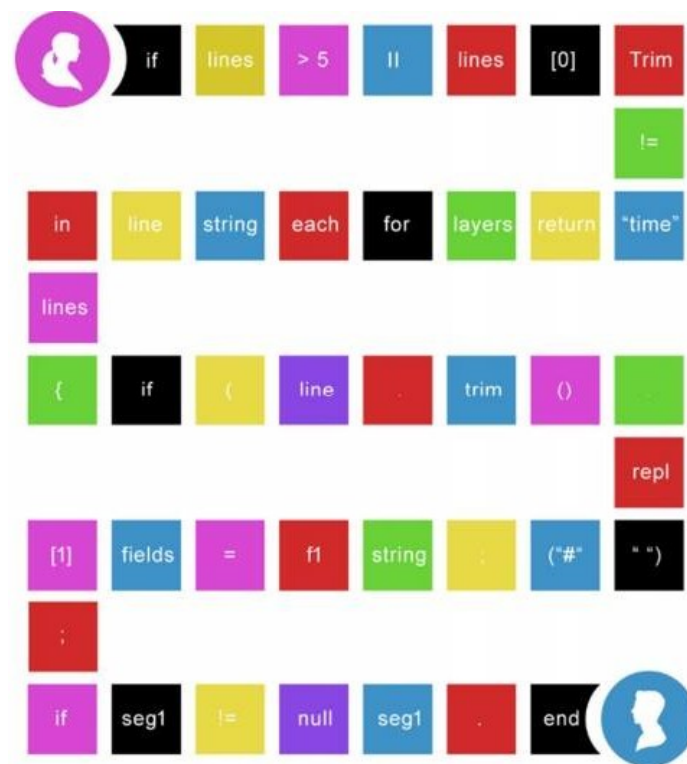


Bazy nagrań mowy

Eksploracja, interpretacja, analiza

Katarzyna Klessa
klessa@amu.edu.pl



Wykład otwarty, UAM, Wydział Neofilologii, dn. 16.03.2006



Bazy nagrań mowy – plan wykładu

- Wybrane etapy poprzedzające proces nagrywania mowy a ich skuteczna eksploracja. Wielokrotny użytek i współdzielenie zasobów.
- Interpretacja – przykłady baz nagrań mowy. Jakie dane badamy, a jakie chcielibyśmy badać?
- Jak technologia może usprawniać proces tworzenia zasobów na różnych jego etapach, np. anotacja nagrań, zarządzanie danymi.



Bazy nagrań mowy – dla kogo mogą być użyteczne?

specjaliści z dziedziny technologii mowy
akustycy informatycy lekarze
językoznawcy socjologowie terapeuci
fonetycy psycholingwiści
uczniowie i studenci historycy
“zwykli” użytkownicy języka
osoby zainteresowane rewitalizacją języków

- Grupa użytkowników a sposób udostępniania prezentacji danych
 - bazy nagrań przeznaczone do badań naukowych nad sygnałem mowy,
 - analiza zaburzeń mowy, porównanie z normą
 - bazy dla technologii: synteza, rozpoznawanie mowy i mówcy,
 - archiwa językowe dla specjalistów i popularyzatorskie.
- Inni użytkownicy niż ci pierwotnie planowani?

Wielokrotny użycie i współdzielenie zasobów

“re-usability”





Wybrane etapy prac

- ☐ wyszukiwanie informacji, np. analiza stanu badań, przeszukiwanie tekstów źródłowych
- ☐ projektowanie ankiety osobowej uczestnika nagrań (metadane) ✓
- ☐ projektowanie scenariusza sesji nagraniowej ✓
- ☐ planowanie zasad dostępu do danych, współdzielenia
- ☐ dobór technologii do zarządzania danymi ✓
- ☐ dobór sprzętu i oprogramowania do nagrań ✓
- ☐ zgoda na nagrania - projekt, nadzorowanie, konsultacje prawne ✓
- ☐ rekrutacja mówców
- ☐ zbieranie danych do kwestionariusza osobowego, ankiety ✓
- ☐ prowadzenie nagrań mowy ✓
- ☐ nadzorowanie wypełnienia ankiety osobowej uczestnika nagrań
- ☐ zarządzanie bazą danych, serwerem
- ☐ tworzenie kopii bezpieczeństwa danych
- ☐ transkrypcja fonetyczna nagrań ✓
- ☐ segmentacja nagrań (podział na głoski, sylaby wg czasu występowania w nagraniu, 'time-aligned') ✓
- ☐ zarządzanie anotacją nagrań ✓



