



ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ

Faculty of English

# The Greater Poland Speech Corpus: Description

Małgorzata Kul & Paulina Zydorowicz



# Acknowledgements

---

- the Corpus of Modern Spoken Polish in the area of Greater Poland
- Research project funded by the Ministry of Science and Higher Education; 0113/NPRH2/H11/81/2013 (2013-2017)
- Special thanks to all the researchers in the academia for all their inspiring feedback and support. We are indebted to dr Kamil Kaźmierski for his advice, guidance and technical assistance.

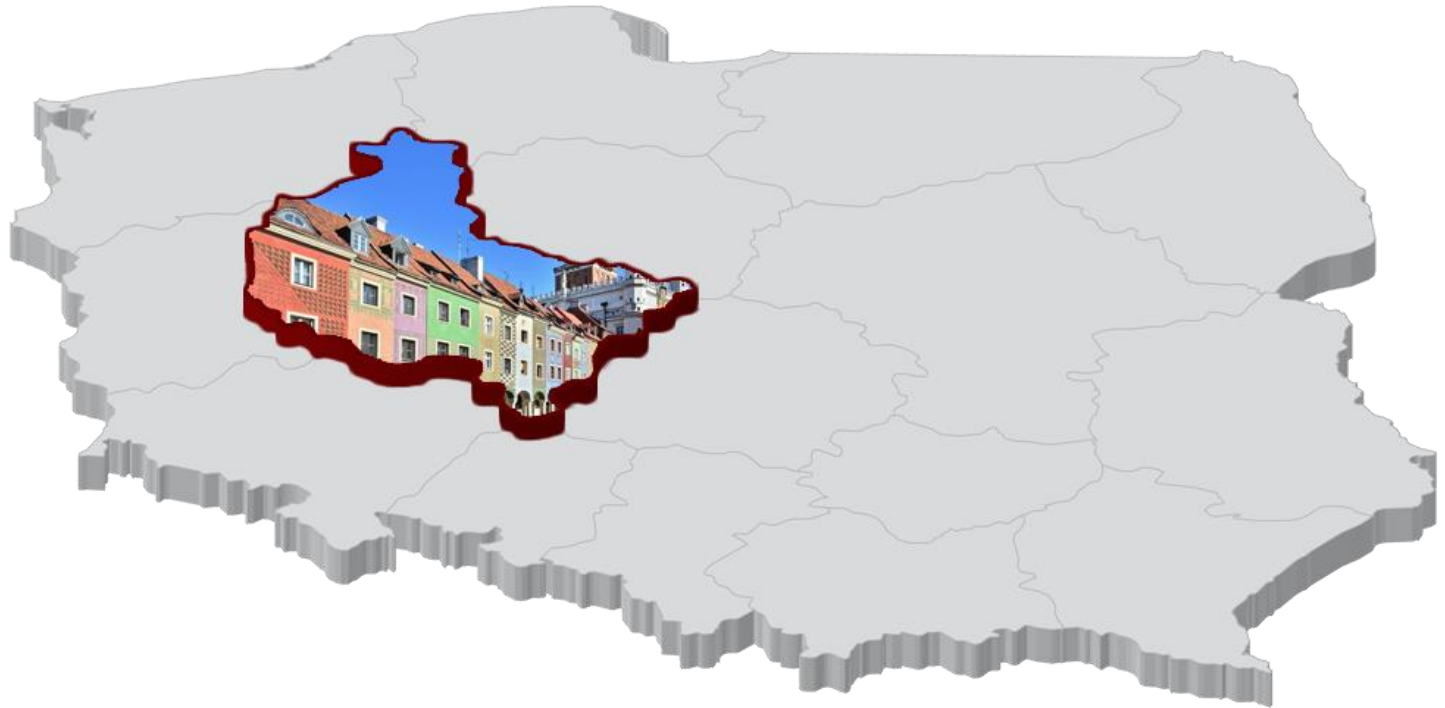


Ministerstwo Nauki  
i Szkolnictwa Wyższego



# Greater Poland

---



# The Corpus: Format

---

- 2+2 interview format
  - 2 interviewers
  - 2 interviewees
- speakers know each other





