varieties of English.

The corpora cited above are examples of speech databases for English (and in the case of CallHome, for other languages as well) that are in the public domain; they are disseminated to the public by a distributor, often times with a licensing fee. The alternative to using an existing corpus is for the researcher to build a corpus from scratch, by recording speech samples directly from speakers recruited for that purpose.

The advantages to using an existing, published corpus are savings in time and money, and with some corpora, access to a much larger database than a single individual researcher could construct. A further advantage to working with a corpus in the public domain is the possibility of building on the work others have done using the same corpus, or using prior results as a benchmark for testing new research methods.

Disadvantages of using existing corpora usually arise when the goals of the research are not adequately served by the speech materials available in existing corpora. For example, at the time of this writing, there is no publicly available database of dysarthric speech that surpasses the one compiled by H. Kim and her colleagues containing just about one hour of speech for each of eighteen talkers (Kim et al. 2008). Likewise, to investigate the phonological structures of a non-standard dialect, the researcher may need access to speech that is produced in a social setting and register that is conducive to the use of that dialect. Speech samples that are recorded in a formal laboratory setting , or through interaction with an unfamiliar investigator who is not part of the target speech community, may fail to fully exhibit the characteristics of the dialect. A related limitation is the simple fact that there are no existing corpora for most languages, and similarly few databases for non-standard or non-prestige varieties of any language..[1] Using a portable digital voice recorder, spontaneous conversational speech data in any speaking style or language may be recorded with no substantial technical effort; most of the effort in acquiring a corpus is spent contacting subjects, acquiring their legal consent, creating a task description that will keep subjects talking long enough to collect the desired speech sample in the desired speaking style, and finally, transcribing the data.

Although spontaneous speech databases are especially relevant to the study of phonetic variation, there are existing corpora for read speech that are appropriate for some research needs. Thus, the Boston University Radio Speech corpus (Ostendorf et al. 1996) is useful for research on prosody because it comes with a detailed, manually produced prosodic transcription, and a phone-level transcription, both of which are aligned with the audio signal. This corpus has been used for research on the acoustic correlates of prosodic features in American English, as they are represented in this style of professionally read speech (e.g., Choi et al. 2005; Cole 2007; Dainora 2001; Kim & Cole 2005; Yoon 2007).

5 Corpus transcription

5.1 Metadata and orthographic transcription

In order for a speech database to be useful for phonological analysis, it is necessary to have some additional information about the content of the speech. Linguistic metadata will provide information about the speakers, such as sex, age, ethnicity, and region of residence. Metadata may also provide information about speaker recruitment and recording procedures .

The most ubiquitous and, often, most useful type of annotation available for any speech corpus is its orthographic word transcription. Using an orthographic transcription together with a pronunciation dictionary, it is possible for the researcher to use simple text search tools in order to find places in the database where specific phonological structures of interest may have occurred, and to focus manual post-hoc analysis exclusively on the selected segments. At its simplest, the transcription is a separate document that specifies the words of each utterance in the database in running text. Much more useful are transcriptions that are time-stamped, so the beginning and end of each word (or sentence, or talker-turn) is indicated, allowing the researcher to locate that word /sentence/turn in the corresponding audio file. A useful method for producing partially time-stamped orthographic transcriptions is to segment the speech data at every silence longer than some threshold (e.g., 500 ms), and then to give the pre-segmented waveforms to transcribers for annotation.

Some corpora do not come with transcriptions, and the researcher must create one, as of course must be done for any corpus that is created by the researcher; working efficiently, it is possible for most annotators to transcribe utterance units in about four times real time, i.e., four minutes of transcriber time for every minute of speech. Although word transcription may seem like a very simple task, in the case of conversational speech complications arise due to disfluencies, hesitations, and speech repairs, or from poor signal quality. For these reasons, transcriptions almost always include questionable entries, where reasonable people disagree about what they hear in the recording. There are also a surprising number of orthographic ambiguities in the transcription of spoken English, e.g., numerical expressions, word fragments, idioms, discourse markers, and proper names each typically have two or more common transliterations. To minimize the impact of errors and uncertainties on the reliability of the transcription, transcription projects will typically rely on a written protocol for the treatment of disfluencies, errors and ambiguous entries, which is used to train the transcribers (e.g., Linguistic Data Consortium 2009).

