

Analiza dialogu na potrzeby tworzenia systemów dialogowych

Jolanta Bachan

Dlaczego analiza dialogu?

- badania są potrzebne? Obecne systemy opierają się głównie na intuicji inżynierów mowy, nie na systematycznej analizie lingwistycznej
- badania są ważne? Modele automatów rzeczywistych dialogów są nową metodą, szczególnie dla języka polskiego, i są potrzebne do modelowania odpowiednich scenariuszy
- badania są trudne? Niezbyt wiele prac zostało przeprowadzonych na ten temat, przeniesienie opisów korpusowych na modele komputerowe, a następnie na system operacyjny jest bardzo skomplikowane

Projektowanie systemu dialogowego

- Komunikacja człowiek-komputer:
 - specyfikacje wymagają
 - lingwistycznego badania korpusowego
 - lingwistycznego modelowania komputerowego
- System dialogowy powinien być:
 - dobrze umotywowane przez
 - teorię dialogu
 - analizę rzeczywistych dialogów
 - testowany w rzeczywistych warunkach

Akty dialogowe

- Przeczytaj:
http://en.wikipedia.org/wiki/Dialog_act
- *Dialogue acts are the smallest functional units of dialogues, and are utterances corresponding to **speech acts** such as ‘greetings’, ‘request’, ‘suggestion’, ‘accept’, ‘confirm’, ‘reject’, ‘thank’, ‘feedback’”(Gibbon et al. 2000:6).*
- DIT++ Taxonomy of Dialogue Acts (Bunt 2010)
<http://dit.uvt.nl/>
- https://pl.wikipedia.org/wiki/Akt_mowy

Mój własny system dialogowy

Główna teza

*Dopasowanie reprezentacji semantycznych
jest niezbędne
do pomyślnej komunikacji w dialogu.*

Scenariusz

- Sytuacja stresowa nagłego wypadku
- Sytuacja publiczna
- Zadanie z mapą
- Dopasowanie reprezentacji semantycznych map

Strategia badań

- Teoretyczna:
 - Selektywne wyszukiwanie prac na temat dialogu
 - Skupienie się na dopasowaniu interlokutorów
- Empiryczna:
 - Korpusowe badania lingwistyczne dialogu
 - do ekstrakcji modeli formalnych
 - jako źródło danych do rozwoju systemu
- Komputerowe:
 - Weryfikacja modeli ewaluowanego systemu

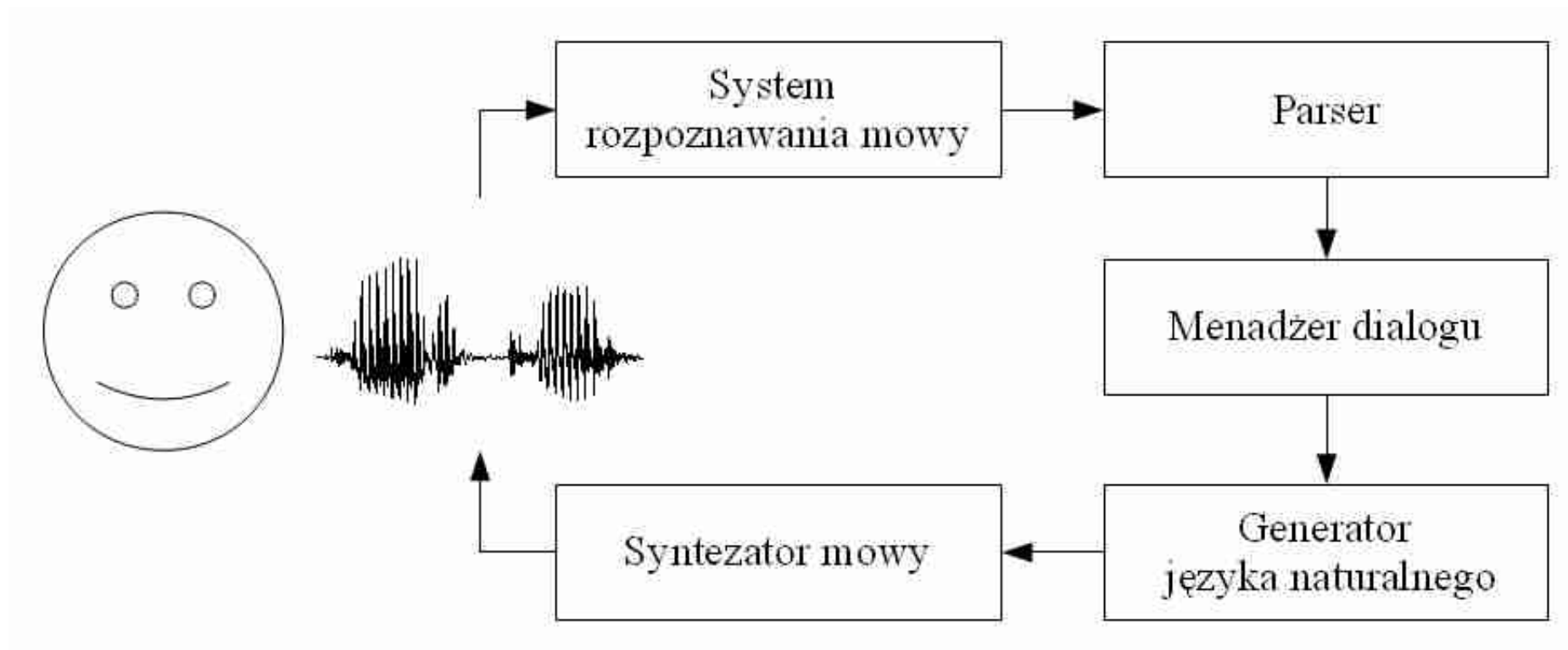
Metodologia

- Językoznawstwo:
 - lingwistyczne + psycholingwistyczne teorie dialogu
 - korpusowo-lingwistyczne opisy dialogu
- Językoznawstwo komputerowe:
 - modelowanie dialogu automatami
- Eksperymentalna technologia mowy:
 - Komponent syntezy mowy systemu dialogowego
 - tworzenie ‘mikrogłosów’: technika + ewaluacja
 - głos dla systemu prototypowego + ewaluacja

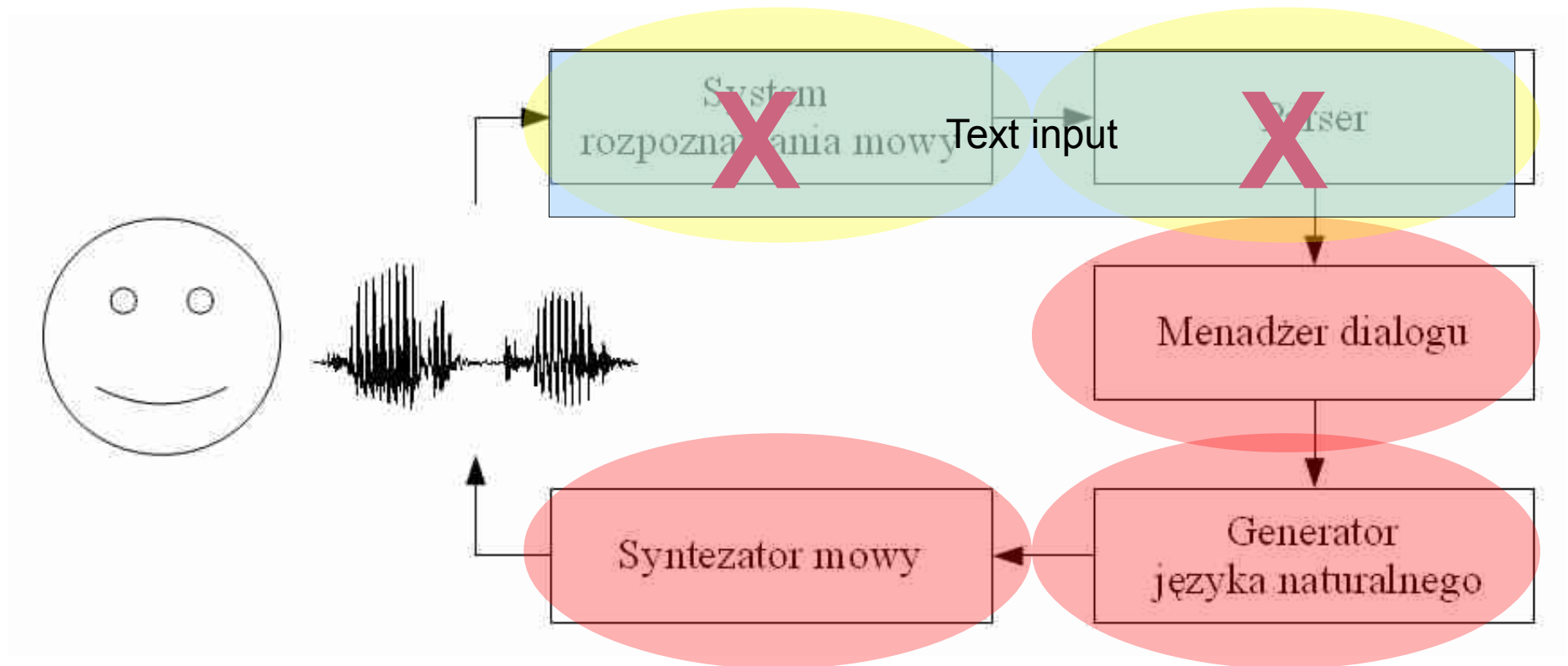
Cele i ograniczenia

- Cele ukierunkowane na:
 - podstawowy system dialogowy do weryfikacji koncepcji, oparty na głównych twierdzeniach i łączący komunikację pisemną i mówioną
 - prototyp demonstrujący:
 - metodologię
 - realizację twierdzenia o dopasowaniu semantycznym
 - w symulowanych, ale realistycznych warunkach stresowych
- Celem nie jest:
 - rozwój pełnego funkcjonalnego systemu dialogowego spełniającego standardy inżynierskie