Dane Metadane

– metadane – dane o danych

- » dla nagrań mowy – dane osobowe mówcy, wiek, płeć, jego stan psychiczny, dane środowiska nagraniowego, mikrofonów, osób nagrywających, innych osób obecnych przy nagraniu, innych czynników...
- » dla danych tekstowych – dane jak np. w katalogu bibliotecznym, informacje o sposobach kolekcjonowania danych...
- » załączniki – np. zgody na wykorzystanie danych
- » uwaga: rozróżnienie nie jest jednoznaczne, jeśli korpus jest wielokrotnego użytku metadane mogą stać się danymi i na odwrót

Metadane a współdzielenie zasobów

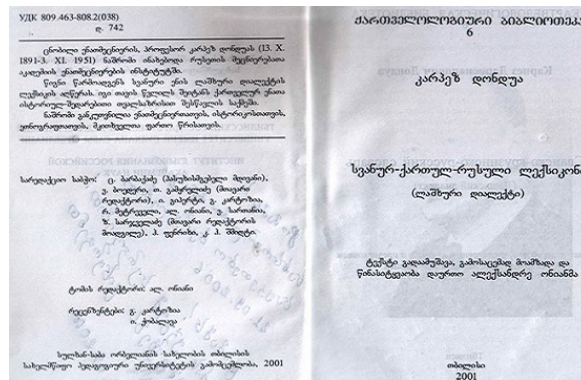
“re-usability”

- Rozbudowany system metadanych może przyczynić się do większej użyteczności korpusu, także w obszarach poza pierwotnym docelowym obszarem zastosowań – np. nagrania mowy na potrzeby badań fonetycznych mogą stać się wartościowym materiałem na potrzeby psychologii, socjologii, terapii języka i mowy, technologii informatycznej i in.
- Właściwy opis i dokumentacja zasobów mogą stanowić o ich większej użyteczności.
- Aspekty współdzielenia danych i komercyjne.



Dane surowe i korpusy językowe

Dane surowe (raw/primary data) – bez obróbki, systematyzacji, nieprzetworzone / przetworzone w niewielkim stopniu



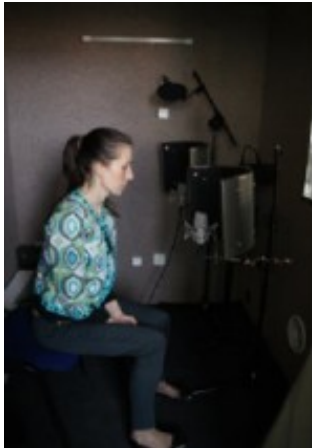
Zdjęcia: languagesindanger.eu



Dane surowe i korpusy językowe

- **Korpus** (ang. *corpus*, *corpora*; z łac. *ciało*, **określona** całość)
 - dane zebrane zwykle wg założonego z góry klucza, w określonym celu
 - zwykle już w trakcie projektowania korpusu planuje się określoną jego **strukturę** oraz **sposób opisu** danych
 - **uporządkowana** kolekcja nagrań, tekstów pisanych, transkrypcji wypowiedzi i / lub powiązanych z nimi nagrań mowy

Różne sposoby pozyskiwania danych



Zdjęcia: languagesindanger.eu



“Paradoks obserwatora” Observer's Paradox

Celem badań językoznawczych prowadzonych w danej społeczności ma być ustalenie, jak ludzie rozmawiają ze sobą wtedy, gdy nikt ich systematycznie nie obserwuje.

Ale takie dane możemy pozyskać tylko przez systematyczną obserwację.

William Labov (1972) *Sociolinguistic Patterns*.

Fot. L. W. Hine, *The Observer and the Observed*.





Zgoda na nagrania

Zgoda na nagrania głosu i ich wykorzystanie - można jej udzielić w formie pisemnej (optymalnie) lub ustnej, jeśli zostanie ona zarejestrowana. Warto sprawdzić aktualne przepisy przed rozpoczęciem nagrań.

