

Akustyczna baza danych zgromadzona na potrzeby układu detekcji wybranych błędów wymowy w mowie angielskiej Polaków

An acoustic database gathered for the purpose of creating a detector of selected pronunciation errors appearing in English spoken by Poles

Grzegorz Krynicki, Dawid Pietrala,
Katarzyna Dziubalska-Kołaczyk, Mikołaj Wypych

Interdyscyplinarne Centrum Przetwarzania Mowy i Języka, Polska
krynicki@ifa.amu.edu.pl
daphon@ifa.amu.edu.pl
dkasia@ifa.amu.edu.pl
wypych@polfonetika.com

STRESZCZENIE

Niniejszy artykuł przedstawia założenia i wstępne wyniki projektu, którego celem jest stworzenie układu detekcji wybranych błędów wymowy w mowie angielskiej Polaków. Ważnym krokiem do osiągnięcia tego celu a jednocześnie cennym zasobem jest prezentowana tu obszerna akustyczna baza danych składająca się z nagrań Polaków mówiących po angielsku i po polsku. Nagrania te powstały w oparciu o równoważone fonetycznie zdania wybrane z korpusów języka angielskiego i polskiego oraz z listy zdań angielskich wykorzystanych w projekcie TIMIT [1]. Angielska część bazy została anotowana ze względu na wybrane błędy wymowy. Artykuł przedstawia analizę ilościową wybranych błędów występujących w zgromadzonej bazie.

ABSTRACT

The following article presents the assumptions and preliminary results of a project whose goal is to create a detection system for pronunciation errors made by Poles speaking English. An important step towards reaching this goal is an extensive acoustic database consisting of recordings of Poles speaking English and Polish. The database was recorded on the basis of phonetically balanced sentences extracted from English and Polish corpora and on the basis of a list of English sentences used in the TIMIT project. The English part of the database was annotated with respect to selected pronunciation errors typical to Polish learners of English. The article presents the quantitative analysis of selected errors appearing in the database.

1. Informacje ogólne

Celem naukowym projektu¹ prezentowanego w niniejszym artykule było stworzenie automatycznego układu detekcji błędów fonetycznych w mowie angielskiej osób, dla

¹Grant naukowo-badawczy Ministerstwa Nauki i Szkolnictwa Wyższego numer N519 016 31/2965.

których językiem ojczystym jest język polski. Dla osiągnięcia tego celu konieczne było zebranie korpusu mowy angielskiej Polaków oraz stworzenie technologii detekcji błędów wymowy w oparciu o układ rozpoznawania mowy angielskiej.

Stworzony w wyniku projektu prototyp detektora błędów wymowy posłużył z jednej strony do stworzenia oprogramowania dydaktycznego nowej generacji, pozwalającego na samodzielne udoskonalanie uczenia się wymowy w języku obcym jak też wspomagającego nauczyciela w ocenie postępów ucznia. Z drugiej strony, stanowi on przyczynek do udoskonalania systemów rozpoznawania mowy akcentowanej. Zebrany i opracowany korpus mowy akcentowanej pozwoli z kolei na wykonanie wielorakich analiz fonetycznych.

Wykonawcy projektu dysponowali zapleczem i doświadczeniem zdobytymi we wcześniejszych projektach oraz wiedzą zdobytą dzięki współpracy z Center for Spoken Language Research Uniwersytetu w Colorado. Zasadniczym ułatwieniem dla wykonania projektu było laboratorium nagraniowe przy Interdyscyplinarnym Centrum Przetwarzania Mowy i Języka UAM wyposażone w komorę bezchłową oraz odpowiedni sprzęt audio i komputerowy.

2. Ekspertyza i inwentarz błędów wymowy

Pierwszym krokiem w projekcie skupionym na detekcji błędów wymowy było opracowanie inwentarza typowych i istotnych błędów fonetycznych popełnianych przez Polaków uczących się języka angielskiego. Stworzenie inwentarza poprzedzono ekspertyzą znanych oraz dostępnych w opracowaniach błędów, opartą na porównaniu fonologii języków angielskiego i polskiego w trzech aspektach: uniwersalnym, typologicznym i systemowym (m.in. [9, 2, 3, 4]). Wynikiem porównania są przewidywania wystąpienia danych typów błędów.