Ogólny schemat systemu dialogowego



Schemat Text-In-Voice-Out



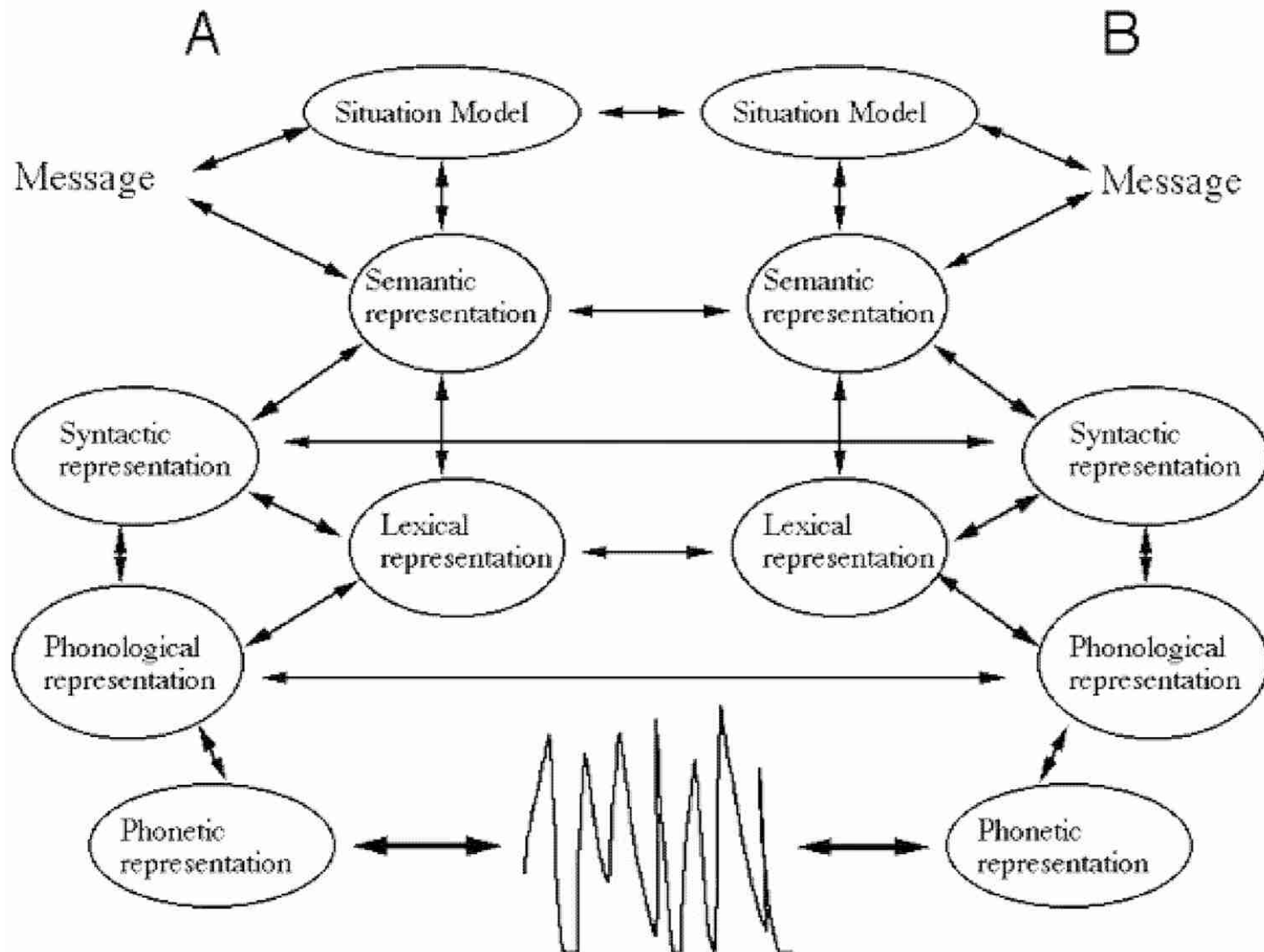
Schemat TIVO jest często stosowany w systemach poleceń i kontroli oraz w systemach protezowych (por. Stephen Hawking)

Dopasowanie interaktywne

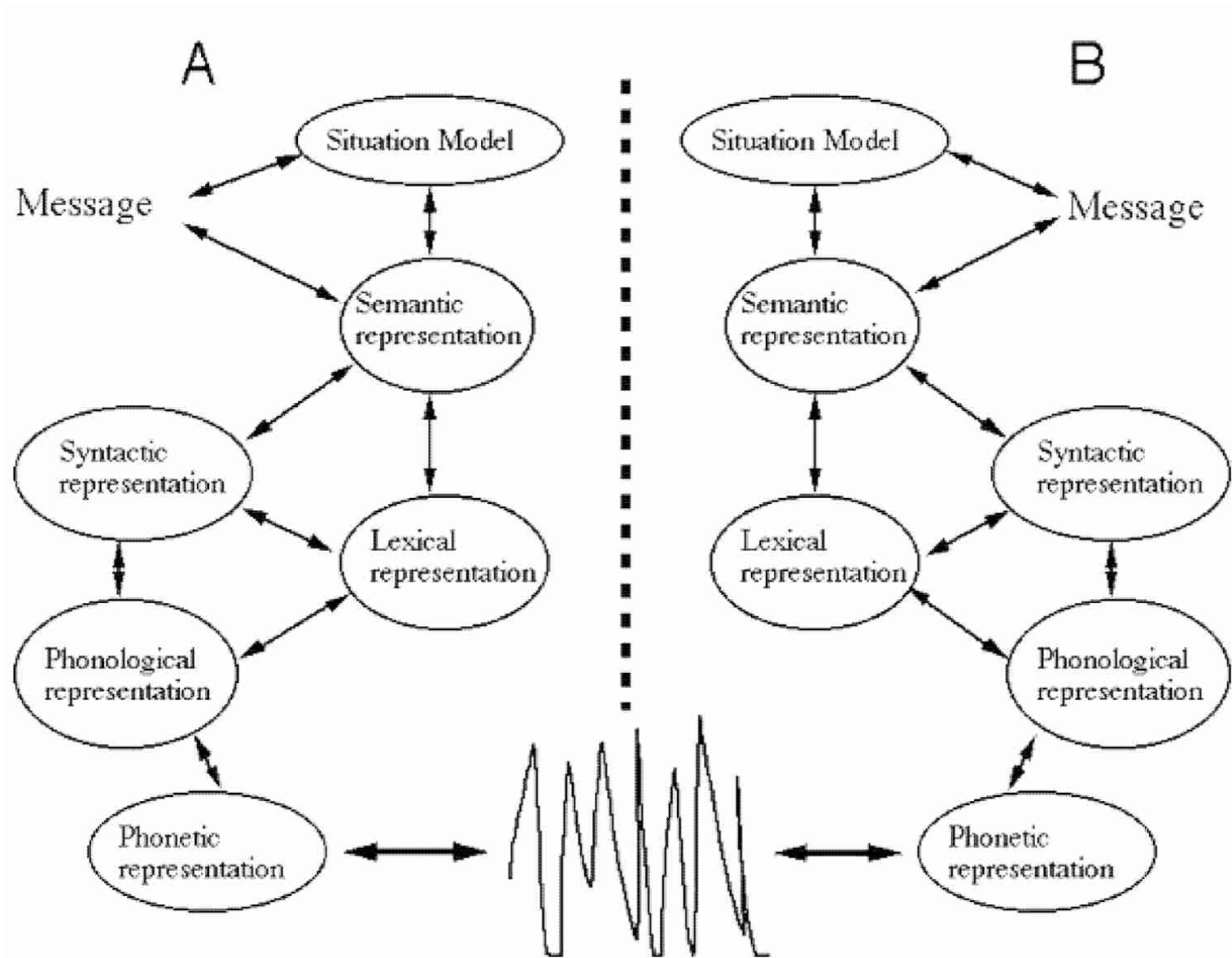
- Definicja ‘dopasowania’:
 - podstawową funkcją dopasowania jest koordynacja pomiędzy interlokutorami w celu uzyskania pomyślnego wyniku komunikacji
- Typy dopasowania:
 - dopasowanie w dialogu odbywa się na poziomie semantycznym, syntaktycznym i pragmatycznym (Pickering & Garrod 2004)

Model dopasowania interaktywnego

(Pickering & Garrod 2004)



Ale może jednak transmisja jest autonomiczna...?