Tekst zgody powinien być jasny i wolny od specjalistycznej terminologii. Sporządź tekst tak, żeby podpisując zgodę, uczestnik nagrania:

- potwierdzał, że dobrowolnie zgadza się na udział w sesji nagraniowej;

mógł wyrazić osobną zgodę na:

- nagrywanie we wszystkich trybach (audio / wideo);
- użycie danych do celów badawczych;
- publikację danych;
- archiwizację danych.

Pomyśl nad stworzeniem dwóch kopii formularza dla siebie i osoby, która ma być nagrywana. Pamiętaj o podpisach!





Mowa spontaniczna: baza JST / CREST

- Baza mowy spontanicznej i ekspresywnej JST/CREST.
- Baza złożona z szeregu podzbiorów, z których jeden zebrano tak, że ochotnikom zamontowano małe przenośne urządzenia do nagrywania. Urządzenia rejestrowały spontaniczną mowę tych osób podczas codziennych czynności (w pracy, domu czy szkole) przez stosunkowo długi okres czasu (np. kilka miesięcy).





Mowa spontaniczna: język opiekuńczy - mikrokorpus UAM

- 4 matki mówiące do swoich dzieci, niemowląt <1 r.ż
- Okoliczności: karmienie, przewijanie, kładzenie do snu i in. codzienne czynności





Mowa spontaniczna a kwestie techniczne, prawne i etyczne

- Zarządzanie?
- Kontrola?
- Metadane?
- Przetwarzanie?
- Kwestie etyczne
- Ochrona danych osobowych





Scenariusze nagraniowe

- Wypowiedzi (dialogi, monologi):
 - spontaniczne
 - quasi-spontaniczne, wywoływane, zadaniowe
 - czytane (teksty ciągłe, listy wyrazowe, jednostki izolowane)
- Rejestracja zachowań językowych
- Wiedza o języku, świadomość języka



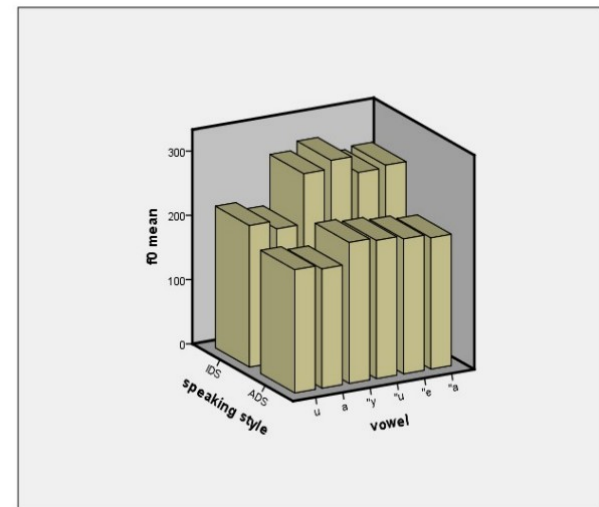
Mowa wywoływana, skierowana do dzieci i dorosłych - IDS vs. ADS korpus NeuroPerKog

Wypowiedzi izolowane (sylaby, krótkie słowa / logatomy) realizowane w sposób neutralny oraz jako skierowane do dzieci
 - instrukcja + zdjęcie jako sposób wywołania specyficznej realizacji wypowiedzi



Photo: Piotr Ciuchta, freeimages.com

Figure 1: F_0 means in IDS and ADS: in stressed (/e/, /a/, /y/, /u/) unstressed (/u/, /a/) vowels

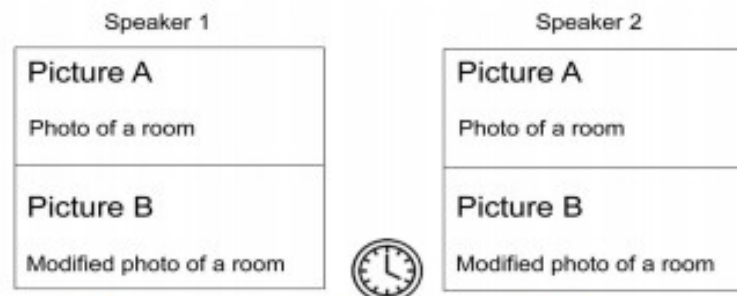


Dialogi zadaniowe, telefoniczne: Paralingua Dial

- Kilka etapów sesji nagraniowej – różne zadania: czytanie tekstu, 4 zadania dialogowe, ponowne czytanie tekstu
- Rozmowa przez telefon komórkowy, brak kontaktu wzrokowego
- Nagrania za pomocą kamer, mikrofonów i telefonów komórkowych.