Z uniwersalnego punktu widzenia, język polski, w przeciwieństwie do angielskiego, jest nienacechowany jeśli chodzi o proces ubezdźwięcznienia wygłosowych obstruentów jak też pod względem zgodności dźwięczności w zbitkach spółgłoskowych. W konsekwencji, przewidywalne jest stosowanie procesu ubezdźwięcznienia oraz upodobnienia dźwięczności przez Polaków w języku angielskim (interferencja procesów uniwersalnych, wzmocniona jeszcze interferencją z języka ojczystego, por. niżej cechy systemowe).

Typologicznie, organizacja rytmiczna języka polskiego nie jest oparta na akcencie (w odróżnieniu od angielskiego). W związku z tym, samogłoski utrzymują swą jakość, akcent wyrazowy jest stały, zbitki spółgłoskowe są rozbudowane, iloczas nie odgrywa roli, a inwentarz samogłosek i spółgłosek jest liczbowo znacznie bliższy średniej równowadze (tzn. 6 samogłosek i ok. 20 spółgłosek). Oczekuje się zatem trudności w redukowaniu nieakcentowanych samogłosek oraz pozycjonowaniu akcentu w wyrazach w angielskiej wymowie Polaków.

Inwentarze spółgłoskowe obu języków są całkowicie różne. W inwentarzach spółgłosek występują różnice zarówno systemowe jak i dystrybucyjne, np. polski posiada spółgłoski zębowe laminalne, a angielski – apikalne; dystrybucja welarnej nosówki ograniczona jest w języku polskim do kontekstu przed homorganicznymi plozywami. Typowym błędem Polaka jest zastępowanie zębowej apikalnej (tzw. <th>) przez polskie zębowe lub labialno-zębowe obstruenty.

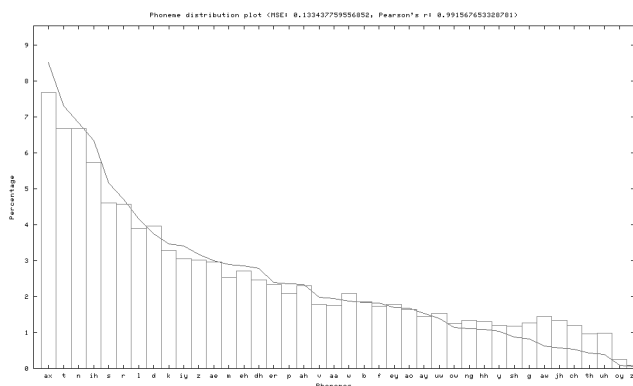
Podsumowując, niektóre błędy wymowy popełniane przez użytkowników języka polskiego w języku angielskim będą miały podłoże w języku ojczystym lub wynikać będą ze stosowania się procesów natury typologicznej lub uniwersalnej.

3. Korpusy tekstów

W celu stworzenia podstawy dla wygenerowania fonetycznie zrównoważonych zdań polskich i angielskich zebrano materiał korpusowy dla języka angielskiego i polskiego. Materiał zebrano automatycznie polegając na źródłach dostępnych w Internecie. Otrzymano w ten sposób korpus angielski wielkości ponad 52 milionów tokenów i polski wielkości 24 milionów tokenów. Z nich wybrano kolejno 162 664 i 39 194 zdania niezawierające nazw własnych, cyfr ani wulgaryzmów. Te zbiory stanowiły podstawę dla dalszego przetwarzania.

4. Zbiory zdań do nagrywania

Dla wiarygodnej estymacji własności akustycznych wszystkich fonemów w korpusie akustycznym wymagana jest minimalna liczba wystąpień każdego z nich. Zwiększenie minimalnej oczekiwanej liczby wystąpień dowolnego fonemu w korpusie można uzyskać bez zwiększania rozmiaru korpusu. W tym celu konieczna jest selekcja treści korpusu w taki sposób, by różnice w częstościach wystąpień poszczególnych fonemów zostały zredukowane. Zbiór zdań, w którym rozkład fonemów jest bardziej równomierny niż w rozkładzie Yule'a nazywać będziemy zbiorem *zrównoważonym fonematycznie*. Rozkład Yule'a jest dobrą aproksymacją rozkładu częstości wystąpień fonemów w tekście bieżącym [5]. Zgodnie z rozkładem Yula prawdopodobieństwo wystąpienia najrzadszych fonemów jest o rzędy wielkości mniejsze od prawdopodobieństwa wystąpienia fonemów najczęstszych.



