Jak zdobylam finansowanie wniosku z NPRH

Studenckie Koło Językoznawcze "Remat"
Instytut Językoznawstwa UAM

REMAT

18.06.13

Malgorzata Kul
Wydział Anglistyki

Cel wystapienia

- Przybliżę Wam, jak dostałam finansowanie projektu „Internetowy korpus współczesnego, mówionego standardowego języka polskiego i gwary na obszarze Wielkopolski”

Cel wystapienia

- Projekt badawczy finansowany ze środków Ministerstwa Nauki i Szkolnictwa Wyższego w ramach Narodowego Programu Rozwoju Humanistyki

- Okres realizacji: 03.04. 2013–02.08.2016

Ministerstwo Nauki i Szkolnictwa Wyższego

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU

WYDZIAŁ ANGLISTYKI

Plan

1. Cele i zastosowania projektu
2. Zespół
3. Metodologia korpusu
4. Zbieżność celów z profilem NPRH
5. Harmonogram

1. Cele projektu

Celem projektu jest zarejestrowanie współczesnej mowy z obszaru Wielkopolski w odmianie standardowej i gwarowej oraz 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 projektu

- Ze względu na objętość korpusu i liczbę uczestników nagrań, korpus doskonale nadaje się do trenowania systemów rozpoznawania mowy i podniesienia ich efektywności (Wypych 1999).

- 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.

1. Cele 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

1. Cele i zastosowania projektu
2. Zespół
3. Metodologia korpusu
4. Zbieżność celów z profilem NPRH
5. Harmonogram

2. 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.

2. Zespół

- Wykonawcami projektu są:

- dr Małgorzata Kul (kierownik projektu, Wydział Anglistyki UAM)
- dr Paulina Zydorowicz (Wydział Anglistyki UAM)
- dr Liliana Madelska (Instytut Slawistyki UWien)
- prof. UAM dr hab. Goeffrey Schwartz (Wydział Anglistyki UAM)

Plan

1. Cele i zastosowania projektu
2. Zespół
3. Metodologia korpusu
4. Zbieżność celów z profilem NPRH
5. Harmonogram

3. Metodologia korpusu

- **Struktura korpusu**

- Korpus będzie stanowiło kilkadziesiąt godzin nagrań 76 użytkowników języka z obszaru Wielkopolski, zbieranych w okresie 3 lat.

3. Metodologia korpusu

- **Uczestnicy**

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

3. Metodologia korpusu

- **Nagrania**

- Wszystkie nagrania zostaną przeprowadzone przez wykonawców projektu w możliwie nieformalnym miejscu (np. pracy czy zamieszkania) i atmosferze, co zapewni ich autentyczność.

3. Metodologia korpusu

- Internetowy korpus współczesnego, mówionego języka polskiego i gwary na obszarze Wielkopolski zostanie 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). Metodologia ta oparta jest na klasycznych opracowaniach Labova (1972) i Milroya (1980).

3. Metodologia korpusu

- Wg Milroya (1980), jedna wytypowana osoba zaangażuje w nagrania członków swojej rodziny oraz znajomych tak, aby nagrania zostały przeprowadzone w trakcie rozmowy pomiędzy znanymi sobie osobami (maksymalnie dwiema) w obecności wykonawcy projektu, zapewniając swobodny charakter rozmowy. Nagrania zostaną przeprowadzone w możliwie nieformalnym miejscu (np. kawiarnia, dom czy miejsce pracy w przypadku osób posługujących się gwarą) przez wykonawców projektu.

3. Metodologia korpusu

- Rodzaje nagrań:
- odczytanie listy słów
- odczytanie krótkiego tekstu
- przeprowadzenie rozmowy *formalnej*
- przeprowadzenie rozmowy *spontanicznej*

3. Metodologia korpusu

- Zapisy ortograficzne nagrań przy zachowaniu pełnej anonimowości będą umieszczone w elektronicznej bazie naukowej i udostępniane wyłącznie w celach badawczych.

Plan

1. Cele i zastosowania projektu
2. Zespół
3. Metodologia korpusu
4. Zbieżność celów z profilem NPRH
5. Harmonogram

4. Zbieżność celów z profilem NPRH

- Dziedzictwo narodowe (gwara i mowa)
- Bogata bibliografia
- Przegląd wcześniejszej literatury
- Uzasadniona krytyka istniejącego korpusu (wg mnie, to przeważyło)

4. Zbieżność celów z profilem NPRH

- Istnieje już Narodowy Korpus Języka Polskiego (http://nkjp.pl/), jednak analiza założeń i struktury NKJP w pełni uzasadnia potrzebę zbudowania korpusu całkowicie poświęconego współczesnej, polszczyźnie mówionej:
- a) zaledwie 10 % NKJP to język mówiony, resztę stanowi język pisany (Górski 2008, Przepiórkowski et al. 2009, Przepiórkowski et al. 2010, Pęzik 2011). Wśród 10%-wego komponentu mówionego, „jedynie 1% całego korpusu, będą stanowiły transkrypcje spontanicznych „podsłuchanych" rozmów" (Przepiórkowski et al. 2009). 9 % przypada na przemówienia publiczne, wywiady, stenogramy przesłuchań, programy telewizyjne czy audycje radiowe, nie oddające nieformalnego charakteru potocznego języka.

