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

dr Małgorzata Kul i dr Paulina Zydorowicz

Faculty of English, Adam Mickiewicz University in Poznań

Projekt badawczy

Projekt badawczy finansowany ze środków Ministerstwa Nauki i Szkolnictwa Wyższego w ramach Narodowego Programu Rozwoju Humanistyki (moduł 1.1. dziedzictwo kulturowe), nr 0113/NPRH2/H11/81/2013

Ministerstwo Nauki i Szkolnictwa Wyższego

Plan prezentacji

1. Cele i zastosowania projektu
2. Metodologia korpusu
3. Stan obecny
4. Zespół
5. Badania prowadzone z użyciem korpusu

1. Cele i zastosowania projektu

Cele:

- zarejestrowanie współczesnej mowy z obszaru Wielkopolski w odmianie standardowej i gwarowej
- udostępnienie nagrań i transkryptów w formie internetowego korpusu dla celów badawczych i kulturowych

1. Cele i zastosowania projektu

Zastosowania:

- dziedzictwo kulturowe
- uwiecznienie języka mówionego na obszarze Wielkopolski
- cyfrowy zapis gwary

1. Cele i zastosowania projektu

Nauka:

- dokumentacja współczesnej, regionalnej polszczyzny
- najnowsza norma referencyjna poznańskiej mowy
- możliwość prowadzenia badań porównawczych (zmiany językowe)
- bogaty materiał do badań dialektologicznych, leksykalnych, fonetycznych i socjolingwistycznych

1. Cele i zastosowania projektu

- trenowanie systemów rozpoznawania mowy i podniesienia ich efektywności
- ukierunkowane rozmowy pozwolą zarówno na porównania fonologiczne między użytkownikami języka, jak i badanie wariantywności fonetycznej tego samego mówcy
- zebrany korpus będzie doskonałym źródłem do badań fonostylistycznych i fonotaktycznych (i nie tylko)

1. Cele i zastosowania projektu

- Możliwość prowadzenia badań porównawczych ze starszymi danymi oraz na obserwacje ewolucji polszczyzny. Podobne badania porównawcze były już prowadzone w przypadku zmian językowych w angielszczyźnie (Labov 1994, Preston 2003) oraz w przypadku języka niemieckiego (Piroth i Skupinsky 2010)
- istotny punkt odniesienia dla badań zaburzeń mowy (Połczyńska i Tobin 2011)
- Ponadto, brak jest w aktualnej literaturze norm referencyjnych, takich jak VOT (np. Waniek-Klimczak 2011, Wrembel 2011 odnoszą się do Keating et al. 1981)

Plan prezentacji

1. Cele i zastosowania projektu
2. Metodologia korpusu
3. Zespół
4. Stan obecny
5. Badania prowadzone z użyciem korpusu

2. Metodologia korpusu

Internetowy korpus współczesnego, mówionego języka polskiego i gwary na obszarze Wielkopolski został zbudowany zgodnie z metodologią, zastosowaną w korpusach współczesnego języka angielskiego (La Phonologie de l’Anglais Contemporain, Durand i Pukli 2004) i francuskiego (Phonologie du Français Contemporain, Durand, Laks i Lyche 2002)

2. Metodologia korpusu

Struktura korpusu

Korpus będzie stanowiło kilkadziesiąt godzin (minimum 60) nagrań 76 użytkowników języka z obszaru Wielkopolski (początek nagrań: listopad 2013)

2. Metodologia korpusu

określenie „standardowy” w tym projekcie

rozumiane jest w opozycji do gwary

2. Metodologia korpusu

Uczestnicy

Uczestnicy nagrań reprezentować będą zarówno regionalną mowę (56 osób), jak i gwarę poznańską (20 osób)

2. Metodologia korpusu

Uczestnicy

- Format 2+2
- Dr Zydorowicz, dr Kul
- Para znanych sobie osób

2. Metodologia korpusu

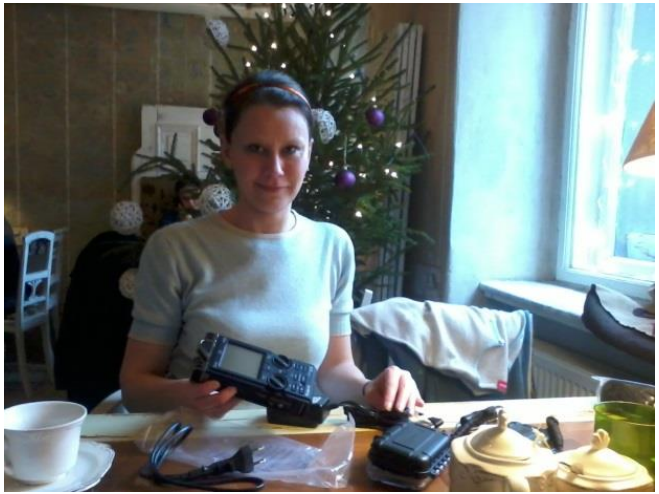
Nagrania

- Nagrania użytkowników standardowych zostały przeprowadzone w laboratorium mowy
- nagrania części gwarowej zostaną przeprowadzone przez
- wykonawców projektu w możliwie nieformalnym miejscu (np. pracy
- czy zamieszkania) i atmosferze, co zapewni ich autentyczność