5.2 Transcription of sub-word units: phones and features

When the research plan is to investigate phonetic variation at a level smaller than the word, such as the phone or syllable level, an additional layer of transcription is needed to identify such units within each word. Phone-level transcription is the most common sub-word level that is labeled in existing corpora, but transcriptions of this sort for large databases (anything over about 1,000 words) are rare. Because it is a very time-intensive task that requires phonetic training , phone-level transcription is rarely done by hand. Rather, an initial pass at transcription is made with the use of automated methods. Working from an orthographic word-level transcription, the phones for each word can be retrieved from a digital pronunciation dictionary and automatically inserted into the transcription, as a further specification for each word. This step is followed by a procedure of *forced alignment,* by which each phone in the dictionary form of a word is mapped onto some portion of the acoustic signal for that word.

Forced alignment is done using algorithms from Automatic Speech Recognition (ASR), and is most successful when each phone associated with the word in its dictionary form is actually fully pronounced. But this is not always the case, and indeed, full pronunciation is not even typically the case for words in spontaneous speech (Greenberg and Fosler-Lussier 2000). Forced alignment can be improved by systems that explicitly model the most common patterns of pronunciation variation, but much more research is needed in this area to improve the reliability of the time-aligned phone labeling using this method. Some of the corpora mentioned above use forced alignment followed by a process of manual correction which can correct many if not all of the resulting errors (e.g., the Buckeye corpus). Manual correction is still a slow and costly procedure, but this dual approach using automatic labeling with manual correction is often an excellent compromise to the much more costly alternative of a full manual transcription.

The need for a digital dictionary for the use of forced alignment means that automatic phone labeling can be applied only to those languages for which such resources exist. Fortunately, there are efforts underway to produce such resources for an increasing pool of languages (eg., Hussain et al. 2005).

Looking below the level of the phone, transcription can also specify smaller units such as phonological distinctive features or articulatory gestures. For example, phones specified for a given word in the pronunciation dictionary can be mapped onto distinctive features, and then automatic methods can be used to locate the distinctive features in the speech stream using acoustic landmarks. This approach has been demonstrated for many of the distinctive features used to encode lexical contrast (Stevens 2002; Livescu et al. 2007)

5.3 Prosody transcription

Corpus-based analyses have proved beneficial for the study of speech prosody, but introduce the need for an additional level of prosodic transcription . Using transcription methods such as the Tones and Break Indices (ToBI) system (Beckman et al. 2005), the locations of phrasal prominence and phonological phrase boundaries are identified, along with a tonal specification marking the associated pitch movement (see **PostandNolan_chap21** for other approaches to prosody transcription). Prosody transcription is a complex task that incorporates the transcriber's auditory impression of prominence and phrasal juncture with visual inspection of the graphical speech display (including at least the pitch track, waveform and spectrogram), and requires specialized training. In the case of ToBI transcription it is also a slow task, taking anywhere from 10 to 100 times the duration of the speech recording, and requires first having a reliable time-aligned word transcription. And, like other forms of transcription, prosody transcription is error prone and different transcribers can perceive the prosodic features of an utterance differently. Reliability studies of several ToBI transcription projects show that agreement rates between transcribers are impressively high—Pitrelli et al. (1994) report agreement rates of up to 81% for tone label, and 92% for the break index coding the level of phrasal juncture—but the potential for errors and uncertainty remains.

Many researchers have looked at ways to automate prosody transcription, primarily by identifying a set of acoustic correlates of prosody and using these features to train a classifier that takes as its input the word sequence, the acoustic speech signal, and sometimes additional information about part-of-speech or shallow syntactic features and returns a prosody annotation for each word or sub-word unit (eg., Wightman et al. 1994; Syrdal et al. 2001; Chen et al. 2004; Ananthakrishnan and Narayanan 2008). These efforts have contributed greatly to the understanding of how prosody is encoded in the acoustic signal, but so far have not been successfully tested on spontaneous speech data.

