

BUDOWA BAZY DIALOGÓW TELEFONICZNEJ MOWY POLSKIEJ W RAMACH PROJEKTU EC LUNA

Krzysztof Marasek, Ryszard Gubrynowicz
PJWSTK, Warszawa



Polsko-Japońska Wyższa
Szkoła Technik Komputerowych



Ogólne zadania realizowane w ramach projektu LUNA

Cele naukowe

- Metody reprezentacji znaczenia
- Proces interpretacji
- Niezawodność, metody oceny niezależne od języka
- Metody uczenia, ujednolicony proces anotacji

Cele technologiczne

- Rozwój nowych metod, algorytmów i narzędzi niezbędnych do szybkiego rozwoju niezawodnych elementów składowych systemów wielojęzycznych rozumienia mowy, stosowanych w zautomatyzowanych telefonicznych serwisach informacyjnych

Spoken **L**anguage **UN**derstanding in Multilingu**A**I Communication Systems

- Partnerzy

Przemysłowi	Akademiccy
	University of Avignon
Loquendo (coordinator)	RWTH Aachen
France Telecom R&D	University of Trento
Consorzio per il Sistema Informativo piemontese – CSI-Piemonte	Instytut Podstaw Informatyki PAN
	Polsko-Japońska Wyższa Szkoła Technik Komputerowych

Partnerzy industrialni w LUNA



System 3000

System 1014

Opinion analysis



Dialogue platform

language and conceptual models in new
languages



Zainteresowani aplikacjami dla języka polskiego

Korpusy dialogów w LUNA

French

- FT 3000 Voice Agency corpus
- FT 1014 Customer Care corpus data
- FT Stock Exchange corpus data
- UAPV media corpus

