Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), Genova, Italy, May 24-26, 2006

"Casselberveetovallarga" and Other Unpronounceable Places: The CrossTowns Corpus

Stefan Schaden, Ute Jekosch*

* Lehrstuhl für Kommunikationsakustik, Technische Universität Dresden, Germany
stefan.schaden@freenet.de ute.jekosch@tu-dresden.de

Abstract

This paper presents a corpus of non-native speech that contains pronunciation variants of European city names from five countries spoken by speakers of four native languages. It was originally designed as a research tool for the study of phonetic variants of non-native speakers in the pronunciation of foreign city names. The corpus has now been released. Following a brief sketch of the research context in which this data collection was established, the first part of this paper describes the contents and technical specifications of the corpus (design, speakers, language material, recording conditions). Compared to corpora of native speech, non-native speech compilations raise a number of additional difficulties that require specific attention and methodology. Therefore, the second part of the paper aims to point out some of these general issues from the perspective of the experience gained in our research. Strategies to deal with these difficulties will be explored along with their specific benefits and shortfalls, concluding that non-native speech corpora require a number of specific design guidelines which are often difficult to put into practice.

1. Introduction

The speech data collection described in this paper has a very specific focus of interest with respect to both vocabulary and speakers. The CrossTowns corpus is devoted to the pronunciation of foreign city names by non-native speakers from various European languages. A general overview of the purpose and design of this corpus was already given in an earlier paper, when the data collection was still in its initial stage (Schaden, 2002). A subset of this data has now been released. The corpus may be used for research and development purposes in the contexts of speech recognition or speech synthesis, for empirical linguistic research on L2 pronunciation, or in any other related field in which the pronunciation of names by non-native speakers is a relevant issue.

The focus of the present article is twofold: In the first part, a description of the corpus design, contents and technical specifications in its current release version is provided. Following that, some general design issues and particular problems in designing and building non-native speech corpora (as opposed to resources of native speech) that we encountered during the data collection stage will be discussed. In order to avoid extensive overlaps with previous publications, the general research context in which the corpus was created will only be briefly touched upon. However, complementary information on the corpus and its design can be found in the previously mentioned article (Schaden, 2002). For an outline of the data analysis, the reader is referred to Schaden (2003b), or, for a more in-depth account of the research, Schaden (2006, forthcoming).

2. Corpus Design

2.1 General Remarks

With its particular domain of speech, the CrossTowns corpus does not enter entirely new ground. During the last ten years, a range of non-native speech corpora have been established (e.g. Benarousse et al., 2001; Byrne et al., 1998; Lamel et al., 1994; Menzel et al., 2000; the 'Strange Corpus' I and II provided by the Bavarian Archive for Speech Signals, and the Verbmobil 'Denglish' subcorpus), some of which were made available to the research community.

While most of these corpora have no particular emphasis in their vocabulary, other previous research has explicitly focused the domain of proper names. City names and their potential pronunciations by non-natives were covered e.g. by the *Onomastica Interlanguage Pronunciation Lexicon* (Onomastica Consortium, 1995), which contains phonetic transcriptions of non-native pronunciation variants of city names for a variety of European languages. Studies by Fitt (1995) dealt with the pronunciation of European town names by speakers of Scottish English. Both activities, however, did not result in publicly available corpora. Therefore, material for empirical studies in this domain is still sparse.

The release of the CrossTowns corpus can be viewed as an initial effort to remedy this situation. However, there is a caveat to be pointed out in advance: The corpus was not designed as a self-contained language resource from the start. Initially, it was set up as a research tool to identify common pronunciation errors of non-native speakers and to model these errors by rules on the lexicon level (see Schaden, 2003a). Therefore, just like many other project-specific data collections, the corpus reflects to some degree the (narrowed) research aims for which it was designed.