Rycina 1. Rozkład częstości fonemów w korpusie tekstowym (lamana) oraz w zbiorze zrównoważonym fonetycznie (słupki), alfabet DARPA, język amerykański.

W przedstawionych badaniach wykorzystano 1) zbiory zrównoważone fonematycznie zawierające zdania wybrane z korpusu tekstów oraz 2) zbiory zrównoważone fonematycznie złożone ze zdań ułożonych przez eksperta. Pierwszą grupę zdań reprezentują zbiory enGB-pelt, enUS-pelt oraz pl-pelt (wspólnie oznaczane dalej jako *-pelt). Drugą grupę zdań reprezentuje 4200 zdań ze zbioru TIMIT [1].

Zbiory *-pelt zostały utworzone automatycznie ze zdań występujących w korpusach tekstowych przedstawionych w punkcie 3. Z wejściowych korpusów tekstowych losowo wybrano po 38 000 dla każdego z języków (dialektów). Powstałe w ten sposób korpusy, poddano transkrypcji fonematycznej (odpowiednio w języku angielskim brytyjskim, angielskim amerykańskim oraz polskim za pomocą narzędzi przedstawionych [9] oraz [2]). Każdemu z przetranskrybowanych zdań przypisano zbiory fonemów i difonów oraz zastosowano autorski wariant zachłannego algorytmu selekcji zdań przedstawionego w [3]. W wyniku przedstawionej procedury powstały zbiory enGB-pelt, enUS-pelt oraz pl-pelt zawierające po 6400 zdań każdy. Wykres 1 przedstawia naturalny oraz zrównoważony rozkład częstości występowania fonemów.

5. Sesje nagraniowe

Sesje nagraniowe korpusu odbywały się od 15.01.2007 do 01.04.2007 w Instytucie Filologii Angielskiej na Wydziale Neofilologii Uniwersytetu im. Adama Mickiewicza w Poznaniu. W ramach tych prac zebrano korpus nagrań 74 osób, z których każda zarejestrowała w sumie ok. 120 zdań:

- 40 naturalnych zrównoważonych fonetycznie zdań angielskich,
- 40 sztucznie wygenerowanych bogatych fonetycznie zdań angielskich (TIMIT),
- 40 naturalnych zrównoważonych fonetycznie zdań polskich.

Parlatorami, którzy zostali zarejestrowani w trakcie nagrań byli studenci Filologii Angielskiej w Poznaniu. Każda z osób przed rozpoczęciem nagrań poproszona została o podanie informacji dotyczących jej wieku, języków, którymi posługuje się w stopniu co najmniej średnio-zaawansowanym, regionu pochodzenia oraz zobowiązana była do zadeklarowania modelu wymowy, który stara się naśladować.

6. Anotacja błędów wymowy

W ramach prac nad projektem stworzono nową metodę oznaczania błędów wymowy w angielskim materiale nagraniowym. Proces anotowania materiału dźwiękowego odbywał się dwuetapowo: 1) anotowanie w trakcie sesji nagraniowych, oraz 2) anotowanie w trakcie sesji anotacyjnych. Anotacje powstające w pierwszym etapie odnosiły się do parlatorów oraz do zdań rozpatrywanych w całości. W drugim etapie powstały anotacje dotyczące poszczególnych segmentów fonetycznych w zdaniach. Dla redukcji kosztów anotacji etap drugi przeprowadzono częściowo automatycznie.

Nagraną i podzieloną na zdania akustyczną bazę danych w całości przekazano jednocześnie każdemu z członków zespołu anotującego. Pliki w wewnętrznie ustalonym formacie *tags* zawierające transkrypcję zdania i charakterystykę mówcy je wypowiadającego pogrupowano w 100 zrandomizowanych zbiorów zawierających średnio po 65 plików

tags. Zbiory te były sukcesywnie przekazywane anotatorom przez koordynatora, który jednocześnie dokonywał niezbędnych korekt i udzielał informacji zwrotnej na temat jakości dokonywanej anotacji. Korekta zmierzała do osiągnięcia następujących celów:

- Usunięcie błędów anotacji względem ustalonego formalizmu.
- Uzupełnienie niedostrzeżonych przez anotatora błędów automatycznej transkrypcji oraz błędów wymowy przy pomocy środków opisanych w ustalonym formalizmie.
- Modyfikacje ustalonego formalizmu anotacji oraz wyjściowego zakresu uwzględnianych błędów wymowy w celu pełniejszego zobrazowania błędów wymowy.
- Ujednolicenie poziomu szczegółowości anotacji pomiędzy anotatorami.

