FE logo, marks and signatures in accordance with AMU visual identity system – **LEAVE THIS SPACE BLANK**

Building a corpus of spontaneous speech

Małgorzata Kul, Paulina Zydorowicz and Geoffrey Schwartz (Faculty of English, Adam Mickiewicz University)
in collaboration with Liliana Madelska (Wien University)

Research project funded by the Ministry of Science and Higher Education (03.04. 2013–02.08.2016)

Website: http://wa.amu.edu.pl/korpuswlkp/ **Email:** korpuswlkp@wa.amu.edu.pl

1. The stimuli

- 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 maintain fair representation of education, age and gender across speakers
- To differentiate between standard and substandard speech

2. The aim

Building a corpus of modern spontaneous speech from the area of the Greater Poland as well as publishing the corpus online for both academic (research-oriented purposes) and non-academic (heritage-oriented) audience

3. The structure

The corpus is designed as 50 hours of recording of 76 participants from the region of the Greater Poland, recorded during the period of three years. Apart from sound files, orthographic transcripts are going to be provided

4. The participants

We aim to collect speech from both educated and uneducated subjects, varied across age and gender and representing both the standard and substandard speech

5. The recordings

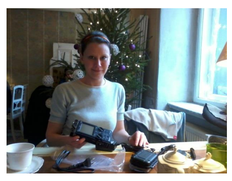
We make the recordings in a possibly informal environment, either in classrooms or workplaces of the subjects. A recording session has a classic Labovian design, including a) a wordlist b) a formal conversation c) an informal conversation

6. The present state of the corpus

- 20 speakers so far, recorded in pairs in a 2x2 conversation format (= 2 interviewers and 2 speakers)
- Conversation: 357.5 minutes
- Word list / carrier phrase: 34 minutes

M8: albo na przykład wystarczy zalajkować takie strony jak BBC czy CNN i tak naprawdę, wszystkie najważniejsze wiadomości się wyświetlają więc to też jest dosyć przydatne…i jak cos się dzieje, to od razu wiadomo, że coś się dzieje, bo…
M7: znaczy, na przykład, jak pada śnieg, nie? <śmiech>
M8: na Ukrainie.

M7: Znaczy no, chodzi o to, że ciężko jest na przykład, że ja wiem, że się muszę czegoś nauczyć i w pewnym momencie otwieram tego facebooka i jakoś już się trudno jest oderwać i potem ludzie do mnie piszą, co robisz, mówię, no uczę się, no jak się uczysz <śmiech> na facebooku się uczysz?