. Recording scenario Dial 1A session settings (a different picture for each speaker), no time restrictions.



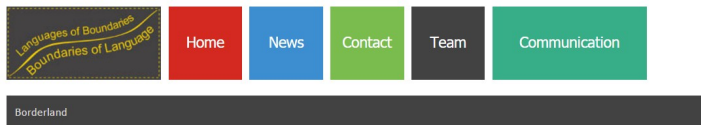
Recording scenario Dial 1B session settings (the same picture for both speakers). Time restrictions applied.



Dialogi zadaniowe

Korpusy multimodalne DiaGest, Borderland

- Zadania dialogowe zaprojektowane w celu wywołania gestykulacji – np. tzw. „origami task”, „maptask”
- Rejestracja audio i wideo, badanie interakcji mowy z innymi, językowymi i pozajęzykowymi składowymi wypowiedzi



Language of Boundaries and Boundaries of Language

The aims of the project are the documentation and interdisciplinary analysis of phenomena related to interpersonal communication on the border of languages and cultures.

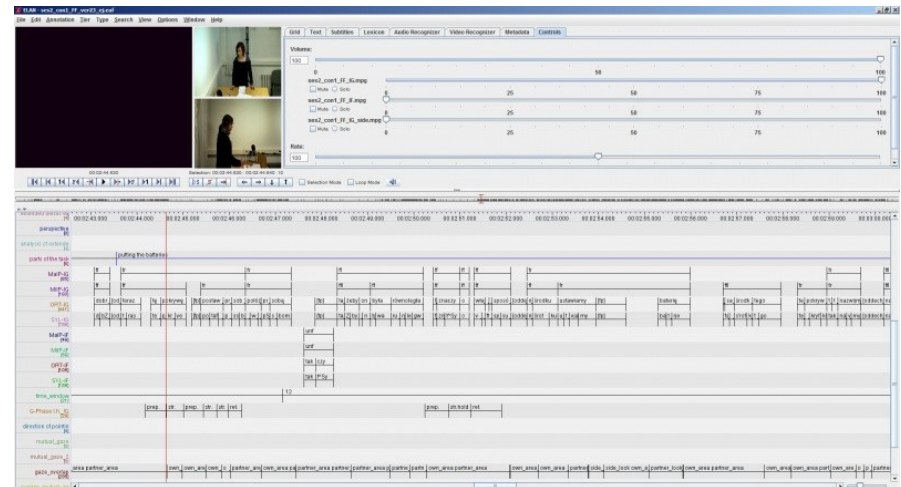
The main research focus is on paralinguistic phenomena (prosody, gestures) and accommodation within the communicative order and Słubice constitute an investigation and testing ground for fieldwork analyses of an exceptionally close co-existence

The project's goals include: studies of linguistic and cultural awareness, the sense of identity, the awareness of language communities of Słubice and Frankfurt Oder, the studies of the dynamics of the processes of mutual accommodation (cor the prosody and gestures, as well as empirical studies dedicated to the realisation of selected categories of gestures used



NARODOWY PROGRAM ROZWOJU HUMANISTYKI

Narodowy Program Rozwoju Humanistyki powstał w oparciu o konsultację umożliwiającą lepszą integrację polskiej humanistyki z humanistyką europejską. Programowi Rozwoju Humanistyki przyznano na I rozwój badań humanistycznych o zasadniczym znaczeniu dla zachowania



Www.bordland.amu.edu.pl,
screenshot z progr. E LAN
z languagesindanger.eu, dzięki
uprzejmości M. Karpińskiego

Portrety emocjonalne: Paralingua Emo

- Emotion portrayals, por. GEMEP Geneva corpus (Scherer et al.)
- Wypowiedzi neutralne leksykalnie, co najmniej jedna fraza, wyrazy o wysokiej częstotliwości występowania w codziennym użyciu, wszyscy mówcy – te same teksty wypowiedzi
- Emocje podstawowe i emocje o zróżnicowanym nacechowaniu i stopniu pobudzenia (*valence, activation*)

Teraz wszystko rozumiem.