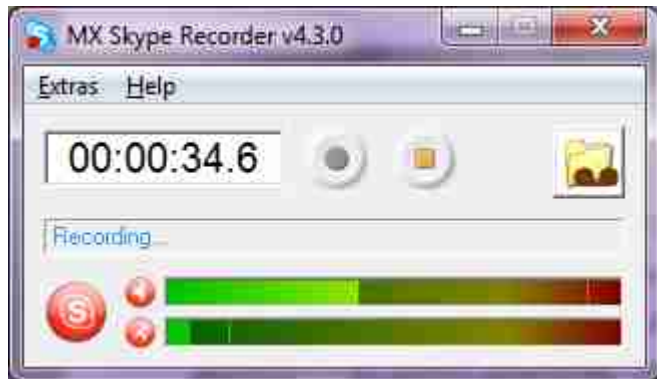


Projektowanie korpusu

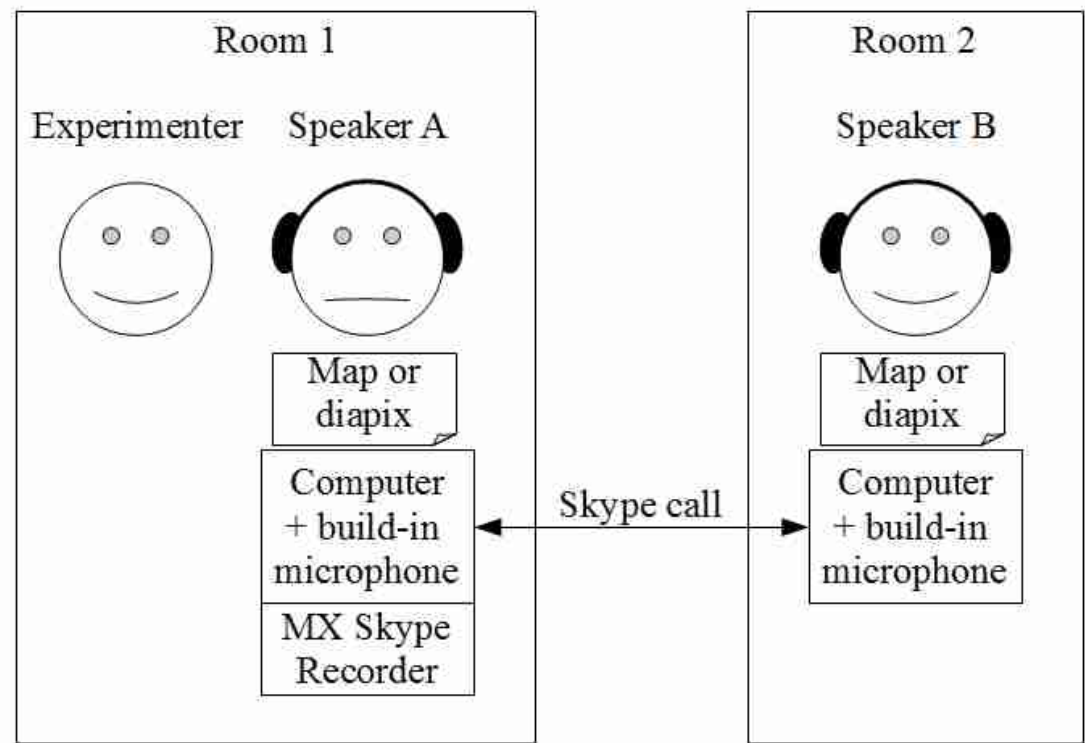
- osoby nagrywane
 - 12 par & 3 pary kontrolne
 - 15 mężczyzn & 15 kobiet, M:M, M:K, K:K
- sytuacja publiczna
 - osoby, które się nie znały
 - osoby z dużą różnicą wieku
 - różny status akademicki: nauczyciel – student
- scenariusze: nagłego wypadku & neutralny
- metody: map-task & diapix
- nagranie neutralne: “Sto lat samotności” (1967)
Gabriel García Márquez