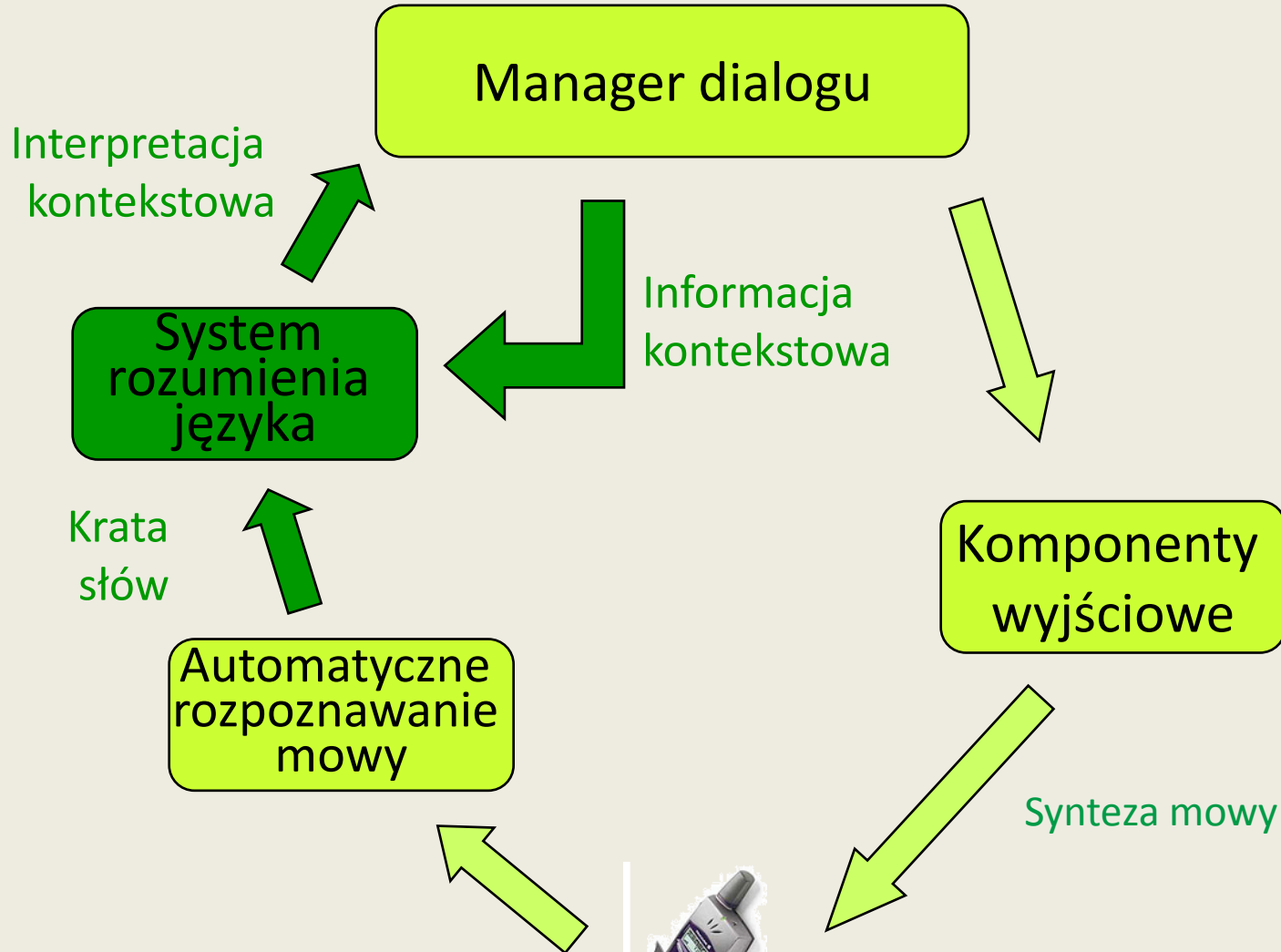
Italian

- CSI's help desk corpus data

Polish

- PJIIT-PAS's corpus data transportation help desk

Wzajemna komunikacja człowiek – maszyna za pomocą mowy



Zadania PJWSTK w projekcie LUNA

Nagranie i anotacja 500 dialogów człowiek - człowiek

- Transkrypcja i anotacja na poziomie akustycznym (anotacja semantyczna IPI)

Nagranie i anotacja 500 dialogów człowiek - maszyna

- Realizacja w środowisku Wizard-of-Oz
- Transkrypcja i anotacja

Przygotowanie demo

- PJIIT/IPI z pozostałymi partnerami, przede wszystkim Loquendo
- Zastosowanie modułów opracowanych w ramach projektu