Dzisiaj jest poniedziałek.

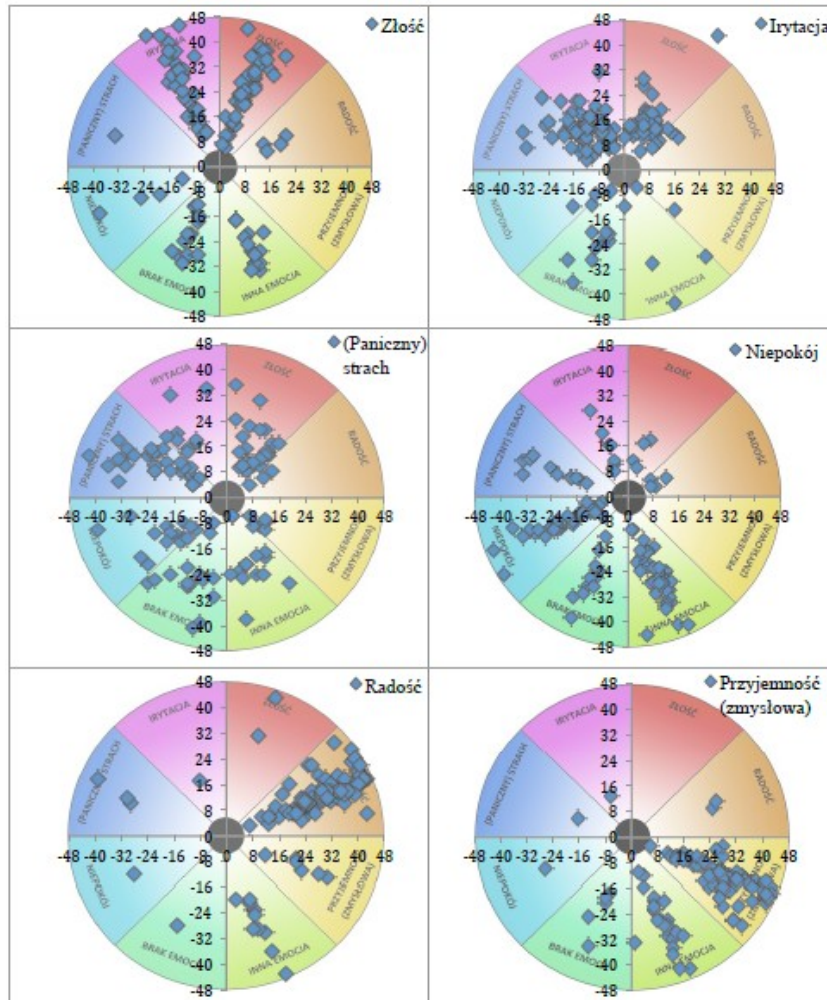
Od rana pada deszcz.

Powiedział, że nic się nie stało.

Jedziemy na wycieczkę do Grecji.



Portrety emocjonalne: Paralingua Emo





Archiwa językowe, dokumentacja języków

- Cele gromadzenia danych na potrzeby archiwizacji i dokumentacji oraz badań naukowych (fonetycznych, psycholingwistycznych, socjologicznych...) mogą być **różne**, ale wiele metod i technik **pokrywa się**
- Archiwa językowe np.
 - » <http://www.dialektologia.uw.edu.pl/index.php>
 - » <http://dobes.mpi.nl/>
 - » <https://www.soas.ac.uk/library/archives/>
 - » <http://inne-jezyki.amu.edu.pl/Frontend/>
- Wykorzystanie danych na potrzeby popularyzacji wiedzy:
<http://pl.languagesindanger.eu/map-html/pl/>

Dziedzictwo językowe Rzeczypospolitej

Baza dokumentacji zagrożonych języków

Baza dokumentacji zagrożonych języków zawiera informacje i materiały dotyczące zagrożonych odmian językowych stanowiących diachronicznym i synchronicznym, (z wyjątkiem gwar i dialektów samej polszczyzny), świadczących o bogactwie językowym Polski