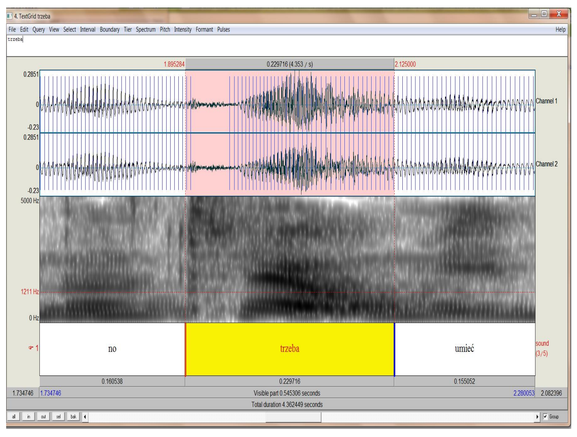
7. Prominent features of modern Polish

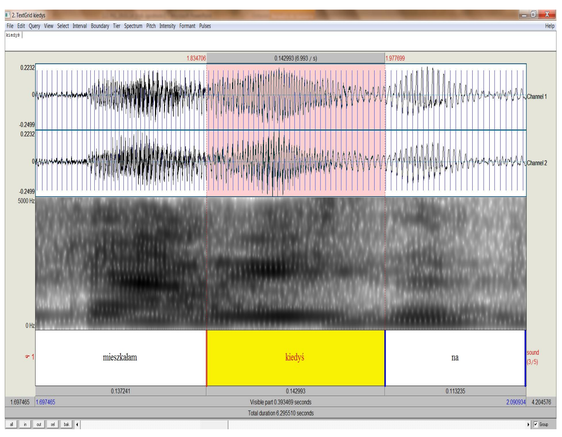
Reduced forms

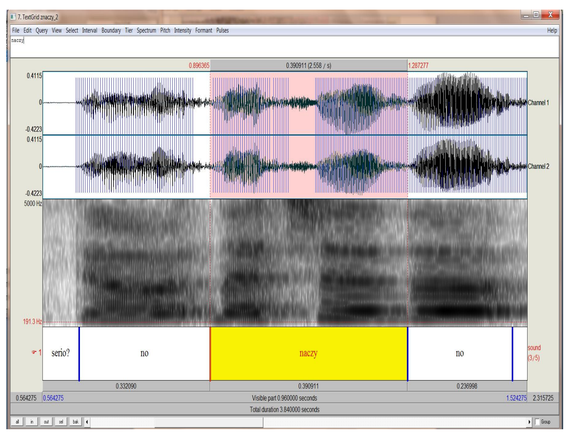
Kiedyś - kieś

znaczy się - naczy

Trzeba - czea







MK: pyszne pralinki. Ja mam takie pytanie, może niedyskretne, jak scena klubowa wygląda w tej chwili bo…tego panie nie wspomniałyście <śmiech>
M4: ja nie mam pojęcia, szczerze mówiąc
M3: ja też nie, nie lubię chodzić do klubów, uważam to za jeden wielki za przeproszeniem syf <śmiech> jest tam strasznie duszno, aam, ciężko o/o jakieś przyzwoite warunki w toaletach, po prostu no odpycha/odpychają mnie kluby. Wiem, że to jest takie dziwne, młoda osoba powinna iść do klubu i się wyszaleć ale ja wolę bardziej higieniczne <śmiech> środowisko.

MK: Jeszcze może takie niedyskretne pytanie, który kierunek jest taki imprezowy, niech Pani nam zdradzi.
M6: Bezpieczeństwo narodowe.
<śmiech>
PZ: Ok.
MK: Ok.
PZ: Dobrze.
M5: Stosunki międzynarodowe podobno też. Mój kolega powiedział, że za dużo alkoholu się wkradło w jego życie i że to już chyba podchodzi pod alkoholizm i że cały semestr przeimprezował i tylko teraz na sesji trochę ciężko.

PZ: mhm, okay, a ma pani jakieś negatywne doświadczenia z Jeżycami?
M3: Eee, tam są dziwni ludzie bardzo. <śmiech>
PZ: Proszę ich opisać.
M3: Mmm, uważam, że Jeżyce za bardzo podejrzaną dzielnicę Poznania, nawet czytałam artykuł, że jest tam zaraz po Wildzie, największa ilość popełnianych zbrodni [PZ: mhm] nie, wolę omijać Jeżyce szerokim łukiem.

M7: Tak samo, jak imprezę, ludzie organizują imprezę i zrobią wydarzenie i nie zrobią ustawienia, że to impreza jest tylko dla zaproszonych …więc były przecież już chyba w Polsce dwa takie przypadki, ze komuś z piętnaście tysięcy osób przyszło do domu <śmiech>
MK: i wtedy wkroczyła policja, to jest nie/nielegalne zgromadzenie [M7: tak] bo one wyszły w sumie później gdzieś tam na zewnątrz [M7: tak] i gdzieś palili ogniska też
M7: tak. Cała/ cała ulica podobno się bawiła tak, że [MK: ooo, to całe miasteczko chyba, nie?] noo [MK: gdzieś/gdzieś tam] więc może być niebezpiecznie bardzo na przykład
PZ: może być niebezpiecznie. A to są chyba ekstremalne przypadki, prawda?

8. The impact

a) heritage

To record and advance our knowledge of the speech in the area of the Greater Poland
To digitalize the regional substandard

b) research

to provide an up-to-date norm/reference point of the Greater Poland variety
to provide ample material for dialectological, lexical, phonetic and sociolinguistic studies

FE logo, marks and signatures in accordance with AMU visual identity system – **LEAVE THIS SPACE BLANK**