Cele akwizycji baz nagrań

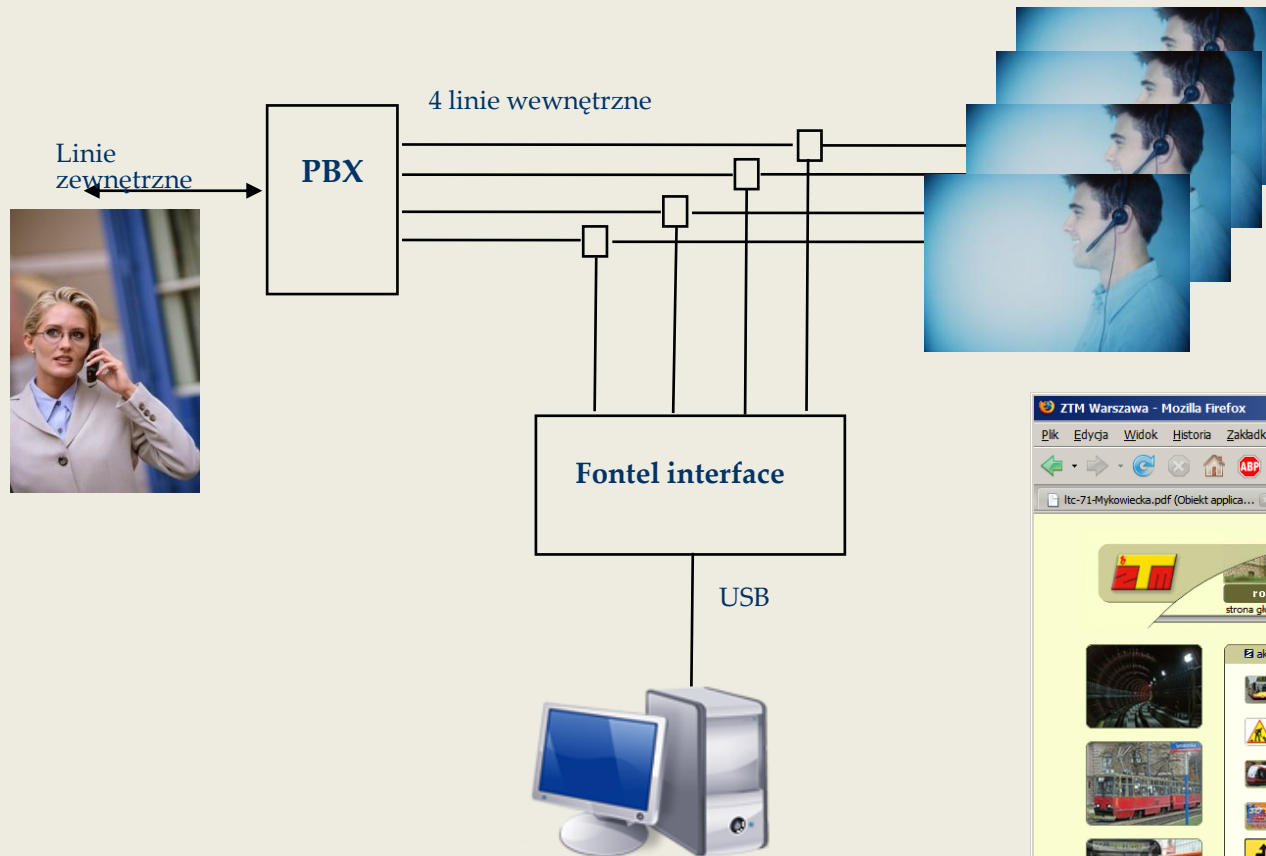
Zebranie danych do
stworzenia
wiarygodnych i
realistycznych modeli
pojęciowych dialogu

- Staranna anotacja lingwistyczna
- Wyróżnienie konceptów, części dialogu, itp.

Zebranie danych do
uczenia systemów
ASR

- Staranna anotacja na poziomie akustycznym (hałasy, błędy wymowy, itp.)
- Podział na wypowiedzi poszczególnych mówców

Schemat systemu nagrywania rozmów



próbkowanie 16 bitowe, 16 kHz, liniowe bez kompresji



Struktura dialogów w bazie danych

- W sumie 15000 rozmów w infolinii ZTM zostało zarejestrowanych (b. duża różnorodność rodzajów poszukiwanych informacji)
- Wybrano 500 dialogów, które podzielono na 5 grup wg typu

Informacji dotyczącej trasy komunikacyjnej między dwoma określonymi punktami w mieście: np. „jak dojechać z do ...”

- rozkładu jazdy dla określonego przystanku i linii komunikacyjnej oraz dotyczącej czasu przejazdu: np. „o której odjeżdża najbliższy autobus z przystanku” „jak długo będę jechał do”

- trasy określonej linii komunikacyjnej: np. „czy tramwajem dojadę do”

- przystanków (najbliżej położony określonego miejsca w mieście, przystanków danej linii, przystanki „przesiadkowe” itp.): np. „na jakim przystanku mam wysiąść ... by...”

- zniżek i przejazdu dla określonych grup pasażerów komunikacji miejskiej (dzieci, młodzież, emeryci, inwalidzi itp.): np. „czy osoba 72-letnia ma prawo do”

Podział korpusu wg typu informacji

Rodzaj udzielanej informacji	Liczba nagrań	kobiety	mężczyźni
trasy komunikacyjne	93	53	40
trasy dojazdu	140	78	62
rozkłady jazdy	112	60	52
przystanki	55	24	31
Zniżki, bezpłatne	101	61	40

Dialogi są klasyfikowane wg dominującej w rozmowie poszukiwanej informacji, lub wymienionej na pierwszym miejscu.