With regard to its general applicability, both benefits and drawbacks result from this descent. Since our research was led by the idea to build a unified rule-based framework that can be extended to new languages, it was decided to include multiple languages and language combinations from the start. The corpus currently covers sixteen L1/L2 combinations (*language directions*) altogether, as detailed in section 2.2. Thus, with respect to the number of languages covered, it can be viewed as a fairly rich language resource.

However, in order to avoid an explosion of the overall corpus size, some trade-offs had to be made. First, the vocabulary size had to be limited to a rather small set of 43 names per target language (see section 2.3.1 for details and examples). Secondly, the number of speakers per native language (L1) had to be restricted. The release version of the corpus includes 61 speakers. Yet despite these re-

strictions, the overall corpus includes about 26.000 utterances or 16 hrs. of speech material.

2.2 Languages and Language Directions

The material used for the central part of the data collection consists of 43 city names from each of the countries (1) Germany, (2) France, (3) Britain, (4) Italy, and (5) The Netherlands. This amounts to a total of 215 city names for the entire corpus. This material was spoken by native speakers of (a) German, (b) French, (c) Italian, and (d) Spanish in a controlled recording setting (see section 2.5 for details).

Not each possible L1/L2 combination resulting from this cross-language design is covered. In the present release, English and Dutch are included only as target languages (i.e. there are no native speakers of these L2s), while Spanish is present only as a native language. Accordingly, the corpus covers a total of 16 language directions, as shown in the following listing:

Native language (L1)	**Target languages (L2)**
(1) German	English, French, Italian, Dutch
(2) French	German, English, Italian, Dutch
(3) Italian	German, English, French, Dutch
(4) Spanish	German, English, French, Dutch

Table 1: Language pairs and directions in the corpus, number of speakers per native language

In addition to the cross-language recordings, some speakers were asked to pronounce the names of their own native language. Some of this material was applied as acoustic stimuli for parts of the recordings (see section 2.5). These recordings are not part of the corpus distribution (but can partially be made available upon request).

2.3 Linguistic Material

2.3.1 City Names

The selection of prompts is a crucial issue for non-native speech recordings, since the prompts themselves determine to some extent which pronunciation errors will actually be observable in the data.

Perhaps the most obvious selection criterion that suggests itself is the importance or international relevance of cities. Following this guideline, we could have compiled lists of the *N* largest cities of each country to apply them as recording prompts. However, this approach would probably have missed some of the more interesting features of non-native pronunciations, since names such as *Hamburg*, *London*, *Marseille, Amsterdam* tend to have a relatively stable and standardised pronunciation in other languages (e.g. due to their frequent appearance in the media, which often has a considerable normative effect on pronunciation). As a consequence, it can be expected that their pronunciations will be rather uniform or even predictable. This would be a major drawback for studies in which inter-individual variability of pronunciation is a focus of interest, as it was in our own research.

Moreover, many large or internationally relevant cities have lexicalised *exonyms* in foreign languages (e.g. English *Cologne* for German *Köln*; German *Genua* for Italian *Genova*). Stimuli of this kind were avoided as well to rule out the possibility that speakers replace the native forms (*endonyms*) of the names by the corresponding exonyms during the recordings.

After applying these exclusion criteria, an important positive selection criterion was an adequate coverage of orthographic or phonetic elements that were expected to raise particular difficulties for non-native speakers. Obviously, this criterion is based on subjective preconceptions about non-native speech (which may be mistaken from the start or falsified during the subsequent data analysis). It could be objected, for instance, that the 'degree of difficulty' of particular L2 phonemes strongly depends on the speakers' L1 and its sound inventory and that a selection of prompts ought to be adapted to individual L1s for this reason[1]. Nonetheless, we chose the same L2 material for all native speaker groups. Otherwise there would have been no way to compare the pronunciations of individual words across different native languages at the data analysis stage.