Nagrania

005:00



48kHz sampling rate

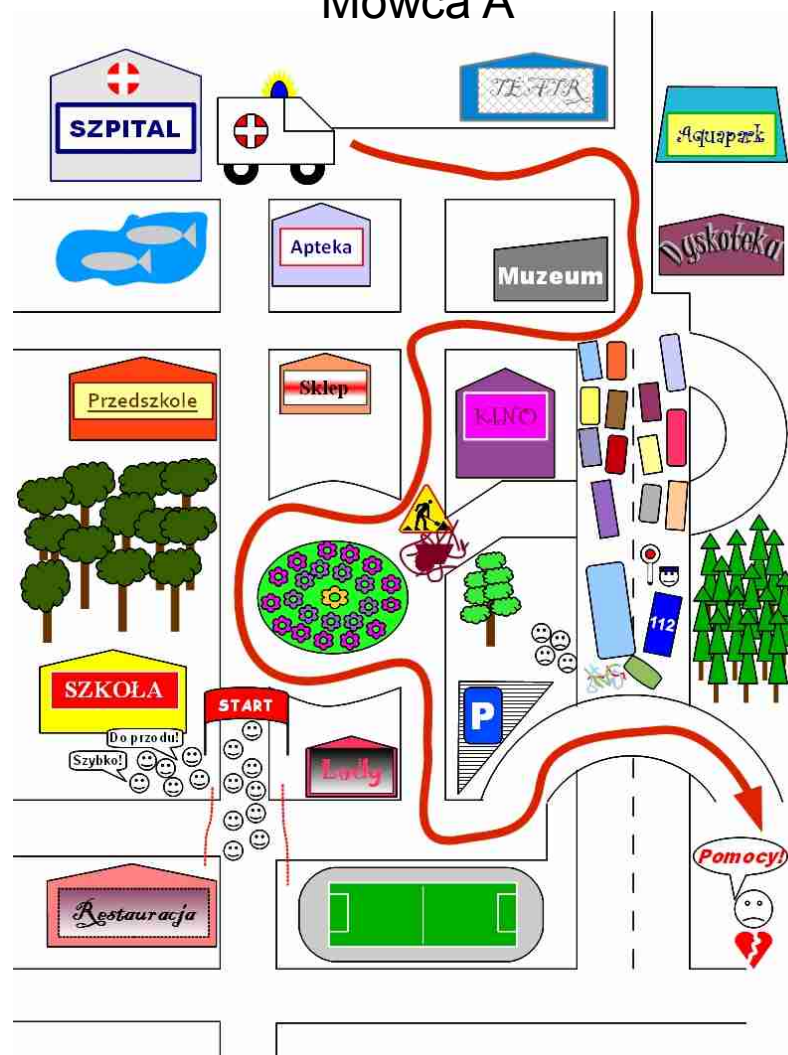


Materiały do pozyskiwania mowy

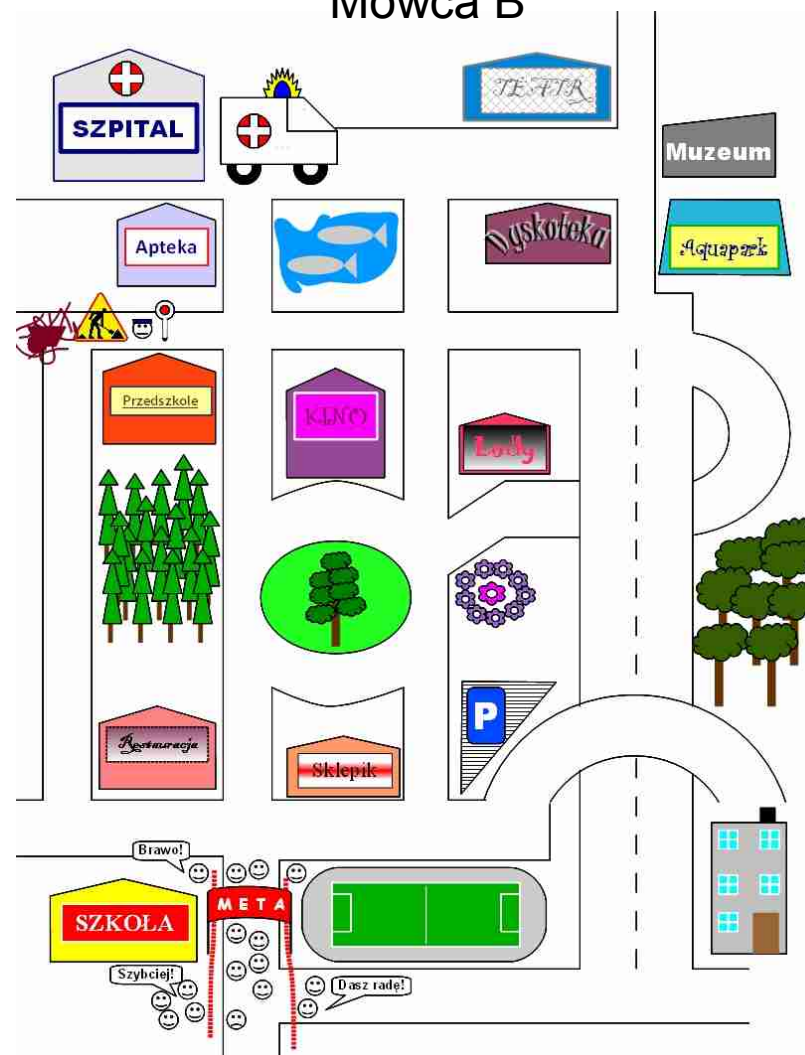
1. Scenariusz nagłego wypadku

Mapa scenariusza nagłego wypadku dopasowanie + rozbieżność

Mówca A



Mówca B



Mapa scenariusza nagłego wypadku dopasowanie + rozbieżność

Przykład rozbieżności

A: spk//znaczy* oczywiście nie ma nie ma pan
właśnie wyjścia tam na rondzie yyy są roboty
więc w prawo

B: acha

A: spk

B: spk//bo właśnie w* z* z mojej mapy wynika
że nie ma robót

Diapixy scenariusza nagłego wypadku

Mówca A



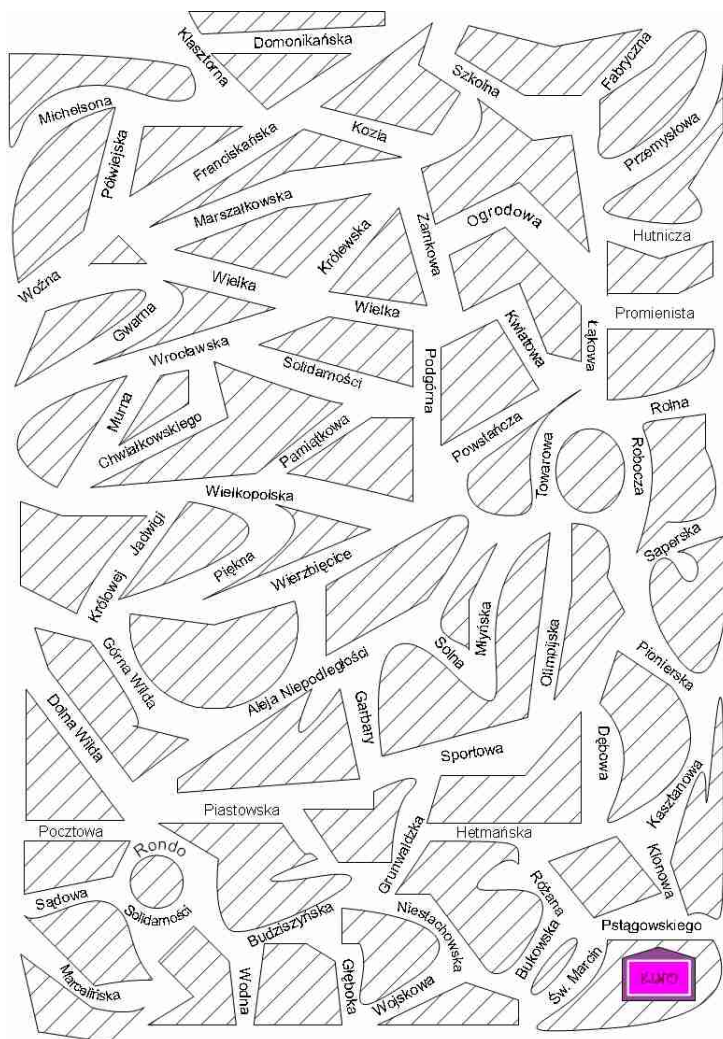
Mówca B



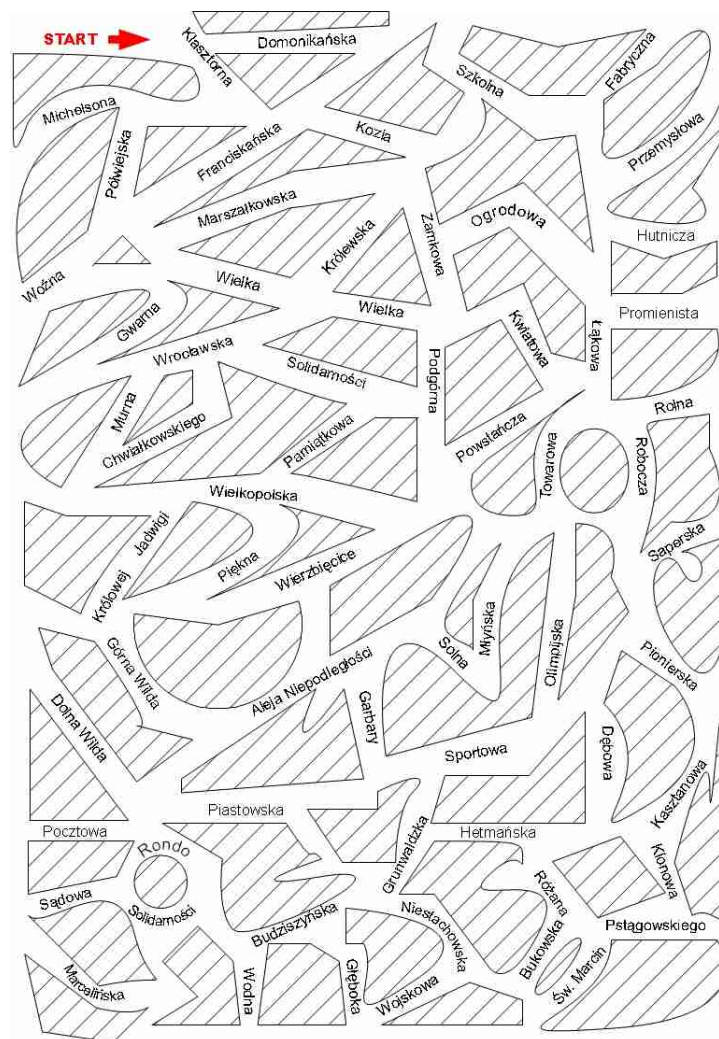
2. Scenariusz neutralny

Mapy scenariusza neutralnego

Mówca A



Mówca B



Diapixy scenariusza neutralnego (Bradlow et al. 2007)

Mówca A



Mówca B



Przechowywanie odpowiedzi

1 odpowiedź



3 nakładające się odpowiedzi



Dane korpusu (w sekundach)

	Map task: ER	Diapix: ER	Map task: Neutral	Diapix: Neutral	Reading
All	2325	3957	2353	5391	1111
Min	54	59	72	150	31
Max	325	310	251	568	49
Mean	155	264	157	359	37
Overall	<i>15137sec = 4h 12m</i> <i>2726298KB = 2.6GB</i>				