Poziomy anotacji

Audio file	Folder	File	Content
name.wav	name	name.trs	Transcription
		name_words.xml	Words
		name_turns.xml	Dialogue turns
		name_pos.xml	Part of Speech
		name_chunks.xml	Chunks
		name_attvalue.xml	Domain attribute-value pairs
		name_frames.xml	Predicate structure
		name_coref.xml	Coreference
		name_dialacts.xml	Dialogue acts

Narzędzie transkrypcji - Transcriber

Transcriber 1.5.1

File Edit Signal Segmentation Options Help

report - male user, mid recording quality

(no speaker)

operator

[pron=*-]dzień[-pron=] dobry xxx słucham

user

[noise-] [pron=*-]dzień[-pron=] dobry panie xxx [lex=FIL] moje pytanie jest takie [lex=FIL] o autobus sto dwadzieścia siedem chciałbym jechać w kierunku Dworca Zachodniego [lex=~]g[-lex=~]dzieś w pobliżu Starego Miasta chciałbym się przesiąść to jak tam [lex=?] [babble-] który to [pron=**] [-babble] [-noise]

operator

[babble-] Plac [-babble] Bankowy [noise-] na przykład [-noise]

user

a wcześniej [lex=?] [sil]

operator

[noise-] [lex=FIL] [sil] wcześniej zaraz powiem [sil] [-noise]

user

[noise-] [lex=~]g[-lex=~]zieś Bonifraterska czy [babble-] jeździ tamtędy [lex=?] [-babble] [-noise]

operator

przystanek Świętojerska na ulicy [noise-] Świętojerskiej albo [lex=FIL] na albo przystanek Muranowska na Bonifraterskiej [sil] [-noise]

user

aha [sil] [noise] [lex=FIL] rozumiem czyli [lex=FIL] Świętojerska bądź [pron=**] potem on jedzie już Andersa tak [lex=?]

operator

no z [lex=FIL] z Bonifraterskiej skręca w prawo w Świętojerską a ze Świętojerskiej w [lex=FIL] w lewo w Andersa

user

aha [sil] [noise-] [lex=FIL] dobrze [sil] [-noise] to w takim razie jakiś rozkład gdyby pan znalaz[lex=~]i[-lex=~] taki po osiemnastej no powiedzmy ze Świętojerskiej [sil] [noise] [babble-] w stronę dworca Zachod[lex=~]niego[-lex=~]

