

## Converging results from phonological and phonetic analyses: the role of corpora

Renate Raffelsiefen and Anja Geumann (IDS Mannheim)

English-German cognates such as *thief* - *Dieb*, *shoe* - *Schuh* suggest a close phonological similarity between the respective high vowels, reflected in near-identical representations in phonological and phonetic vowel charts in comparative grammars (König & Gast 2012). However a phonological analysis based on an examination of neutralization patterns indicates the development of fundamental structural differences, which are reflected in significant differences in the dynamic spectral changes of the vowels in question. Section A briefly motivates the structural differences along with the relevant phonetic correlates, section B details the role of corpora in the investigation.

A. All German vowels contrast before other vowels (e.g. *Ch/a.ɔ/s* <Chaos>, *F/i.a/sko* <Fiasko>), whereas in English only diphthongs and high tense vowels occur before other vowels (e.g. *Isr/i.ə/l* <Israel>, *kar/i.əu/ke* <karaoke>). Given the universal affinity between high vowels and syllable margins the relevant restriction in English indicates the association of high vowels with both nucleus and coda in that language, analogous to the representation of diphthongs, as opposed to the organization of other monophthongs with the nucleus alone. The uniform behavior of all German vowels in the hiatus context indicates a uniform prosodic organization of all monophthongs as nuclei in German. Whereas phonetic comparisons of the steady states of the vowels in question fail to indicate relevant differences in the two languages, comparisons of dynamic spectral changes do show such differences: German /i/ and /u/ exhibit relatively short trajectories (in the F1 X F2 vowel space) directed towards the center of the vowel space, whereas English /i/ and /u/, like diphthongs, exhibit rather long trajectories directed towards the outer limits of the vowel space. The phonological and phonetic evidences hence converge to indicate an additional margin association of English /i/ and /u/, as opposed to German /i/ and /u/, which form exclusive nuclei.

B. To establish the distributional regularities in question, specifically the absence of all monophthongs other than /i/ and /u/ in prevocalic position in English, a searchable electronic data base consisting of phonetic transcriptions (Baayen et al. 1995) was used.

Our further analysis is based on acoustic corpora. Based on the findings in Morrison & Assmann (2013) that English monophthongs can be better described using additional dynamic information, we used a similar methodology for comparing German and English vowels described above. The important prerequisite for this kind of analysis (beside a certain recording quality) is to have fairly precise temporal annotations for the data. The chances of having available large acoustic data sets for different languages set up uniformly for direct comparison are sparse. Thus we use here different types of corpora for comparison. The data set for American English was originally designed and used for automatic speech recognition training and testing purposes. Good quality acoustic recordings together with hand verified segmental annotations are made available in the TIMIT database (Garofolo et al. 1993). The German data stem from the Deutsch Heute (Brinckmann et al. 2008) corpus as well as the Kiel Corpus of Read Speech (Kohler, ed. 1992). The Deutsch Heute data have been partially further segmentally annotated, so that metadata are provided for looking at fine phonetic detail.

We argue here for using corpora and metadata that match certain quality standards. In addition we note the importance of making the time consuming effort to segment data in accordance with phonemic analysis, either manually or semi-automatically in an open format.

### References

- Baayen, R. H., Piepenbrock, R.; Gulikers, L. (1995) CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium.
- Brinckmann, C.; Kleiner, S.; Knöbl, R.; Berend, N. (2008) German Today: an areally extensive corpus of spoken Standard German. In: Proceedings 6th International Conference on Language Resources and Evaluation (LREC 2008), Marrakesch, Marokko.

- Garofolo, J., et al. (1993) TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium.
- König, E.; Gast, V. (2009) Understanding English-German Contrasts. Erich Schmidt Verlag: Berlin.
- Kohler, Klaus J., Ed. (1992) Phonetisch-Akustische Datenbasis des Hochdeutschen. Kieler Arbeiten zu den PHONDAT-Projekten 1989-1992. (=Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 26).
- Morrison, G. S.; Assmann, P. F., eds. (2013) Vowel Inherent Spectral Change. Springer: Berlin, Heidelberg.