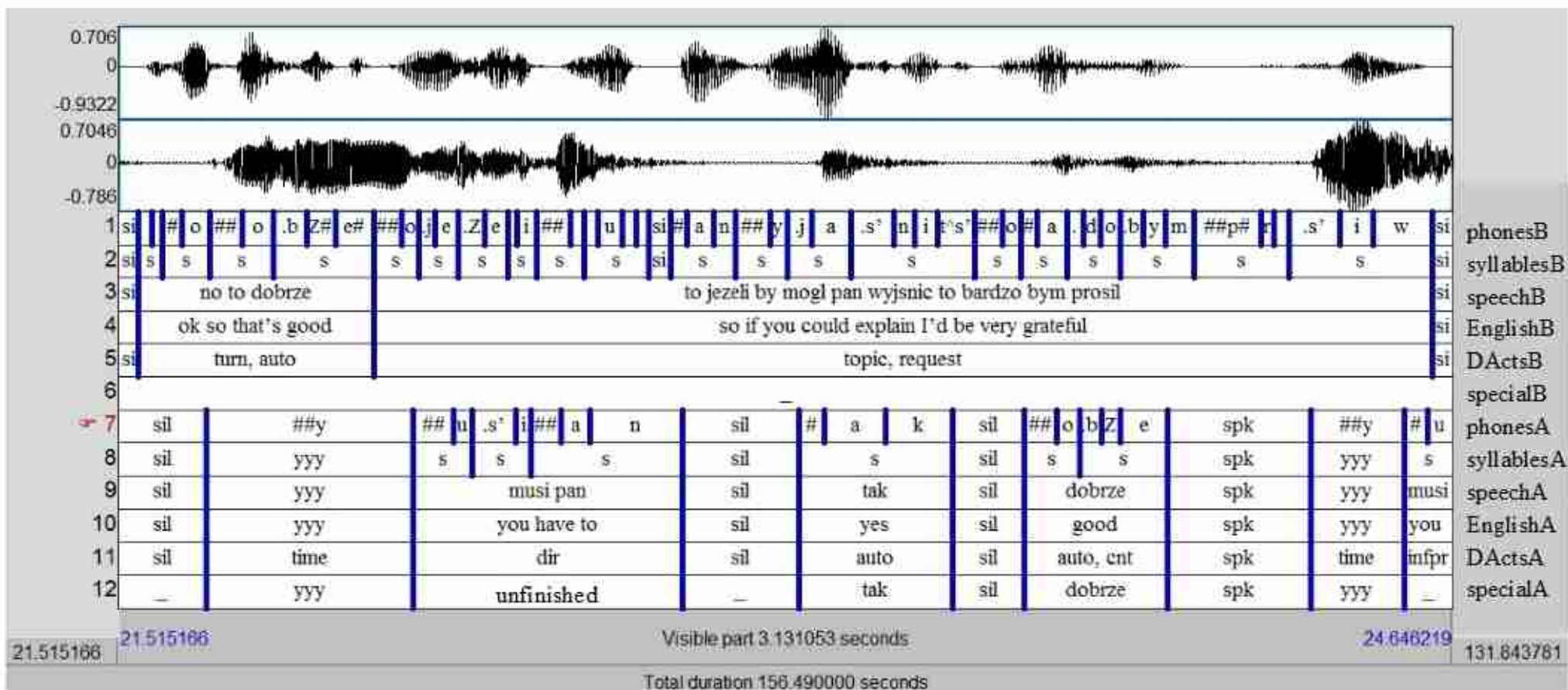
Metadane

- Date
 - Folder name, e.g. 2011-02-15_03_MM.
 - Subject's ID
 - Subject's sex
 - Subject's name
 - Subject's age
 - Subject's height
 - Subject's weight
 - Subject's education
 - Subjects occupation, e.g. Ph.D. student
 - Notes
 - Place of the recording
 - Length of recording
 - Size of recording in KB
- youngest: 19, oldest: 53,
max age difference: 32

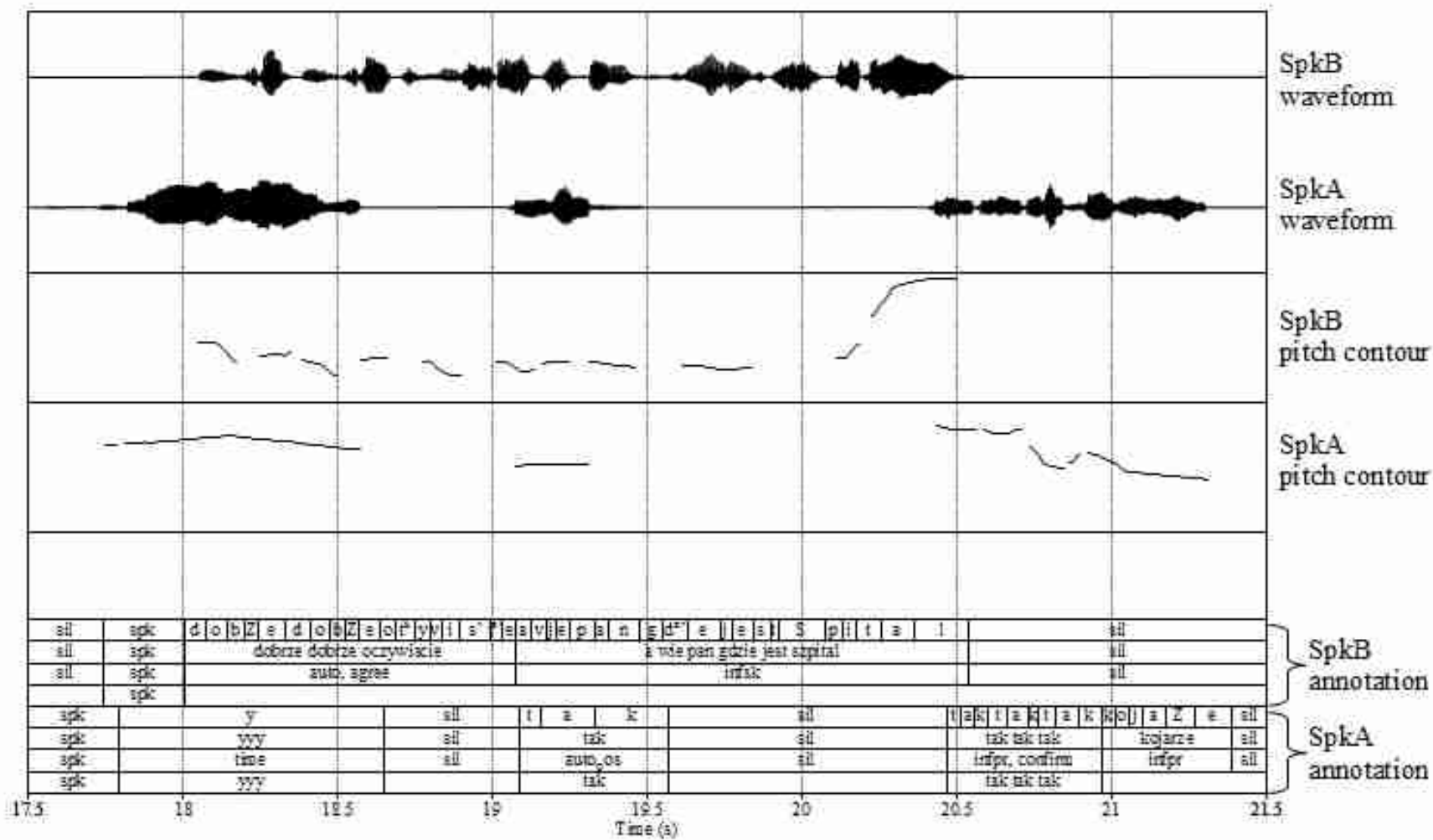
Analiza ogólna

- Map task: nagły wypadek – szybka mowa, formalna, dominacja mówcy A
- Diapix: nagły wypadek – mowa formalna, dialog kooperatywny
- Map task: kino – mowa nieformalna, dużo wtrąceń (auto-feedback) mówcy B, zabawna rozmowa
- Diapix: zakupy – mowa nieformalna, dialog kooperatywny, zdrobnienia