Dyneburg p. Wanda i p. Jan



Kraslaw - Krypta Platerów



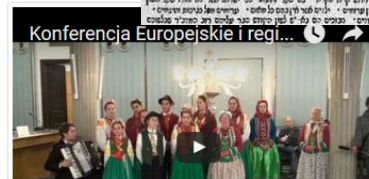
Molenna Starowierów w Lucynie



Lista języków

- iatgalski *
- (polski) jidysz *
- wilamowski *
- hałcnowski i bielsko-bialska wyspa językowa *
- armeno-kipczacki / ormiański-kipczacki
- białoruskie gwary
- czeskie gwary
- dolnoniemieckie dialekty
- karaimski
- łaskie gwary
- litewskie odmiany
- mazurskich Starobrzędowców gwara
- niemieckie dialekty Śląska

Na terenie Polski oraz w krajach sąsiadujących (odnoszących się do tytułowego języka Rzeczypospolitej) używanych jest wiele języków, o różnej - często niewielkiej - liczbie tej pory dokładnie udokumentowane, a ich istnienie jest zagrożone. Języki te są świadectwem dawnej Rzeczypospolitej i stanowią tym samym ważny element polskiego dziedzictwa językowego terytorium Rzeczypospolitej stanowiło i wciąż jeszcze stanowi (choć pod względem ilościowym, jak i jakościowym) obszar współwystępowania, różnymi wspólnotami komunikacyjnymi.





Reprezentatywność korpusów

Reprezentatywność – zakres w jakim korpus informuje o zjawiskach zachodzących w całej populacji; w przypadku korpusów językowych – w języku lub jego specjalizowanym podzbiorze

- **Odróżnia korpusy od archiwów językowych** *(jednak por. **korpusy** mowy spontanicznej, jak określić ich reprezentatywność?)
- **Możliwość generalizacji, wyciągania uogólnionych wniosków**



Korpusy statyczne vs. dynamiczne

Styczne – zbiór danych językowych dający informację na temat elementów systemu językowego w określonym momencie

- łatwiej kontrolować zrównoważenie zawartości korpusu,
- użyteczność może być bardziej ograniczona (dezaktualizacja).

Dynamiczne

- umożliwiają obserwację zmienności zjawisk językowych w czasie,
- trudniejsze w utrzymaniu, droższe,
- trudniejsza kontrola zrównoważenia i jakości danych
- Archiwa – często o charakterze dynamicznym, niekiedy każde nagranie „na wagę złota”.



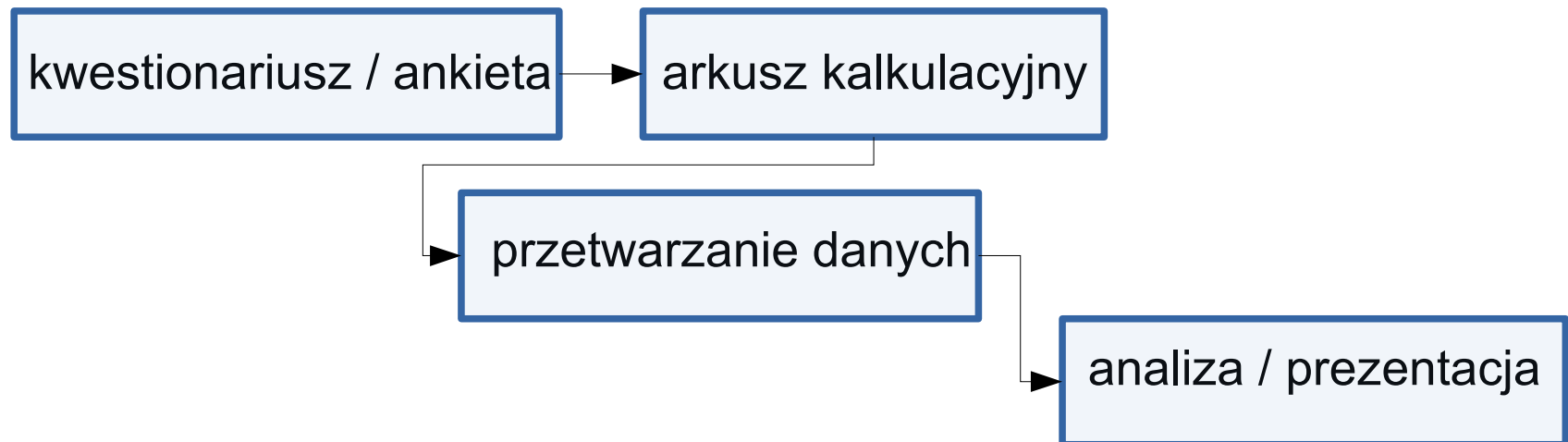
Opis i systematyzacja danych w korpusach - wybór metod i technologii