operator

2_2007-04-13_10_47_30

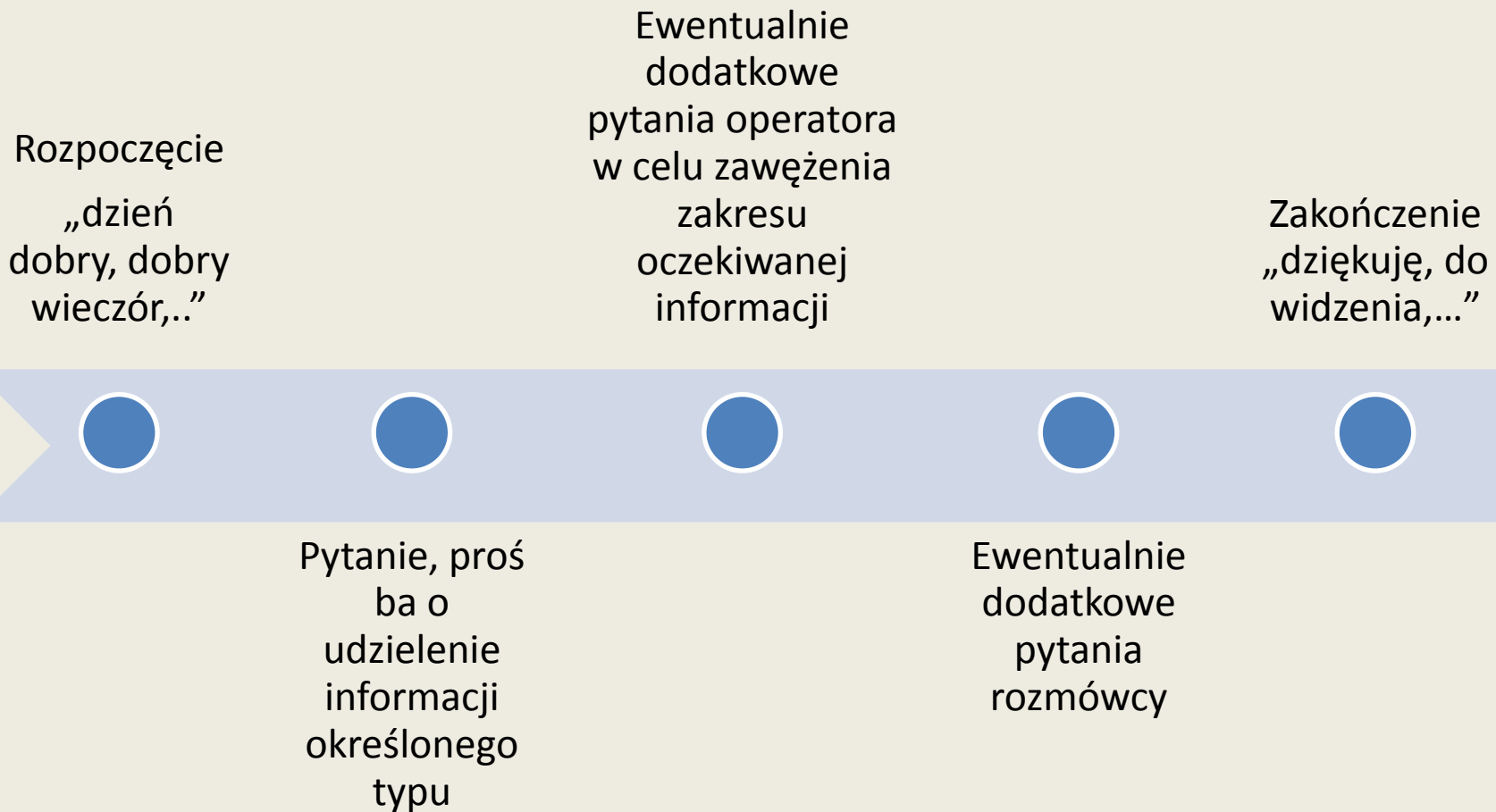


male user, mid recording quality

(no.	operator	user	operator	user	operator	user
	[pron=*-] słucham	[noise-] [pron=*-]dzień[-pron=] dobry panie xxx [lex=FIL] moje pytanie jest takie [lex=FIL] o autobus sto...	[babble-] ... [-noise]	a ... [sil]	[noise-] [lex=FIL] [sil] wcześniej zaraz powiem [sil] [-noise]	[noise-] ... [-babble] [-noise]

0 5 10 15 20 25 30

Struktura dialogów



Zniekształcenia w strukturze dialogów

Brak powitania

Pytania nie związane z typem udzielanej informacji :

„ czy mógłby mnie pan poratować ...”, „co to jest za bezmózgowiec, co zmienił trasę autobusu 105 ...”

Pretensje dotyczące zachowań kierowców, reklamacje

Brak zakończenia

..i wiele innych

Przykład dialogu (trasa autobusowa)

DIALOG 2_2007-04-13_10_47_30

Operator [2_2007-04-13_10_47_30.trs](#)

Dzień dobry xxx słucham

1) rozmówca (pytanie złożone – trasa autobusu i przystanek)

..... chodzi o autobus 127 chciałbym jechać w kierunku dworca Zachodniego gdzieś w pobliżu Starego Miasta chciałbym się przesiąść, (Plac Bankowy)..... lub wcześniej

operator

... wcześniej gdzie?.....