Anotacja dialogu



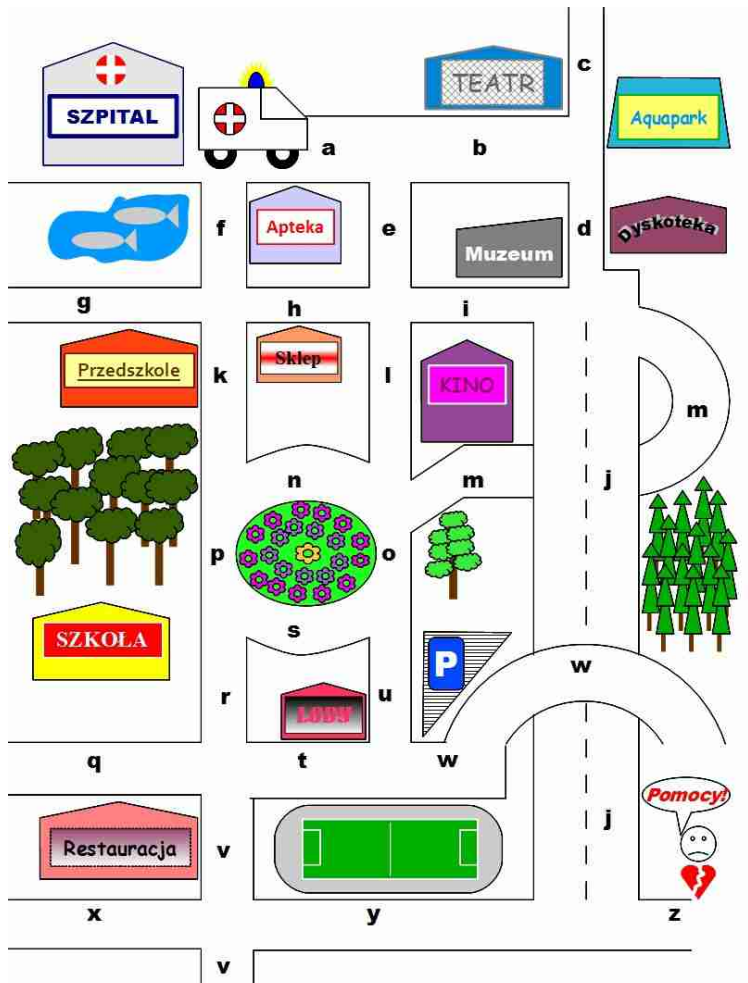
Analiza F0



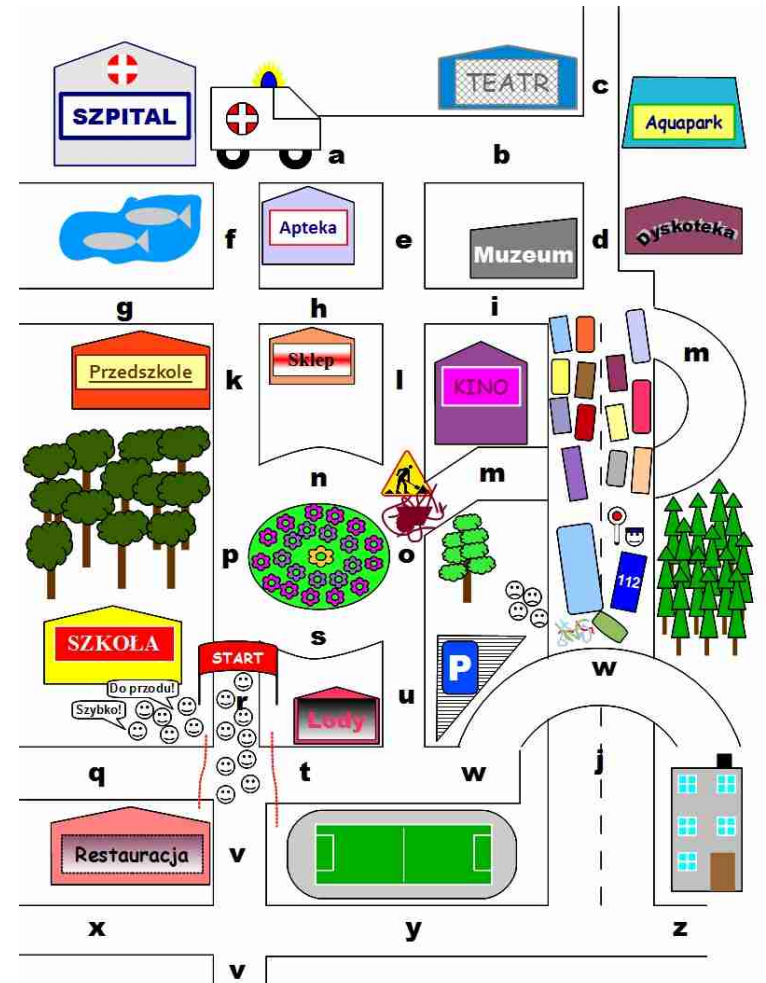
System dialogowy

Semantyczne dopasowanie + rozbieżność

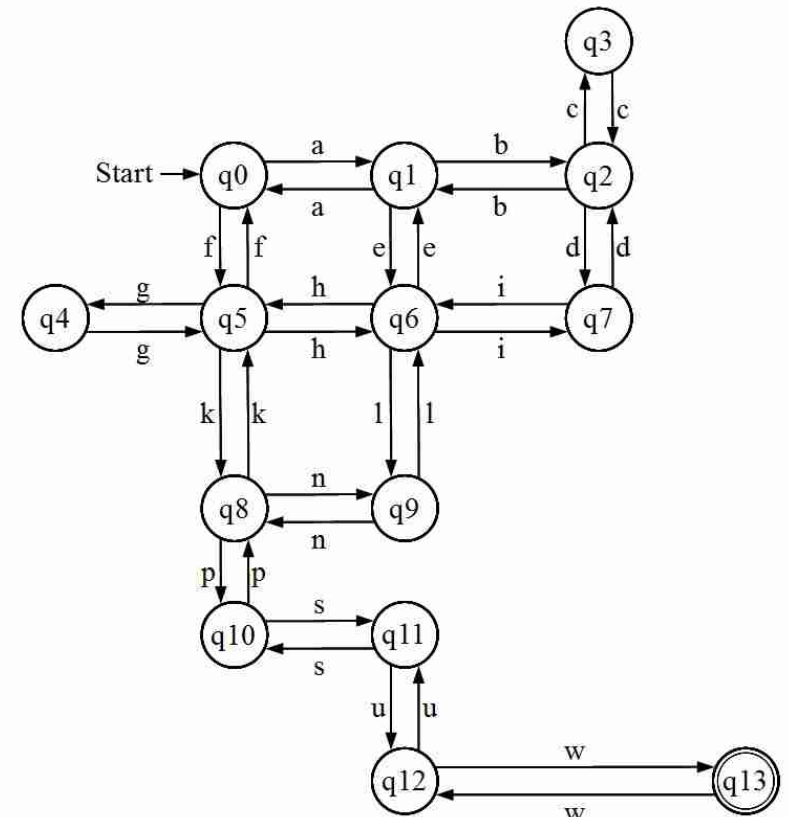
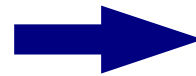
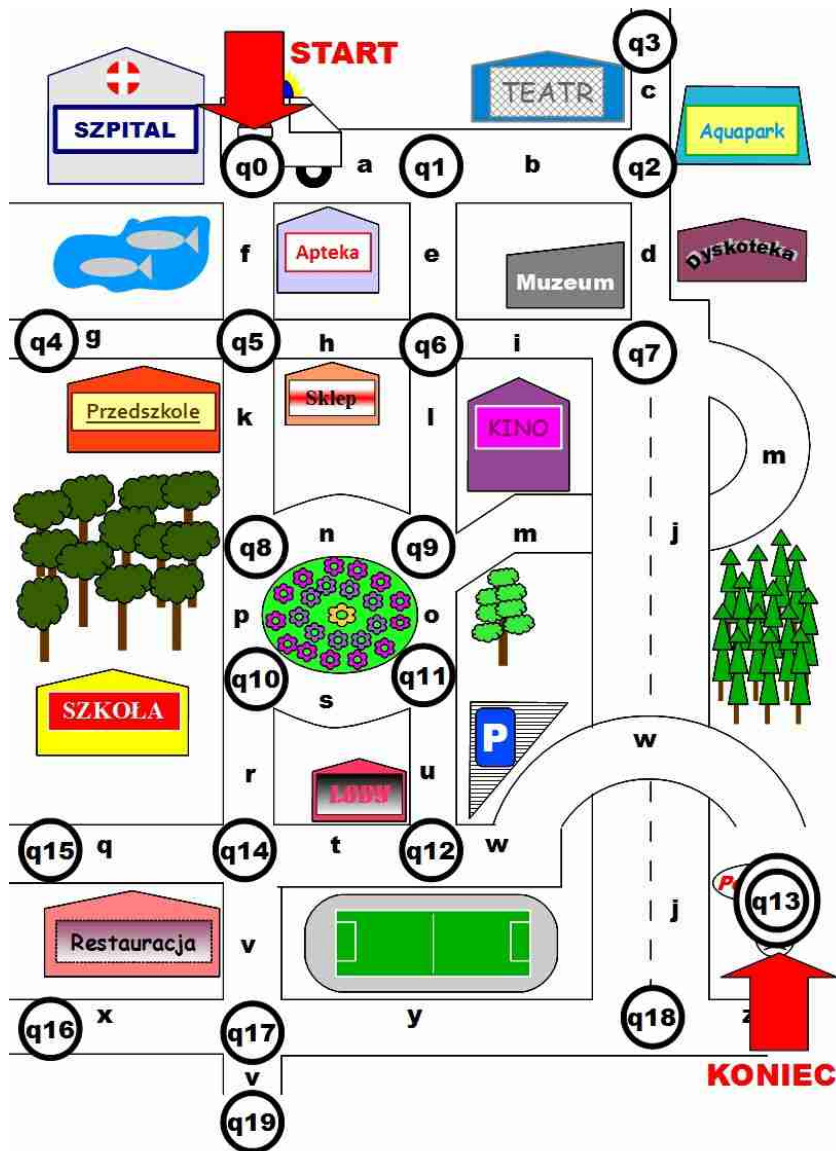
Człowiek