Among the characteristics expected to be particularly critical for non-native speakers, there are features such as complex consonant clusters in German (e.g. *Fünfstetten*), complex orthographic vowel clusters in French (e.g. *Riom Châtelguyon*), or diacritically marked grapheme characters with special phonetic functions, such as <ü, ö, ë, ç, é, è, â>, of which all included languages apart from English make some use.

According to these criteria, we compiled a list of city names which are (a) likely to be unfamiliar to most non-native speakers and (b) contain features that will introduce some degree of difficulty to their pronunciation. Examples are:

German *Gellmersdorf, Aindling, Thränitz, Fünfstetten, Blankenstein*

English *Cheltenham, Hemel Hempstead, Shrewsbury, Toddington, Plymouth*

French *Fougères, Issoire, Questembert, Fréjus, Riom Châtelguyon*

Dutch *Bloemwijk, Purmerend, Enkhuizen, Goes, Arnemuiden*

Italian *Pegognaga, Roncobilaccio, Faenza, Chiavari, Battipaglia*

No attempt was made to obtain a phonetically balanced corpus. Phonetic balance is a valuable feature in some domains (e.g. ASR training databases), but it is not necessarily a useful criterion for the present type of corpus. For one thing, it is likely that particularly the less frequent L2 phonemes (underrepresented in phonetically balanced sets) will be interesting from the perspective of phonetic research. But more importantly, it is hardly possible to introduce true phonetic balance into a non-native speech corpus simply because we do not know in advance which speech sounds will actually be produced by the speakers.

[1] For example, it could be argued that the Dutch fricative [χ] will be easier to approximate by speakers of an L1 that has a similar sound, such as German [x]. In practice, however, this similarity may lead to effects of 'equivalence classification' which often *prevent* the exact reproduction of an L2 sound instead of facilitating it (see Flege, 1987).

Phonetically balanced *prompts* will therefore not necessarily result in phonetically balanced *speech data*.

2.3.2 The Sentences Subcorpus

In addition to the city names segment, which forms the central part of the data collection, the corpus includes a complementary subcorpus of 10 short sentences *without* city names from each of the target languages (a total of 50 sentences). The sentences were read by the speakers. This additional data section was initially compiled as a means to reassess the speakers' target language proficiency using continuously spoken utterances. However, since it may also be used as an independent collection of non-native speech data, it is included in the corpus distribution.

The sentences (extracted from various online newspapers) are not semantically interrelated and have a relatively straightforward syntactic structure, as illustrated by the following examples:

German	*Ich bin allerdings anderer Meinung.*
English	*This is how the market works*
French	*Qu'est-ce qu'on va faire?*
Dutch	*We wachten op concrete plannen.*
Italian	*La storia d'Europa ricominciava.*

Although speakers and language directions are basically identical to the city names segment, the number of speakers is smaller for the sentences subcorpus. This restriction is caused by the fact that unlike the city names tasks, the sentence task was not obligatory for all speakers and all target languages. Subjects were only asked to read the sentences if they had (at least) some basic knowledge of the target language(s) and felt comfortable with the task. As a result, many speakers performed the sentence reading task for the target languages German and English, whereas relatively few speakers did so for Italian, French, and particularly Dutch.

2.4 Speakers

The majority of speakers were university students who were either exchange students from various European countries or – for the group of German speakers – native Germans. For this reason, there is a strong prevalence of the age group 20–30 years in the overall corpus (see 3.2 for a discussion). In order to correct this imbalance at least partially, an additional group of 20 native speakers of Italian was recorded in their native country (Udine, North Italy) at a later stage. The structure of the latter group differs from the former with regard to age and L2 proficiency levels. Within this group, there are 10 subjects aged 40–65 years with significantly lower proficiency levels in most target languages. The following table shows an overview of all speakers in the corpus:

L1	**No. of speakers**	**male / female**	**age group 19-31**	**age group 40-65**
German	24	17 / 7	24	--
French	4	3 / 1	4	--
Italian	24	8 / 16	14	10
Spanish	9	3 / 6	9	--
TOTAL	**61**	**31 / 30**	**51**	**10**