7. Charakterystyka ilościowa korpusu oraz wstępne wyniki analizy ilościowej anotacji błędów wymowy

Krótką charakterystykę zebranego korpusu przedstawiają poniższe dane:

- zdania nagrane: 6123, w tym:
 - zdania odrzucone przed etapem anotacji (ze względu na powtórzenia, przeinaczenia, przejęzyczenie itd.): 646,
 - zdania zaanotowane: 5477;
- segmenty zawarte w zaanotowanych zdaniach: 276 396:
 - w tym błędne (wstawienie, usunięcie, zamiana): 47 060.

Analiza stu błędów najczęściej występujących w korpusie PELT wyłoniła następujące kategorie problemów w kolejności od sprawiających uczniom najwięcej trudności (jeden błąd mógł należeć do więcej niż jednej kategorii):

Tabela 1. Częstotliwość występowania różnych kategorii błędów w korpusie PELT

| Kategoria błędu | Ilość wystąpień | Udział wzgl. błędów w ww. kategoriach | Udział wzgl. wszystkich błędów |
|--------------------------------|-----------------|---------------------------------------|--------------------------------|
| Jakość i długość samogłosek | 15 431 | 37,9% | 32,8% |
| Inne problemy ze spółgłoskami | 7980 | 19,6% | 17,0% |
| Formy słabe samogłosek i schwa | 5305 | 13,0% | 11,3% |
| Ubezdźwięcznienia spółgłosek | 4560 | 11,2% | 9,7% |
| Epentezy | 3617 | 8,9% | 7,7% |
| Elizje | 2110 | 5,2% | 4,5% |
| Problemy z /θ/ i /ð/ | 900 | 2,2% | 1,9% |
| Problemy z /r/ | 806 | 2,0% | 1,7% |
| Sumy | 40 709 | 100,0% | 86,5% |

Powyższa tabela obejmuje 37 938 unikalnych błędów co stanowi 80,6% wszystkich błędów. Częstotliwość objętych nią kategorii potwierdza wiodącą trudność zachowania poprawnej jakości samogłosek w wymowie angielskiej Polaków. Badanie zwraca także uwagę na ważność problemu miejsca i sposobu artykulacji spółgłosek angielskich, ubezdźwięcznienia wygłosowych obstruentów oraz problem zgodności

dźwięczności w zbitkach spółgłoskowych. Ważnym problemem jest także redukowanie nieakcentowanych samogłosek, często związanym z problemem niepoprawnego pozycjonowania akcentu wyrazowego.

Wyniki badania wraz z korpusem PELT, obok bezpośredniego zastosowania w trenowaniu mechanizmu automatycznej detekcji błędów wymowy, może posłużyć w dydaktyce fonetyki skierowanej do Polaków uczących się języka angielskiego.

BIBLIOGRAFIA

- [1] W. M. Fisher, G. R. Doddington, K. M. Goudie-Marshall. 1986. "The DARPA Speech Recognition Research Database: Specifications and Status," *Proceedings of DARPA Workshop on Speech Recognition*, pp. 93–99.
- [2] A. Reszkiewicz. 1981. *Correct your English pronunciation*, PWN: Warszawa.
- [3] S. Bałutowa. 1990. *Wymowa angielska dla wszystkich*, Wiedza Powszechna: Warszawa.
- [4] W. Sobkowiak. 2001. *English phonetics for Poles*. (2nd ed.), Pp. 309. Wydawnictwo Poznańskie: Poznań.
- [5] Y. Tambovtsev, C. Martindale, "Phoneme Frequencies Follow a Yule Distribution". 2007. *SKASE Journal of Theoretical Linguistics*, vol. 4, no. 2.
- [6] A. W. Black, P. Taylor, R. Caley. 2002. *The Festival Speech Synthesis System/System documentation*.
- [7] M. Wypych, M. Szczyszek, E. Szalkowska. 2007. *Dokumentacja systemu transkrypcji fonemetycznej "polf"*, Poznań.
- [8] J. P. H. van Santen, A. L. Buchsbaum, "Methods for Optimal Text Selection". 1997. *Proceedings of Eurospeech '97*, Rhodes, Greece.
- [9] W. Jassem. 1973. *Podręcznik wymowy angielskiej*, PWN: Warszawa.