5.4 Assessing transcription reliability

No corpus should be publicly released without at least two levels of quality validation. First, automatic verification using standard methodsshould be applied to any corpus prior to release. The energy of each waveform should be computed, in order to verify that every file in the distributed corpus contains speech. Transcription files should be spell-checked. The Linguistic Data Consortium (2004) recommends running a "syntax check" that searches transcription files for timestamps without text, illegal characters, ill-formed symbols (e.g., ill-formed foreign speech transcriptions or non-speech transcriptions), bad spacing around punctuation, and numerical utterances that are entered using digits rather than full orthographic words.

Second, any coding system that requires rater training (including phoneme, distinctive feature, and prosodic transcriptions) should be evaluated by measuring inter-transcriber agreement. It is usually impractical to duplicate transcriber effort for the entire corpus, but the general validity of the transcription system can be measured by assigning more than one transcribers to re-code a small portion of the corpus. Cohen's kappa (Cohen 1960) or Fleiss' kappa (Fleiss 1971) are statistical methods that can be used to test the reliability of transcriptions across two or more transcribers.

6 Pronunciation dictionaries and lexica

It is possible to perform phonology research using a database of recorded speech, an orthographic transcription, and a pronunciation dictionary. It usually takes less time to write a dictionary than it would take to phonemically transcribe the entire corpus, but writing a dictionary is, itself, a time consuming task. For this reason, until recently, the pronunciation dictionaries distributed with most speech technology applications (synthesizers and recognizers) were considered to be valuable pieces of intellectual property, protected by the full weight of international copyright law. Recently, encouraged by a few widely cited examples (Weide 1995), increasing numbers of dictionaries are being released to the public. These efforts are supported by the publication of open source licenses appropriate to the distribution of text data, e.g., the Creative Commons Share Alike license (Creative Commons, 2009). The Creative Commons licenses allow users to add content to a published work, provided that, if the work is republished, it be republished under the same license with appropriate attribution; for example, Hasegawa-Johnson and Fleck have republished the dictionary for the Carnegie Melon University Pronouncing Dictionary (or *cmudict,* a machine-readable pronunciation dictionary for North American English) with added tags for syllabification, part of speech, and named entities, and with about 100,000 additional entries derived from other open sources (Hasegawa-Johnson and Fleck 2007).

Languages whose letter-to-sound mappings are more predictable than English may be well served by an orthographic dictionary. For example, Hussain et al. have published an Urdu pronouncing dictionary using pronunciation codes based on the traditional Urdu orthography plus vowels (Ijaz and Hussain 2007).

7 Statistical and computational methods for data analysis

After the researcher has obtained a speech corpus, created and assessed a transcription (if needed), and identified regions of interest within the corpus, data collection can begin. A wide variety of data may be extracted for the purpose of phonological investigation, depending on the researcher's specific interests. For instance, data may consist of acoustic measurements taken from the speech signal, articulatory measurements if they are available (e.g., Westbury 1994), measurements of lexical frequency or phonotactic probability, or properties of the phonological, syntactic or discourse context in which a targeted phonological unit occurs. An important detail in coding the data is the assignment of a unique label to each data point that identifies the speech unit (e.g., word, phrase or utterance) from which the measurement is extracted, and for ease of reference, that also identifies the speaker, file number, and any properties of the data or its context that will be considered in the analysis.

A benefit of corpus research in phonology is that it provides a ready training database for an analysis of the category structure of speech—a central concern of phonology. Statistical methods for classification analysis may be used to test how well the observed data can be classified into linguistically meaningful categories (e.g., voiced vs. voiceless stops, urban vs. rural dialect, phrase-final vs. phrase-medial position) based on one or more characteristics inherent in the items. There are many approaches to classification analysis, using linear or non-linear methods e.g., regression, discriminant analysis, support vector machines, k-nearest neighbor, decision trees, neural networks, Bayesian models, Hidden Markov Models (e.g., Webb 1999; see also **chap24** on statistical methods in phonology). Some of these methods are also used in machine learning to create computer algorithms that can automatically learn the distribution of the data items into linguistic categories (Mitchell 1997). These methods of classification analysis align with methods used for the creation of speech technologies, such as speech synthesis and automatic speech recognition, and many of the studies that employ these methods in the analysis of speech corpora simultaneously contribute to linguistic understanding and technology development (e.g., Chen et al. 2006; Liu et al. 2006; Hirschberg et al.2007).