4. Zbieżność celów z profilem NPRH

- b) 1 % nagrań rozmów spontanicznych NKJP reprezentuje różnorodność tematyczną (nie było z góry ustalonych tematów). Wyklucza to możliwość porównania wariantywności fonetycznej czy prowadzenia badań stricte fonetycznych poprzez brak powtarzających się słów, ograniczając zastosowanie korpusu do zbadania ich częstotliwości występowania lub kontekstu semantycznego.

4. Zbieżność celów z profilem NPRH

- c) brak kontroli kontekstu fonetycznego w postaci powtarzających się słów uniemożliwia stworzenie norm referencyjnych dla współczesnej polszczyzny np. uśrednionych wartości formantów samogłosek lub czasu trwania VOT (*Voice Onset Time*). Ponadto, komponent mówiony NKJP nie zawiera listy słów czy tekstu do czytania przez wszystkich nagrywanych, co pozwoliłoby na pełną kontrolę kontekstu np. formanty samogłoski są obniżone przed głoskami /p, b,/ a podwyższone przed /k, g/, stąd potrzeba zachowania tych samych kontekstów dla słów do ich porównania.

4. Zbieżność celów z profilem NPRH

- d) komponent mowy spontanicznej NKJP charakteryzuje się przypadkowym doborem osób nagrywanych pod względem geograficznym: nagrani zostali znajomi oraz członkowie rodzin studentów Uniwersytetu Łódzkiego, którzy mogli pochodzić z różnych miejscowości; część danych, dotyczących miejsca zamieszkania, nie jest kompletna.
- e) w nagraniach przeważają ludzie młodzi, wykształceni, co wpływa na niezrównoważoną reprezentację wybranej grupy społecznej i utrudnia korelację wieku, płci bądź wykształcenia z wymową.
- f) w części spontanicznej korpusu, komponent gwarowy nie jest kontrolowany.
- g) długość nagrań nie jest kontrolowana (od 17 sekund do 2 godzin i 38 minut).
- h) większość nagrań pochodzi z korpusu PELCRA, zebranego w latach 1992-2003 (Pęzik 2005), można zatem kwestionować ich współczesność.

4. Zbieżność celów z profilem NPRH

- Podsumowując, brak spójności tematycznej, geograficznej i reprezentatywności czynników społecznych (wiek, płeć, wykształcenie, standard vs. gwara ponieważ określenie „standardowy” w tym projekcie rozumiany jest w opozycji do gwary) uzasadniają konieczność stworzenia nowego korpusu mówionej polszczyzny.

Plan

1. Cele i zastosowania projektu
2. Zespół
3. Metodologia korpusu
4. Zbieżność celów z profilem NPRH
5. Harmonogram

5. Harmonogram

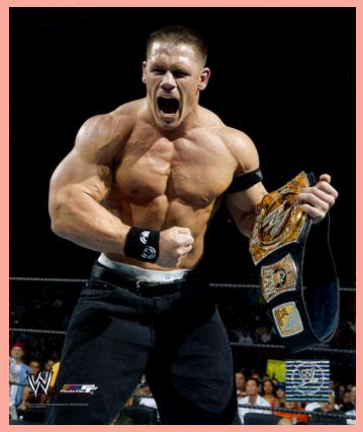
Lp.	Nazwa zadania badawczego	Podmiot realizujący zadanie	Przewidywane koszty (zł)	Planowany okres realizacji (miesiące)
1	2	3	4	5
1	Opracowanie ankiet dotyczących metadanych uczestników projektu	WA, UAM	20552	1
2	Opracowanie danych do nagrań (lista słów, zakres tematyczny rozmów spontanicznych)	WA, UAM	10473	2-3
3	Opracowanie konwencji transkrypcji ortograficznej	WA, UAM	4105	4
4	Przeprowadzenie nagrań pilotażowych	WA, UAM	7954	5
5	Analiza nagrań pilotażowych	WA, UAM	9621	6-7
6	Poszerzenie listy słów oraz opracowanie tekstu czytanego	WA, UAM	10614	8
7	Przeprowadzenie nagrań mowy standardowej – część I	WA, UAM	20679	9-10
8	Transkrypcja danych z nagrań mowy standardowej	WA, UAM	22450	11-14
9	Przeprowadzenie nagrań mowy standardowej – część II	WA, UAM	19089	15-16
10	Delegacja zagraniczna 1	WA, UAM	9868	16
11	Transkrypcja danych z nagrań mowy standardowej	WA, UAM	20526	17-20
12	Przeprowadzenie nagrań mowy gwarowej	WA, UAM	19089	21-22
13	Transkrypcja danych z nagrań mowy gwarowej	WA, UAM	20526	23-27
14	Delegacja zagraniczna 2	WA, UAM	9868	28
15	Udział w konferencjach międzynarodowych	WA, UAM	27411	28
16	Opracowanie konwencji transkrypcji fonetycznej	WA, UAM	16026	29-30
17	Transkrypcja fonetyczna mowy spontanicznej	WA, UAM	38732	31-36
18	Zaprojektowanie i implementacja strona internetowej	WA, UAM	9621	37
19	Edycja oraz udostępnienie plików tekstowych	WA, UAM	6158	38-39
20	Udział w konferencjach międzynarodowych	WA, UAM	27412	40

Ostatni slajd

- Pełen tekst wniosku można znaleźć tu:

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

Zapytania można będzie kierowac tu:

korpuswlkp@wa.amu.edu.pl

Dziękuje za uwagę i good luck!