2) rozmówca (pytanie o trasę autobusu)

... gdzieś Bonifraterska, czy jeździ tamtędy ?

operator

przystanek Świętojska na Świętojskiej albo Muranowska na Bonifraterskiej

3) rozmówca (pytanie o trasę autobusu)

aha rozumiem Świętojska bądź [?] potem jedzie już Andersa tak?

operator

No z Bonifraterskiej skręca w Świętojską a ze Świętojskiej w lewo w Andersa

4) rozmówca (pytanie o rozkład – przystanek Świętojska)

... jakiś rozkład ... taki po osiemnastej ze Świętojskiej w stronę dworca Zachodniego

operator

Osiemnasta jeden ,...szesnaście

5) rozmówca (pytanie o połączenie między wybranymi punktami)

... jak jechać z dworca Zachodniego na Stare Miasto

operator

... bezpośrednio nie ma Na plac Zamkowy? Opis trasy komunikacyjnej

6) rozmówca (pytanie o przystanek)

... aha, na placu Zawiszy jest wspólny przystanek?

12 listopad 2007

K. Marasek, R. Gubrynowicz, PJWSTK, Warszawa

Problem zniekształceń mowy

- **Zakłócenia:** większość rozmów jest prowadzona w niekorzystnych warunkach akustycznych (rozmówcy znajdujących się na ulicach, w środkach komunikacji miejskiej), zniekształcających sygnał mowy.
- **Jednokanałowa komunikacja telefoniczna:** trudności w rozdzieleniu „nakładających się” rozmówców (jednocześnie mówiących)
- **Niska jakość łączy telekomunikacyjnych:** znaczna część rozmów prowadzona jest z komórek, często o niskiej jakości mikrofonów. Typowe zjawiska – chwilowe zanikanie sygnału, lub drastyczne obniżanie się poziomu.
- **Zniekształcenia struktury dźwiękowej wypowiedzi** - nieciągłości, zawahania, powtórzenia, błędy wymowy, agramatyzmy, [żargon](#) , wtrącenia (przerw, paralingwistycznych dźwięków), obserwowane zwłaszcza w mowie spontanicznej. .
- **Silne emocje** (często o negatywnym charakterze) – mogą dodatkowo zniekształcać strukturę dźwiękową i lingwistyczną wypowiedzi.

Transkrypcja i anotacja wypowiedzi I

- W transkrypcji nie uwzględnia się znaków interpukcji, z wyjątkiem pytajnika stawianego w miejscach wyraźnego podniesienia wysokości głosu na końcu zdania
- Duże litery na początku wyrazów stosowane są zgodnie ze standardami obowiązującymi w danym języku, z wyjątkiem początków zdań (małe litery) np. Galeria Mokotów, Bitwy Warszawskiej.
- Liczby są reprezentowane w postaci słownej (łącznie z fleksją)
- Wyrazy nie rozpoznane oznaczone są jako [pron=**]
- Błędnie wypowiedziany wyraz jest zapisany w formie poprawnej z anotacją błędnej wymowy, np. [pron=*-] placu [-pron=*] Narutowicza.
- Przy niepełnym wypowiedzeniu wyrazu („obcięcie”, redukcje na początku, w środku, czy końcu), niewymówiona jego część (głoska, sylaba, fleksja itp.) anotowana jest za pomocą [lex=~-] i [-lex=~]

Transkrypcja i anotacja wypowiedzi II

- Obce wyrazy, akronimy są transkrybowane zgodnie z ich oryginalną ortografią i oznaczane etykietą [lang=] i nazwą języka np. [lang=English-] McDonald [-lang=English], lang=[English-] CD [-lang=English].
- Literowane wyrazy, akronimy są transkrybowane dużymi literami z anotacją - [pron=SPELLED], np. [pron=SPELLED-] AWF [-pron=SPELLED].
- Akronimy wymówione w formie wyrazowej (w sposób ciągły) są zapisywane jak wyżej bez anotacji, np. PEKAES.
- Tag lex=FIL jest stosowany do oznaczania elementów paralingwistycznych (dźwięki nie mające znaczenia, takie jak hmm, eeh itp.) oraz zakłócenia generowane przez rozmówcę - odechy, śmiech, kaszel itp.