Table 2: Speaker distribution, age groups

For all speakers, the following types of information are provided with the corpus:

- Speaker ID
- Age, sex, profession
- Native language (or language spoken during the first 10 years of life)
- Country and region of origin
- L2 proficiency level according to self-judgement
  - (a) for languages covered by the database
  - (b) for additional languages (if applicable)

Language proficiency levels were assessed through self-judgment by the speakers on a scale that ranges from 0 (= no knowledge at all) to 5 (= native speaker of the respective L2). A non-native speaker with excellent L2 proficiency can thus achieve a maximum rating of 4. Since self-judgments yielded a sufficiently precise value for the purposes of our research, more elaborate, formal proficiency tests were not conducted.

2.5 Recording Setting and Tasks

All recording sessions were introduced by an oral instruction in the course of which it was emphasized that the recordings are *not* designed as an L2 proficiency test. We imposed no time restrictions to complete the tasks. If false starts or hesitations occurred, a repetition of the relevant item was not only permitted, but obligatory, since at least one full pronunciation of each item was required from each speaker. The linguistic origins of the material (i.e. the target languages) used were known to the subjects at each stage of the recordings.

All utterances were recorded in an experimental setting that included 3 subtasks, performed in the following order:

Task (1) – city names, read speech. 45 names[2] were read by the subjects as isolated words from a prompt sheet. The names were listed in random order[3].

Task (2) – city names, repetition of native speech. In this task, subjects were asked to repeat 43 names spoken by native speakers of the target language. The stimuli were presented by headphones using an interactive computer-based setup that permitted listening to prompts repeatedly.

[2] In the reading task, two duplicate names were included to control the consistency of individual speakers' pronunciations; therefore the number of names is 45 instead of 43.

[3] The prompt sheets are available independently of the speech corpus (please contact authors).

Task (3) – sentences, read speech. In this optional subtask (see section 2.3.2 above), the subjects read 10 sentences of selected target languages from a prompt sheet.

The same city names were used in both subtasks (1) and (2). In order to avoid potential learning effects caused by previously listening to the correct native pronunciations, subtask (2) was performed subsequent to (1). In addition, the presentation order of the prompts was changed for both subtasks.

As a result of this experimental setting, each spoken name is present in the corpus in two different production modes (read and repeated). This design was chosen to enable the identification of pronunciation errors that are specific of read speech and their distinction from errors that are genuinely phonetically motivated (i.e. independent of the orthographic representation). In fact, the data analysis showed that there are often marked differences in pronunciation between the two production modes.

2.6 Summary of Corpus Specifications

Major parts of the speech recordings were conducted at the University of Bochum (Germany) in a fixed recording setting. As outlined above in 2.4, additional recordings took place in Italy. It was attempted to reproduce the recording conditions in the best possible way for these recordings, but minor differences and reductions in audio quality had to be accepted. The specifications that apply to all recordings (regardless of location) are included in the following table, which at the same time provides a summary of all relevant corpus properties:

Speakers	61
No. of native languages	4 (GER, FR, IT, SPA)
No. of target languages	5 (GER, EN, FR, DUTCH)
Types of utterance (and # items per speaker)	- city names, read (225) - city names, repeated (215) - sentences, read (variable number)
Lexicon	- orthographic - canonical phonetic (SAMPA)
Documentation	- speaker-related information (ID, age, sex, native language, L2 proficiency levels) - prompt sheets - general documentation
Environments	(1) noise-controlled cabin (2) small room (2 x 2 mtrs.)
Microphone	AKG C 414 at 30 cm distance to mouth
Sampling Rate	22,05 kHz, 16 bit resolution
Channels	1 (mono)
Duration	16 hrs. (approx.)
No. of utterances	26.000 (approx.)
Storage media	1-2 DVDs

Table 3: Summary of corpus specifications

3. General Design Issues of Non-Native Corpora

Compared to corpora of native speech, the design of non-native speech data collections raises a number of additional difficulties that require specific attention and methodology. It is the aim of the remaining sections to highlight some of these problems based on the experience of our own data collection efforts. Most of them remained unresolved in the present corpus. So rather than presenting ready-made solutions for these potential problems, the following sections will point to some open questions with respect to the design of non-native speech data resources in general.