Annotation Pro 2.3.1.5 interface showing audio analysis and transcription.

File	FilePart1	FilePart2	FilePart3	Layer	Start	StartInMili	Duration	DurationInMilliseconds	Param1	Param2	Param3	Label
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	6228	141	1323	30				p
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	7551	171	1323	30				r
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	8874	201	1323	30				"o
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	10197	231	1764	40				S
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	11961	271	3528	80				e
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	15489	351	1323	30				p
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	16812	381	3528	80				"a
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	20340	461	1764	40				n
Asia_2_3362_Os_A.ant	Asia	2	3362	Phones	22104	501	3528	80				a

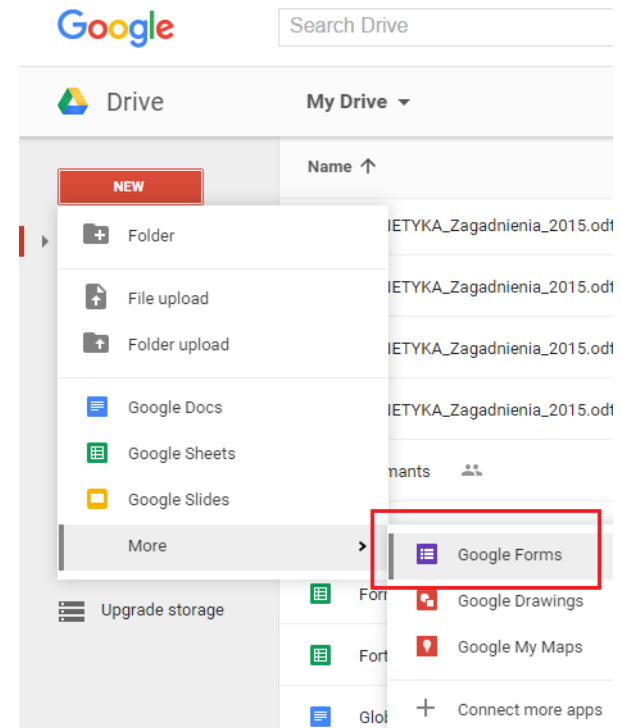
Techniki zbierania (meta)danych

- Kwestionariusze papierowe, elektroniczne, on-line
 - wady i zalety (czasochłonność, koszty, konieczność dostępu do Internetu i zasilania w miejscu nagrań)



Techniki zbierania (meta)danych

- Darmowe, intuicyjne narzędzia do tworzenia ankiet internetowych, np. Google Forms, ankietka.pl



Transkrypcja, segmentacja, anotacja

- Różnice w definiowaniu, wieloznaczność

[Dialekt mazowiecki - Mazury](#)

Tekst gwarowy — Biała Piska 1

PDF PRINT EMAIL

Justyna Garczyńska

Nagranie: Julia Pikacz, Anna Godziuk

Przepisanie: Monika Kresa

Opracowanie: Justyna Garczyńska

Informator: Helena Born, zamieszkała w Białej Piskiej, ur. 20.12.1933 r. w Kaliszkach. Rodzice pochodzili z Kaliszek, pracowali w majątku dziedzica, rodzina mieszkała w czworakach dworskich. Ukończyła sześć klas szkoły podstawowej niemieckiej. Pracowała w Pisku w fabryce drewna, a następnie w przedszkolu jako pomoc.