8 Summary

Speech corpora offer a valuable source of data for phonological investigation, and are arguably an essential resource for the study of sound patterns that arise in connected, casual speech, Relative to corpus-based research in other areas of linguistic inquiry, corpus phonology research is in its infancy, and there remains much to be learned from existing resources. But it is also true that the linguistic coverage of the existing corpora is limited to a fraction of the world's languages, and does not fully represent all the dialectal varieties and speech styles that are of phonological interest. Fortunately, the technology needed to construct a corpus, including recording equipment and digital storage, are fairly inexpensive and easily obtained. On the other hand, a corpus is only as good as its annotation, and the human resources needed to produce a reliable, quality transcription are considerable.

One of the distinguishing features of corpus-based research is the large volume of data that is available for analysis from even a medium-sized corpus, e.g, the Buckeye corpus (Pitt et al. 2007, comprising approximately 20 hours). On the other hand, even with a corpus of this size there may be a scarcity of examples of low-frequency phonological phenomena, reflecting the trade-off between the use of naturalistic speech materials drawn from a corpus and speech materials controlled by the experimenter and elicited in the laboratory.

8.1 References

Ananthakrishnan, S.; Narayanan, Shrikanth S. (2008). Automatic Prosody Labeling using Acoustic, Lexical, and Syntactic Evidence. *IEEE Transactions on Speech, Audio and Language Processing*, 16(1): 216-228.

Anderson, Anne H.; Bader, Miles; Bard, Ellen Gurman; Boyle, Elizabeth; Doherty, Gwyneth; Garrod, Simon; Isard, Stephen; Kowtko, Jacqueline; McAllister, Jan; Miller, Jim; Sotillo, Catherine; Thompson, Henry S.; Weiniert, Regina (1991). The HCRC Map Task Corpus. *Language and Speech*, 34: 351-366.

Archangeli, Diana; Pulleyblank, Douglas (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

Beckman, Mary E.; Hirschberg, Julia; Shattuck-Hufnagel, Stefanie (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (ed.) *Prosodic Typology -- The Phonology of Intonation and Phrasing*, Oxford University Press, Chapter 2: 9-54.

Beddor, Patrice Speeter; Harnsberger, James D.; Lindemann, Stephanie (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics,* 30: 591-627.

Bell, Alan; Jurafsky, Daniel; Fosler-Lussier, Eric; Girand, Cynthia; Gregory, Michelle; and Gildea, Daniel. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America,*113(2): 1001–1024.

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Brown-Schmidt, Sarah; Tanenhaus, Michael K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32: 643-684.

Bybee, Joan. (2001). *Phonology and Language Use.* (Cambridge Studies in Linguistics, 94). Cambridge UK: Cambridge University Press.

Calhoun, Sasha (2006). Information structure and the prosodic structure of English: A Probabilistic relationship. Ph.D. dissertation, University of Edinburgh, UK.

Canavan, Alexandra; Graff, David; and Zipperlen, George (1997). *CALLHOME American English Speech,* Linguistic Data Consortium, Philadelphia.

Chen, Ken; Hasegawa-Johnson, Mark; Cohen, Aaron (2004). "An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, pp. 509–512.

Chen, Ken; Hasegawa-Johnson, Mark; Cohen, Aaron; Borys, Sarah; Kim, Sung-Suk; Cole, Jennifer;
Choi, Jeung-Yoon; Hasegawa-Johnson, Mark; Cole, Jennifer. (2005). Finding Intonational Boundaries Using Acoustic Cues Related to the Voice Source. *Journal of the Acoustical Society of America* 118(4):2579-88.

Choi, Jeung-Yoon (2006). Prosody dependent speech recognition on Radio News corpus of American English. *IEEE Transactions in Speech and Audio Processing,* 14(1): 232-245.

Chen, Matthew (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109-149.

Cieri, Christopher; Miller, David; Walker, Kevin (2004). The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC),* Lisbon.

Cieri , Christopher; Graff, David; Kimball, Owen; Miller, Dave; and Walker, Kevin (2005). *Fisher English Training Speech, Part 2 Transcripts,* Linguistic Data Consortium, Philadelphia.