Anotacja zakłóceń

Anotacja ta służy przede wszystkim do wskazania, że dany segment sygnału mowy może być trudny do rozpoznania dla automatycznego systemu rozpoznawania mowy.

Tag [noise] jest wprowadzany na podstawie subiektywnej oceny on SNR dla danego segmentu sygnału mowy.

Jest opozycja [noise] -[int] – na podstawie czasu trwania zakłócenia; kryterium długi/krótki jest również oceną subiektywną szereg powtórzeń [int] -> [noise]

[noise] (lub [babbling] – mowa w tle) w odcinkach bez sygnału mowy rozmówcy -> [sil] [noise] or [noise] [sil] gdy ich poziom jest stosunkowo wysoki w porównaniu do poziomowi sygnału mowy

Wybrane problemy transkrypcji i anotacji

- Generalna zasada – transkrypcja zarejestrowanych dialogów ma odtwarzać dokładnie to, co zostało powiedziane, bez korygowania błędów rozmówców (błędy wymowy są anotowane, agramatyzmy nie)
- Transkrypcja akronimów – gdy można je wymówić w podobny sposób jak pojedynczy wyraz, to zawsze jest podatny na infleksję. Problem jak wykonywać transkrypcję infleksji.

Przykład akronimu: ZUS -> /zus/ -> transkrypcja ZUS

Wersja z infleksją, np. w ZUS-ie /v zus'e/ -> ZUS-ie (ZUSie ??)

Wersja literowana -> /zet u es/ or /zet u es'e/

W kodzie XML:

[pron=SPELLED-]ZUS[-pron=SPELLED] or [pron=SPELLED-]ZUS[-pron=SPELLED]ie

Ale są trudniejsze przypadki:

Np. skutek infleksji może ulec zmianie jedna z liter akronimu

WAT /vat/ -> (*Wojskowa Akademia Techniczna*)

Dopuszczalne infleksje: WATu /vatu/, /vats'e/ (spółgłoska /t/ -> ts'/)

Mieszanie wymowy wyrazów obcych

- Obce wyrazy są często wymawiane częściowo lub całkowicie zgodnie z Polskimi regułami wymowy (np. rdzeń wymówiony zgodnie z wymową oryginalną + polską końcówką fleksyjną). W pierwszym przypadku wyraz jest anotowany jako błędnie wymówiony, w drugim – nie.