Komputer

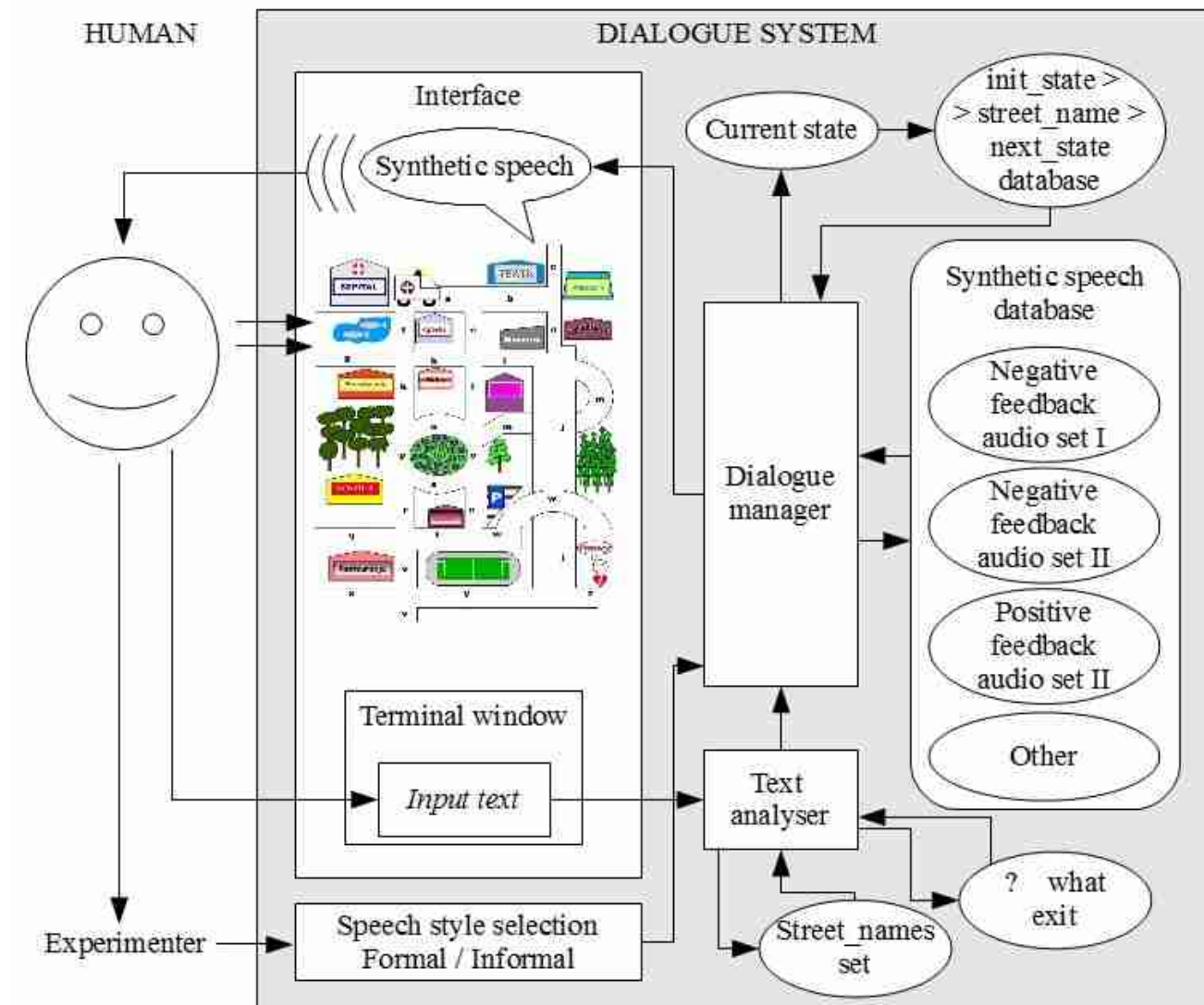


Semantyczny FSA: mapa + droga

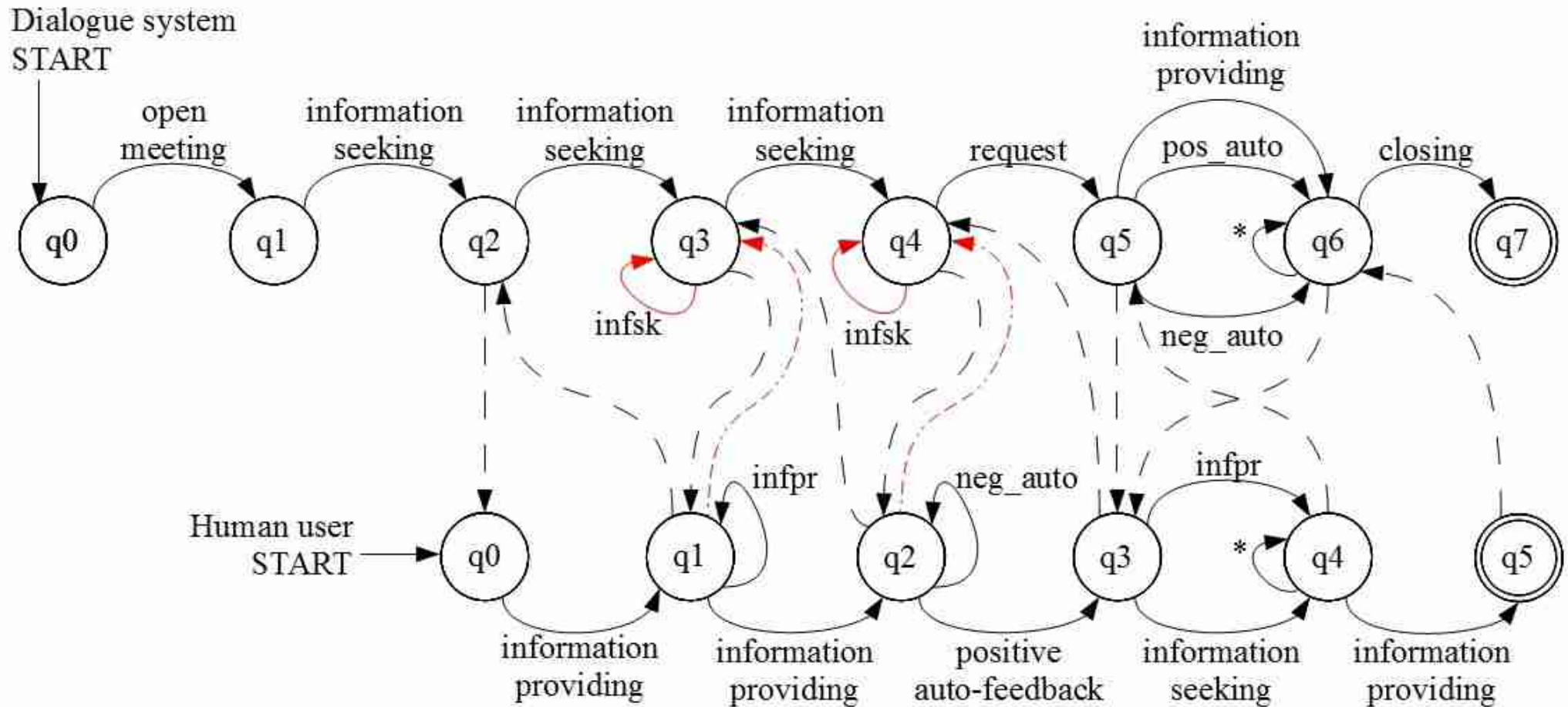


Architektura systemu dialogowego

Text-In-Voice-Out (TIVO)



FSA systemu dialogowego



Linie przerywane pokazują możliwe zmiany kolejek dialogowych

Fragment przykładowego dialogu

Dzień dobry! Witam! Jak się nazywasz?	Anna
Kobieta czy mężczyzna? (<i>Drukowany tekst:</i> k / m)	k
A wie Pani gdzie jest szpital?	tak
No to dobrze. To jeżeli mogłaby Pani wyjaśnić drogę, to bardzo bym prosił.	f
Tak, dobrze.	k
Acha, czyli rozumiem: prosto.	p
A potem?	r
Chyba coś się musiało zmienić – nie mogę tam.	?
<i>Drukowany tekst:</i> Karetka jest na: q10 (+ <i>mapa</i>)	r
Nie, nie, nie. Nie ma właśnie wejścia tam.	co
... No postaramy się. W razie czego będziemy jeszcze dzwonić. Także... proszę czekać. Niedługo przyjeżdżamy. Dziękuję.	-