Cohen, Jacob (1960). *A coefficient of agreement for nominal scales*, Educational and Psychological Measurement Vol.20, No.1, pp.37–46

Cole, Jennifer; Kim, Heejin; Choi, Hansook; Hasegawa-Johnson, Mark (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35: 180-209*.*

Cole, Jennifer; McMurray, Bob; Linebaugh, Gary; and Munson, Cheyenne (in press). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics.*

Creative Commons (2009). *Attribution-Share Alike 3.0*, downloaded April 20, 2009 from http://creativecommons.org/licenses/by-sa/3.0/

Dainora, Audra. 2001. *An empirically based probabilistic model of intonation in English.*

Ph.D. thesis, University of Chicago.

Fleiss, Joseph L. (1971) *Measuring nominal scale agreement among many raters, Psychological Bulletin*, 76(5) 378-382.

Foulkes, Paul (in press). Exploring social-indexical knowledge: A long past but a short history. In *Laboratory Phonology,* vol. 1.

Godfrey, John J.; Holliman, Edward. (1997). *Switchboard-1 Release 2,* Linguistic Data Consortium, Philadelphia.

Greenberg, Steven (1999). Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation*. Speech Communication* 29: 159-176.

Greenberg, Steven; Fosler-Lussier, Eric (2000). The uninvited guest: Information's role in guiding the production of spontaneous speech. *In Proceedings of the Crest Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, 129–132.

Grosz, Barbara; Hirschberg, Julia (1992). Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing.* Banff, October, pp. 429–432.

Hasegawa-Johnson, Mark; Fleck, Margaret. (2007). *The ISLEX Project*, downloaded April 20, 2009 from http://www.isle.uiuc.edu/dict.

Hayes, Bruce; Kirchner, Robert; Steriade, Donca (2004). *Phonetically-based Phonology.* Cambridge, UK: Cambridge University Press.

*HCRC Map Task Corpus. (*1993). *Linguistic Data Consortium*, Philadelphia

Hirschberg, Julia (2000). A corpus-based approach to the study of speaking style. In Horne, Merle (ed.) *Prosody: Theory and Experiment*, pp, 335-350, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hirschberg, Julia; Nakatani, Christine H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the 34th annual meeting on Association for Computational Linguistics,* Santa Cruz, California, pp. 286 – 293.

Hirschberg, Julia; Gravano, Agustin; Nenkova, Ani; Sneed, Elisa; Ward, Gregory (2007). Intonational Overload: Uses of the H* !H* L- L% Contour in Read and Spontaneous

Speech. In Cole, Jennifer and Hualde, José I. (eds.) *Laboratory Phonology 9*, pp. 455-482. Berlin: Mouton de Gruyter.

Holmes, Janet; Vine, Bernadette; Johnson, Gary (1998). *Guide to The Wellington Corpus of Spoken New Zealand English*, School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Hussain, Sarmad; Durrani, Nadir; Gul, Sana (2005). *Pan localization: Survey of language computing in Asia*. Center for Research in Urdu Language Processing, Lahore, Pakistan. [retrieved 4/14/09 from http://www.idrc.ca/uploads/user-S/11446781751Survey.pdf]

Ijaz, Madiha; Hussain, Sarmad (2007). *Corpus-based lexicon development*, in Conference on Language and Technology, University of Peshawar, Pakistan.

Khan, Sameer ud Dowla (2008). *Intonational phonology and focus prosody in Bengali.* Ph.D. dissertation, UCLA.

Kim, Heejin; Cole, Jennifer. 2005. The stress foot as a unit of planned timing: Evidence from shortening in the prosodic phrase. *Proceedings of Interspeech 2005*, Lisbon, Portugal, pp. 2365-2368.

Kim, Heejin; Hasegawa-Johnson, Mark; Perlman, Adriene; Gunderson, Jon; Huang, Thomas; Watkin, Kenneth; Frame, Simone (2008), *Dysarthric Speech Database for Universal Access Research*, in *Proc. Interspeech*.

Kisseberth, Charles; Abasheikh, Mohammad Imam (1974). Vowel length in Chi Mwi:ni—a case study of the role of grammar in phonology. In A. Bruck, R.A. Fox, and M.W. LaGaly (eds.) *Papers from the Parasession on Natural Phonology.* Chicago: Chicago Linguistic Society. 193-200.