3.1 Elicitation Methods

The non-native speech samples for the CrossTowns corpus were recorded in an controlled, 'experimental' setting. This is a common approach in building speech resources. However, although some kind of controlled setting is a necessary requirement in order to produce a structured data collection with predictable linguistic content, it is arguable whether this is an ideal framework to elicit non-native speech.

To begin with, a problem that we regularly encountered during the recordings was that subjects felt embarrassed to some degree when asked to pronounce L2 material for permanent storage on recordings. Although each session was introduced by an oral instruction in which it was (almost over-) emphasized that there is no L2 proficiency test involved, some subjects still appeared to feel uncomfortable with the task (see also Tomokyio & Burger, 1999 on this problem). It seems generally difficult to overcome this problem in controlled recording settings.

There are reasons to assume that particularly for non-native speech, this is more than a negligible psychological side-effect. Rather, it may affect the data itself. According to Krashen (1981), it is a fundamental difference between native and foreign language usage that the latter is far more often subjected to conscious control and reflection about the "correctness" of utterances, whereas the native language is acquired and generally used subconsciously with respect to its linguistic form. The degree of self-monitoring varies from one speaker to another, but often it affects the L2 performance directly.

The divergence between "laboratory" and "natural" settings may be particularly marked for non-native speech. An artificial situation like a recording session is likely to *increase* the tendency of self-monitoring already present in non-native speakers even further, and expectations of being tested will certainly not help to improve this situation. So-called 'natural' non-native speech (outside speech labs or language lessons), in contrast, is characterised by *decreased* self-monitoring, since the speaker's attention will be drawn away from the linguistic form towards the overall act of communication.

So even if we do not know exactly what the specific effects of increased self-monitoring on L2 pronunciation are, we should stay aware that some of these influences may be induced by the recording setting itself. It is an unresolved research question which recording design is optimally suited to control these effects without abandoning the demand for a fixed, predetermined vocabulary in the corpus. One possible future direction might be a variant of

'Map Tasks' (Brown et al., 1984), whose basic idea it is to engage subjects in a cognitively demanding dialogue while studying their speech in order to elicit spontaneous utterances. An adaption of this basic design for the particular requirements of non-native speech research may help to control effects of increased attention to linguistic form during recordings.

3.2 Speaker Recruitment

On-site speech data collections bound to a particular location often face the problem that access to speakers is geographically limited. Since it is usually too costly and inefficient to recruit speakers who have to travel long distances to the recording site, the selection of speakers is often limited by these basic local restrictions.

In the case of non-native speech corpora, this condition obviously turns into a problem. The number of suitable non-native speakers within direct reach is usually severely limited, and even if a sufficient number of speakers can be identified and addressed, their availability for recordings is not granted. Particular cultural problems may further complicate the situation for some non-native speaker groups (see Draxler, 2003).

For the CrossTowns corpus, we therefore recruited speakers among university exchange students from various European countries who stayed in Germany for only a limited period of time. With respect to the efficiency of the data collection, this strategy combines two advantages: First, it is relatively easy to reach sufficiently large groups of native speakers of different languages, since there is a strong social coherence within each of these groups that often results in a "snowball" effect during the recruitment. Secondly, these speakers are usually within local reach, so arrangements for recordings can be readily made.

However, there is an obvious negative effect on the overall speaker distribution with respect to the variables *age* and *L2 proficiency*. Hence, in the CrossTowns corpus, there is a significant overrepresentation of the age group 20 to 30 years (as reflected in table 2). Provided that age is an independent variable that influences the phonetic L2 performance of speakers, as indicated by some of the literature (see e.g. the remarks by Eklund & Lindström, 2001), this can be viewed as a deficit of the present data collection.