2. Metodologia korpusu

Nagrania

Rozmowy rejestrujemy za pomocą rejestratora Roland R-26



Mikrofony przypinane Rhode

typu Lavalier

2. Metodologia korpusu

- Rodzaje nagrań:

- przeprowadzenie rozmowy *spontanicznej (*tematy: studia/praca, życie w Poznaniu, korzystanie z internetu)
- odczytanie listy słów (142 zdania)

Przykładowe nagrania

- rozmowa

- lista słów

2. Metodologia korpusu

Zapisy ortograficzne nagrań przy zachowaniu pełnej anonimowości będą udostępniane wyłącznie w celach badawczych

Plan prezentacji

1. Cele i zastosowania projektu
2. Metodologia korpusu
3. Stan obecny
4. Zespół
5. Stan obecny
6. Badania prowadzone z użyciem korpusu

3. Stan obecny

- Nagrania 73 mówców standardowych (2 pilotażowe; 6 sama rozmowa, na tej podstawie powstała lista; 65: lista, rozmowa,)
- 2 mówców gwarowych (praca w toku)

3. Stan obecny

- Anotacja (przeważnie) ostatnich piętnastu minut rozmowy w pliku tekstowym
- anotacja jest weryfikowana przez drugiego anotującego

2. TextGrid R_M22_23

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

18.900131 (0.053 / s)

0.935
0
-0.7884
Channel 1

0.935
0
-0.7884
Channel 2

(To see the analyses, zoom in to at most 10 seconds,

or raise the "longest analysis" setting with "Show analyses" in the View menu.)

1		sp				s							sp					sp			INT1 (483/1536)
2		sp				s							sp					sp			INT2 (1580)
3		sp			ale co	s		aha					sp					sp			M22 (1677)
4	nie wiem ja raczej jestem taką osobą która ze wszystkimi star	sp	nie wiem jak by to było w naszej	yyy	[un clea	s	chyba wiesz o kim myślę	yyy jeśli chodzi	na przykład o zajęcia gramatyk	[noise]	angielskiej	sp	o te dwie osoby dobra niewa		w każdym razie	sp	są osoby które na pewno dział				M23 (1662)
5																					ACTUAL (16)

502.368154 502.368154 Visible part 18.900131 seconds 521.268285 1122.010083

Total duration 1643.278367 seconds

all in out sel bak Group

PL 23:24 2016-10-16

3. Stan obecny

- przewidywany koniec części standardowej: grudzień 2016
- udostępnimy: nagrania, plik txtgrid, transkrypt ortograficzny, plik exel z metadanymi mówców
- przewidywany koniec części gwarowej: kwiecień/maj 2017

Plan prezentacji

1. Cele i zastosowania projektu
2. Metodologia korpusu
3. Stan obecny
4. Zespół
5. Badania prowadzone z użyciem korpusu

4. Zespół

- Projekt realizowany jest przez Zakład Współczesnego Języka Angielskiego na Wydziale Anglistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu
- Partnerem projektu jest Instytut Slawistyki Uniwersytetu Wiedeńskiego

4. Zespół

Wykonawcy projektu:

- dr Małgorzata Kul (kierownik projektu, Wydział Anglistyki UAM)
- dr Paulina Zydorowicz (Wydział Anglistyki UAM)
- dr Kamil Kaźmierski (Wydział Anglistyki UAM)
- dr Liliana Madelska (Instytut Slawistyki UWien)

Plan prezentacji

1. Cele i zastosowania projektu
2. Metodologia korpusu
3. Stan obecny
4. Zespół
5. Badania prowadzone z użyciem korpusu

5. Badania prowadzone z użyciem korpusu

- **Prowadzone obecnie:**

- Badanie redukcji samogłosek
- Badanie redukcji spółgłosek u tych samych mówców
- Badanie redukcji zbitek spółgłoskowych u tych samych mówców

Planowane:

- Percepcja form zredukowanych (e-prime)

Strona korpusu

- Strona korpusu

http://wa.amu.edu.pl/korpuswlkp/

- Zapraszamy do korzystania!