# The Corpus: Format

---

- reading keywords embedded in a carrier phrase (soundproof lab)
- an informal 20-40-minute conversation (friendly environment, quiet room at university, at home or in a workplace)
- Topics covered:
  - studies / work
  - living in Poznań
  - culture and entertainment in Poznań
  - the Internet
- metadata questionnaire: background information about subjects

# The Corpus: Equipment

---

- recorder: Roland R-26
- microphones: Rode, lavalier





# The Corpus: Speakers

---

94 speakers of Standard Polish as well as the vernacular





## The Corpus: Structure (1)

---

- recordings of 94 speakers
- 73 transcripts of spontaneous conversation (68 speakers with text grids ready to process automatically in systems like LaBBCAT; 34 speakers with .doc files)
- a sample of phonemic transcription (9 speakers); the transcription was added automatically; 3 reduction processes, namely cluster reduction, intervocalic *w*-deletion and assimilation of /t/ + /ɕ/ sequences) were manually annotated
- transcripts and audio files are available in an electronic version for scientific purposes
- website: <http://wa.amu.edu.pl/korpuswlkp/>





## The Corpus: Structure (2)

---

- Folders are coded in the following way: M24\_25, which reads as MÓWCA (=speaker) 24\_25.
- Folders (M3-73) contain sound files with a spontaneous conversation, coded as R\_M24\_25, word lists, coded as L\_M24 and L\_M25, text grids or .doc files with transcripts as well as background questionnaires.
- Folders L01-L21 contain files with a spontaneous conversation and background questionnaires.
- Speakers M1 and M2 were recorded in a cafe (as part of the pilot).
- Speakers M3-73 were recorded in a quiet room (word list in an anechoic chamber).
- Speakers M74\_75 were recorded in a quiet room in their work place.
- Speakers L01-L21 were recorded 'in the field'.



## The Corpus: Aims

---

- to build an **online** corpus of modern spoken Polish
- to capture the **truly spontaneous speech** (without resort to e.g. televised interviews, public or broadcast speeches)
- to compare / control **repeated items** (the same words in the informal speech as in the read part)
- to differentiate between **standard** and **nonstandard** speech
- national heritage: preserving the language and the vernacular
- to prepare the data for automatic processing, e.g. in LabbCAT



## Promotion

---

The corpus itself and the studies conducted on it have been promoted during a series of conferences and lectures.

Institute of Linguistics: Reading group REMAT, Poznań  
Celebrating 111 years of the School of English, Poznań  
(Building a corpus of spontaneous speech)

Phonologie Anglais Contemporain, Toulouse, 2015  
(A corpus-based study of vowel reduction in two speech styles: a comparison between English and Polish)



## Promotion

---

**Phonologie Anglais Contemporain, Aix-en-Provence, 2015**

(A quantitative study of reduction in English and Polish)

**46th Poznań Linguistic Meeting, Poznań, 2016**

(A corpus-based study of vowel reduction in two speech styles: a comparison between English and Polish)

**Polskie Towarzystwo Fonetyczne, Poznań, 2016**

(Internetowy korpus współczesnego mówionego standardowego języka polskiego i gwary na obszarze Wielkopolski)

**Accents , Łódź, 2017**

(A quantitative study of vowel reduction in English and Polish)

**47th Poznań Linguistic Meeting, Poznań, 2017**

(Poznań speech 30 years later)



# Thank you

---

**Małgorzata Kul & Paulina Zydorowicz**  
Adam Mickiewicz University



Contact us @  
[kgosia@wa.amu.edu.pl](mailto:kgosia@wa.amu.edu.pl)  
[zpaula@wa.amu.edu.pl](mailto:zpaula@wa.amu.edu.pl)