A similar imbalance emerges with respect to L2 proficiency levels for some target languages: Compared to larger populations, the foreign language proficiency of university exchange students is likely to be above average. In our own data collection, this is particularly true for the target languages German, English, and, to a lesser degree, French[4]. This imbalance could partially be fixed by the additional recordings in Italy, but for practical reasons, this effort could not be repeated for each relevant country.

Judging from the particular problems that we experienced, it must be concluded that for really large data collections of non-native speech that include multiple native speaker groups, there is currently no viable alternative to international collaborations with distributed recording sites. However, newly emerging methods such as web-based recording set-ups (see Draxler & Schiel, 2002) that allow a geographical detachment of recording site and speakers may develop into a useful technique for future data collections in the domain of non-native speech.

[4] However, for both Italian and Dutch, the average proficiency levels are rather low.

3.3 Data Representativeness

Data representativeness is generally viewed as an essential requirement for speech corpora. The central idea of representativeness is that a corpus, though limited in size, should optimally represent the manner of speaking of larger populations of speakers. In order to achieve this, a set of speaker-related variables, such as age, sex, dialectal origin, or education is applied in such a way that the distribution of these variables is optimally balanced over the entire corpus. This strategy is based on the (implicit or explicit) assumption that each of these variables bears a particular effect on the manner of speaking.

Although these commonly used variables are certainly useful for non-native speech corpora as well, they are probably not sufficient in this domain. In studies of non-native speech, the focus of interest is rarely on all conveiveable aspects of the speech, but rather on its *deviations* relative to some target form (usually the canonical L2 form). So before we can set out to build a representative corpus, we need to answer the fundamental question which speaker-related variables may cause this deviation to what extent.

For non-native speech, however, this question is even more difficult to answer than for native speech. Variables such as *native language*, *age*, *sex*, *education*, *general L2 proficiency*, *initial age* and *mode of L2 acquisition*, *degree of regular exposure to L2*, as well as some (stable or transient) speaker characteristics like *motivation*, *extraversion, empathy with interlocutor*, *degree of attention*, or *degree of anxiety* are just a few of the variables that have been suggested and studied in second language acquisition research as potential influences on the L2 performance of speakers.

However, unless we know exactly which of these variables shape the phonetic form of non-native speech to what extent, defining a representative corpus of non-native speech will be extremely difficult. But even if we do know, the practical aim of building a corpus that is balanced with respect to each of these variables would be an enormous challenge, especially in view of the fact that large amounts of representative data are required for particular applications (e.g. ASR training).

For these reasons, it is currently a more realistic aim to stick to the practice of restricting non-native speech corpora to specific speakers, domains, and situations, instead of attempting to build general-purpose data collections. Even if there is a need to extend the common design specifications when dealing with non-native corpora, this approach is still capable of providing valuable insights about the nature of foreign-accented speech.

4. Concluding Remark: What is "Casselberveetovallarga"?

The word "Casselberveetovallarga" is a pronunciation variant of the Italian town name *Castelbentivoglio* spoken by a native speaker of English (an approximate phonetic transcription would be [kæsɛlbɜviːtovalʼɑːgə]). Although this is certainly not a typical example of the pronunciations we recorded, it illustrates well the important point that non-native pronunciations are sometimes quite dis-

tinct from what theoretical second language research predicts them to be. This is viewed as further evidence that more empirical data from this domain is required in the future.

Acknowledgements

Major parts of this work were supported by the *Deutsche Forschungsgemeinschaft* (DFG) within a research project carried out at the Institute of Communication Acoustics, University of Bochum (Prof. R. Martin, Prof. J. Blauert). The authors would like to thank all colleagues at Bochum and Dresden for their support. Particular thanks go to Anders Krosch, Andreea Niculescu, Katja Kesselmeier and Anneli von Könemann for their assistance in the data collection.