Kochanski, G.; Grabe, E.; Coleman, J.; Rosner, B. (2005). Loudness predicts prominence; fundamental frequency lends little. *J. Acoustical Society of America* 118(2), 1038–1054

Linguistic Data Consortium (2004), *Meeting Room Careful Transcription Guidelines*, technical report version 1.2, 1/16/2004.

Linguistic Data Consortium (2009), *Rapid Transcription Guidelines*, downloaded April 20, 2009 from http://www.ldc.upenn.edu/Transcription/quick-trans/index.html.

Liu, Y.; Shriberg, E.; Stolcke, A.; Hillard, D.; Ostendorf, M.; and Harper, M. (2006). Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies. *IEEE Trans. Audio, Speech and Language Processing*, 14(5): 1526-1540.

Livescu, K.; Bezman, A.; Borges, N.; Yung, L; Cetin, O.; Frankel, J.; King, S.; Magimai-Doss, M.; Chi, X.; and Lavoie, L. (2007). *Manual transcription of conversational speech at the articulatory feature level*,. In Acoustics, Speech and Signal Processing, 2007, vol. 4, pp. 953-956.

Mitchell, Tom M. (1997). *Machine Learning.* New York, NY: McGraw Hill.

Nakatani, Christine H.; Hirschberg, Julia; and Grosz, Barbara J. (1995). *Discourse structure in spoken language: Studies on speech corpora.* Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation.

Ostendorf, Mari; Price, Patti; Shattuck-Hufnagel , Stefanie (1996). *Boston University Radio Speech Corpus,* Linguistic Data Consortium, Philadelphia.

Pitt, Mark A.; Dilley, Laura; Johnson, Keith; Kiesling, Scott; Raymond, William; Hume, Elizabeth; and Fosler-Lussier, Eric (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pitrelli, John F.; Beckman, Mary E.; and Hirschberg, Julia (1994). "Evaluation of prosodic transcription labeling reliability in the tobi framework", In *ICSLP-1994*, 123-126.

Schafer, Amy J., Speer, Shari R., and Warren, Paul (2005). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In Tanenhaus, M. and Trueswell, J. (eds.) *Approaches to Studying World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Tradition*. Cambridge: MIT Press.

Stevens, Kenneth N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features*. Journal of the Acoustical Society of America,* 111(4): 1872-1891.

Syrdal, Ann K.; Hirschberg, Julia; McGory, Julie; Beckman, Mary (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody, *Speech Communication*, 33: 135–151.

Turk, Alice E. and Shattuck-Hufnagel, Stefanie (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics,* 35(4): 445-472.

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

Webb, Andrew. (1999). *Statistical Pattern Recognition.* London: Arnold (Newnes).

Westbury, John R. (1994). *X-ray microbeam speech production database user's handbook, version 1.0*, Madison, WI. [http://www.medsch.wisc.edu/~milenkvc/pdf/ubdbman.pdf].

Weide, Robert (1995). *The Carnegie Mellon Pronouncing Dictionary (cmudict)*, technical report version 1.4, 11/8/1995

Wightman, Colin W.; Shattuck-Hufnagel, Stefanie; Ostendorf, Mari; Price, Patti J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91: 1707–1717.

Wightman, Colin W.; Ostendorf, Mari (1994). Automatic labeling of prosodic patterns. *IEEE Trans. Speech Audio Proces.*, 2(4): 469–481.

Yoon, Tae-Jin (2007). *A Predictive Model of Prosody through Grammatical Interface: A Computational Approach.* Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Yuan, Jiahong; Brenier, Jason M., Jurafsky, Dan (2005). Pitch Accent Prediction: Effects of Genre and Speaker. In *Proceedings of EUROSPEECH-05*

[1] The Linguistic Data Consortium currently distributes speech databases for these languages: Arabic, Croatian, Czech, Dschang, English, Farsi, German, Hindi, Japanese, Korean, Mandarin, Nbomba, Portuguese, Russian, Spanish, Tamil, Turkish, Urdu, and Vietnamese. See also **Loehr and Van Guilder, this volume**, for other languages and resources.