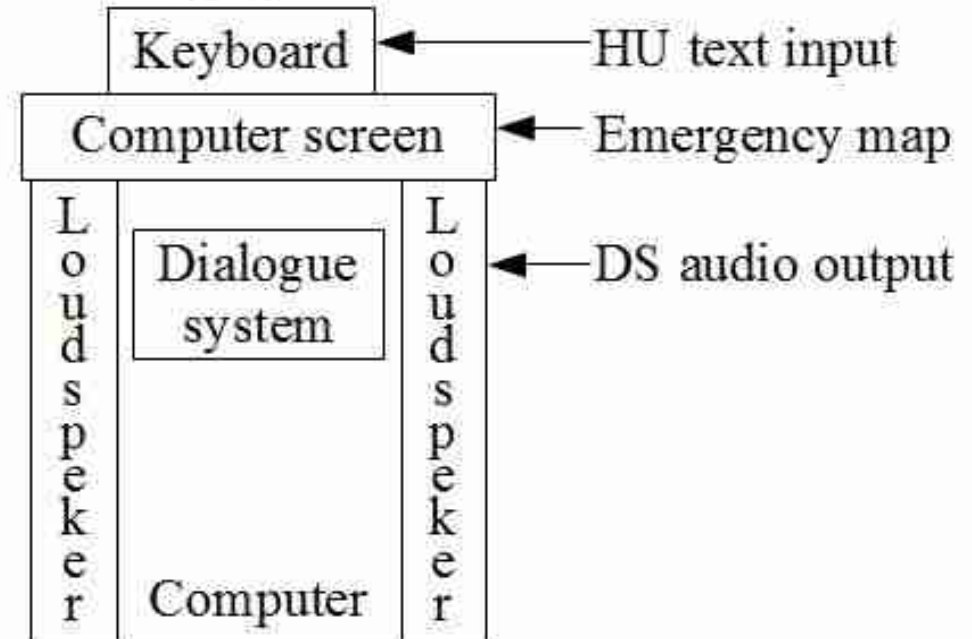
Ewaluacja

- 52 osoby
 - Formalny:
14 kobiet & 12 mężczyzn
 - Nieformalny:
14 kobiet & 12 mężczyzn
 - Wiek średni: 22/23 lata
- Wszystkie odpowiedzi: 1072
- Czas: 152min 22s
 - Min: 1min 51s
 - Max: 4min 47s

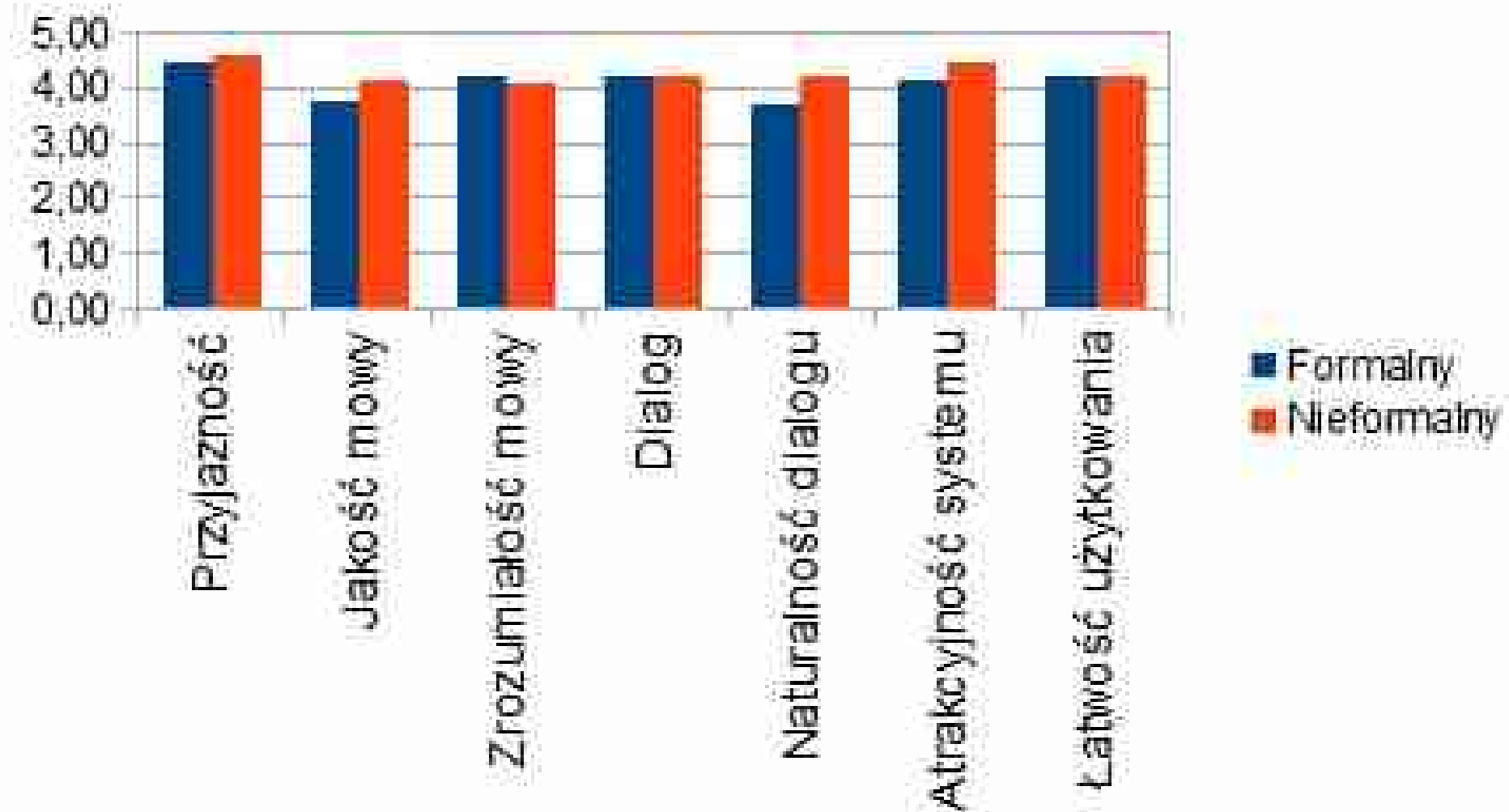
Experimenter



Human User



Graficzna prezentacja wyników skala MOS (Mean Opinion Score)



Podsumowanie zadań rozwojowych (1)

- Specyfikacje:
 - semantyczne dopasowanie w dialogu z wyborem stylu mowy formalnym i nieformalnym
 - realistyczny scenariusz warunków stresowych (dialog w sytuacji alarmowej z centrum telefonicznym szpitala)
- Projektowanie:
 - selekcja formalnych własności automatów
 - automat semantycznego dopasowania
 - automat kontroli dialogu
 - korpusowe badania lingwistyczne (ogólne + konkretny scenariusz)
 - formalna analiza aktów dialogowych + modelowanie skończenie-
stanowe

Podsumowanie zadań rozwojowych (2)

- Zasoby:
 - Lingwistyka korpusowa: korpusy + anotacja + modele
 - ogólna: ...
 - modelowanie automatami: ...
 - konkretne scenariusze: ...
- Implementacja:
 - ekstrakcja informacji z korpusów
 - stworzenie głosu syntetycznego
 - prototyp systemu dialogowego:
 - FSA dopasowania semantycznego
 - FSA menadżera dialogu

Podsumowanie

- Prototyp weryfikujący koncepcję
 - wykazano, że cel modelowania semantycznego dopasowania został osiągnięty
 - poparł twierdzenie, że tradycyjny zestaw emocjonalnych etykiet może być zastąpiony przez formalny-nieformalny styl mowy w systemie dialogowym
- Dopasowanie jest
 - nie tylko odpowiednie dla sytuacji publicznych
 - ludzkie: semantyczne

Osiągnięcia praktyczne

- Synteza mowy
 - ekstraktor zdań bogatych fonetycznie
 - automatyczny ekstraktor difonów
 - syntetyczny głos męski PL2
- Korpus polskich dialogów (4 godz. 12 min):
 - 12 par & 3 pary kontrolne
 - sytuacja publiczna
 - scenariusze: nagły wypadek & neutralny
 - metody: map-task & diapix
 - czytanie neutralne
- Prototypowy system dialogowy

Pytanie

- Jakie mogą być problemy nagrań dialogów telefonicznych, samochodowych, z centrum handlowego?
- Jak wyeliminować te problemy?

Przeanalizuj dialog

- Przeanalizuj dialog z punktu widzenia zaprojektowania systemu dialogowego
- <http://bachan.speechlabs.pl/files/dialogues/dialog.wav>

Do zobaczenia za tydzień!