References

Benarousse, L.; Geoffrois, E.; Grieco, J.; Series, R.; Steeneken, H.; Stumpf, H.; Swail, C.; Thiel, D. (2001): "The NATO native and non-native (N4) speech corpus". Proceedings Workshop on Multilingual Speech and Language Processing, Aalborg [on CD-ROM].

Brown, G.; Anderson, A.; Shillcock, R.; Yule, G. (1984): *Teaching Talk*. Cambridge: Cambridge University Press.

Byrne, W.; Knodt, E.; Khudanpur, S.; Bernstein, J. (1998): "Is automatic speech recognition ready for non-native speech? A data collection effort and initial experiments in modeling conversational hispanic English". Proceedings ESCA Workshop *Speech Technology in Language Learning* (STiLL), Marholmen, Sweden [on CD-ROM].

Draxler, Chr.; Schiel, F. (2002): "Three new corpora at the Bavarian Archive for Speech Signals - and a first step towards distributed web-based recording". *Proceedings Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain.

Draxler, Chr. (2003): "Orientel: Recording telephone speech of Turkish speakers in Germany". *Proceedings Eurospeech 2003*, Geneva, 1557-1560.

Eklund, R.; Lindström, A. (2001): "Xenophones: An investigation of phone set expansion in Swedish and implications for speech recognition and speech synthesis". *Speech Communication* 35 (1-2), 81–102.

Fitt, S. (1995): "The pronunciation of unfamiliar native and non-native town names". Proceedings *Eurospeech 1995*, Madrid, Vol. 3, 2227–30.

Flege, J. E. (1987): "Effects of equivalence classification on the production of foreign language speech sounds". In: James, A; Leather, J. (eds.): *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris, 9-39.

Krashen, S. D. (1981): *Principles and Practice in Second Language Acquisition*. London: Prentice-Hall.

Lamel, L. F.; Schiel F.; Fourcin, A.; Mariani, J.; Tillmann, H.G. (1994): "The Translanguage English Database (TED)". *Proceedings ICSLP 1994*, Yokohama, 1795–1798.

Menzel, W.; Atwell E.; Bonaventura P.; Herron D.; Howarth P.; Morton R.; Souter C. (2000): "The ISLE corpus of non-native spoken English". *Proceedings 2nd International Conference on Language Resources and Evaluation (LREC 2000)*, Athens, Vol. 2, 957–963.

Onomastica Consortium (1995): "The Onomastica Interlanguage Pronunciation Lexicon". *Proceedings Eurospeech 1995*, Madrid, Spain, 829–832.

Schaden, S. (2002): "A database for the analysis of cross-lingual pronunciation variants of European city names". *Proceedings Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain, Vol. 4, 1277-1283.

Schaden, S. (2003a): "Generating non-native pronunciation lexicons by phonological rules". *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 2003)*, Barcelona, 2545-2548.

Schaden, S. (2003b): "Non-native pronunciation variants of city names as a problem for speech technology applications." *Text, Speech and Dialogue. Proceedings of the Sixth International Conference TSD 2003*. České Budějovice, Czech Republic. Berlin, Heidelberg, New York: Springer, 229-236.

Schaden (2006) [forthcoming]: *Regelbasierte Modellierung fremdsprachlich akzentbehafteter Aussprachevarianten. Grundlagen, Entwurf und Implementierung eines Regelsystems für sprachtechnologische Anwendungen*. Ph.D. Thesis, University of Duisburg-Essen, Germany.

Tomokiyo, L.M.; Burger, S. (1999): "Eliciting Natural Speech from Non-Native Users: Collecting Speech Data for LVCSR". *Proceedings ACL-IALL Joint Workshop on Computer Mediated Language Assessment and Evaluation in NLP*, Maryland.