Przykład: plac Wilsona – polska wymowa-> /plats vilsona/,

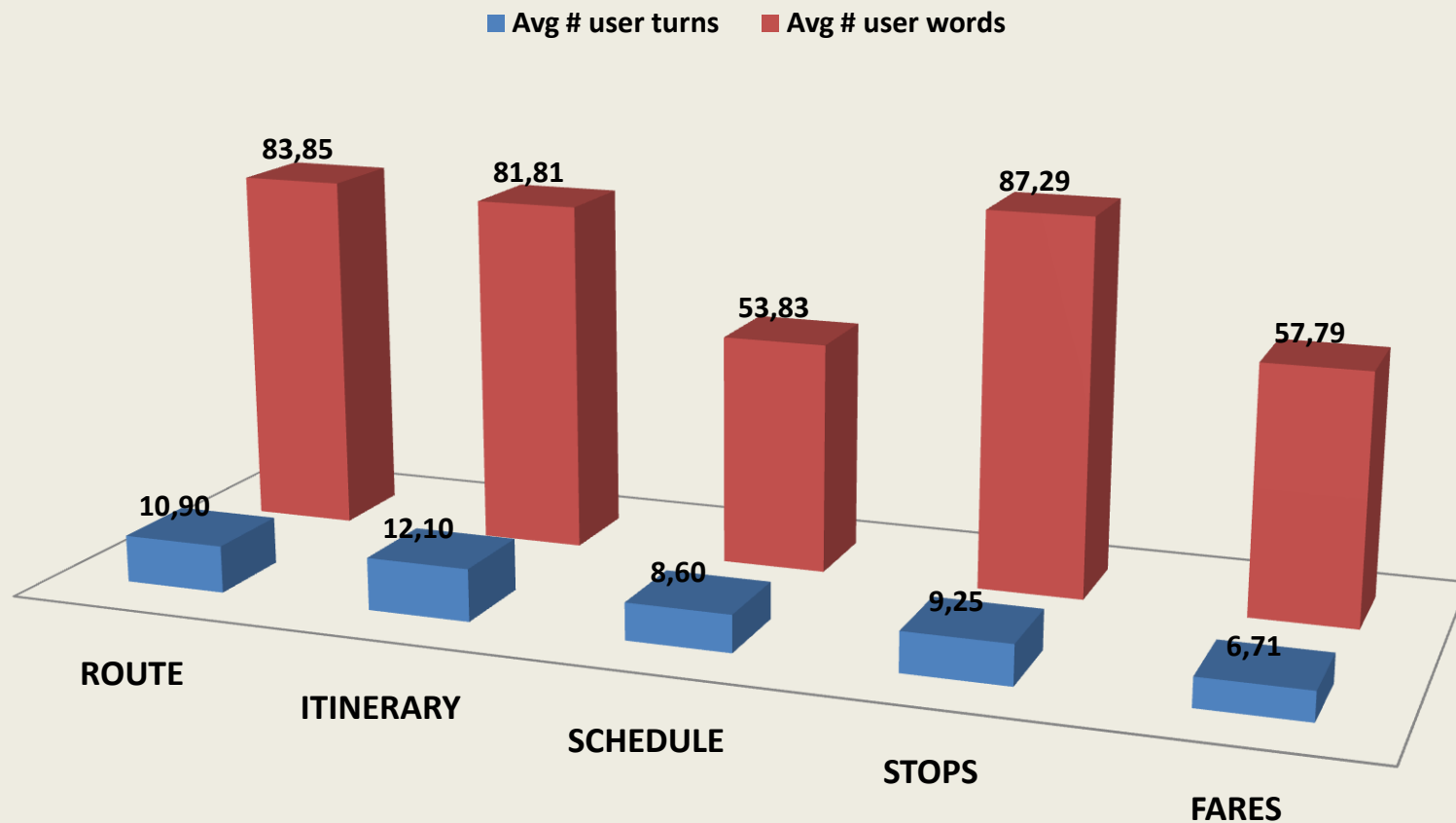
Variant niepoprawny -> /plats **w**ilsona/ (pierwsza sylaba artykułowana zgodnie z wymową angielską)

Dane statystyczne korpusu dialogów

Liczba dialogów	Łączny czas wypowiedzi	Liczba wypowiedzi	Łączna liczba wyrazów	Rozmiar słownika
500	670 min	12788	78082	7768

Statystyka (c.d.)

Średnia liczba wypowiedzi i wyrazów przypadająca na jednego rozmówcę



Obecny stan tworzenia korpusu dialogów

- Zbiór dialogów człowiek – człowiek jest już skompletowany
- 500 dialogów z transkrypcją i anotacją na poziomie akustycznym jest na serwerze PJWSTK
- Stosunkowo niska jakość nagrań, jednak zgodna z rzeczywistymi warunkami panującymi w polskich sieciach telekomunikacyjnych
- Bogactwo tematów, różnorodność rozmówców
- Problemy z transkrypcją wyrazów obcego pochodzenia
- Należy spodziewać się trudności na etapie analizy semantycznej spowodowanych niekorzystnymi cechami mowy spontanicznej (zwłaszcza agramatyzmy)
- Przygotowania do tworzenia bazy dialogów człowiek – maszyna

Podziękowanie dla ZTM Warszawa i pracowników infolinii 94-84 za współpracę