O wojnie

Urodziłam sie w Kaliszkach i tażeriska forma liczebnikowa dwie zastąpiona przez formę męską dwa lat. No ji tam mjeszkałam do czasu wojnyi. I po wojn..., jak wojna przeszła, to nazad z pegieerzu z Kaliszk to nam ten pan kazał take wozy przyszykować drabiniaste, nakładli słomy, tygo

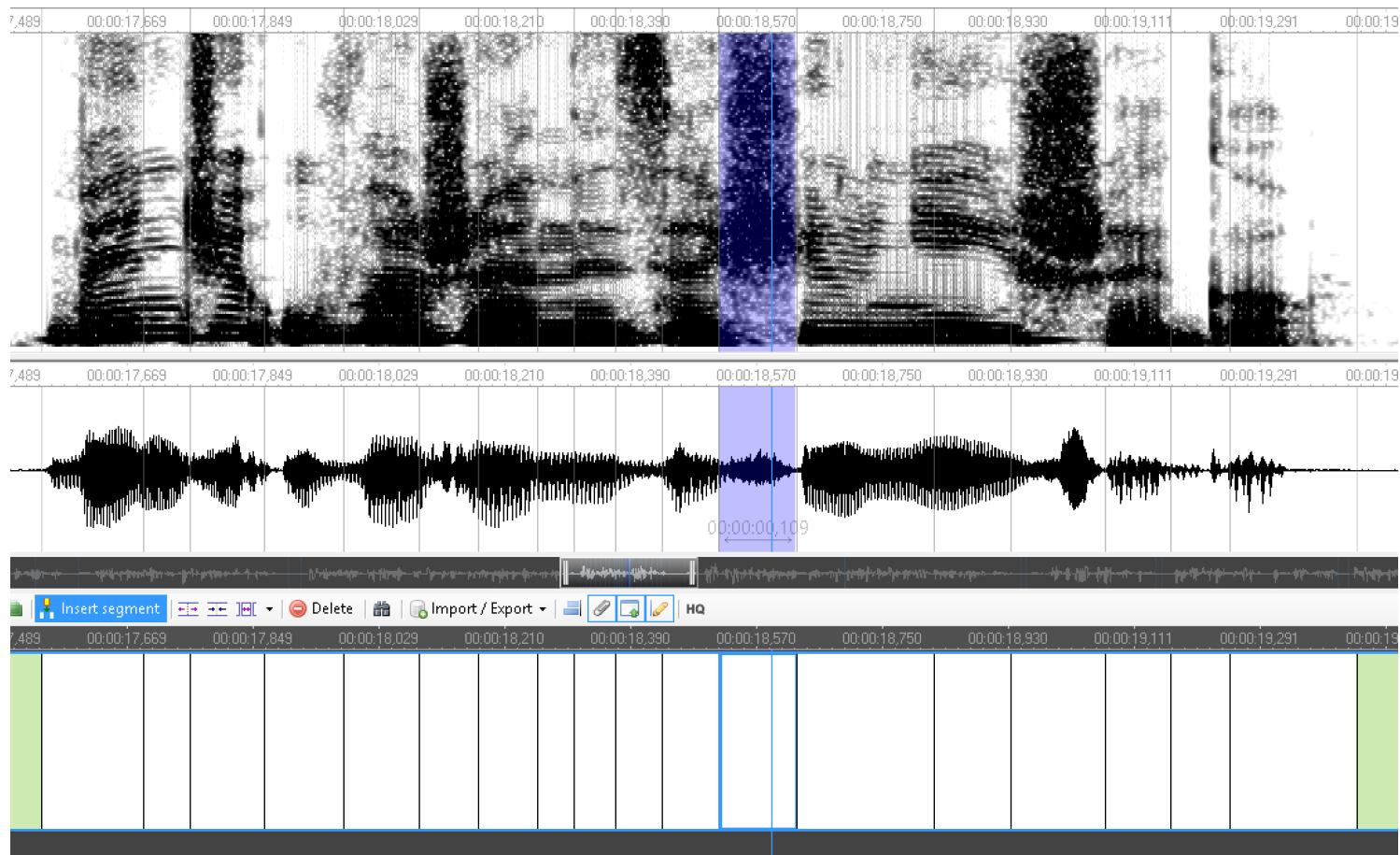




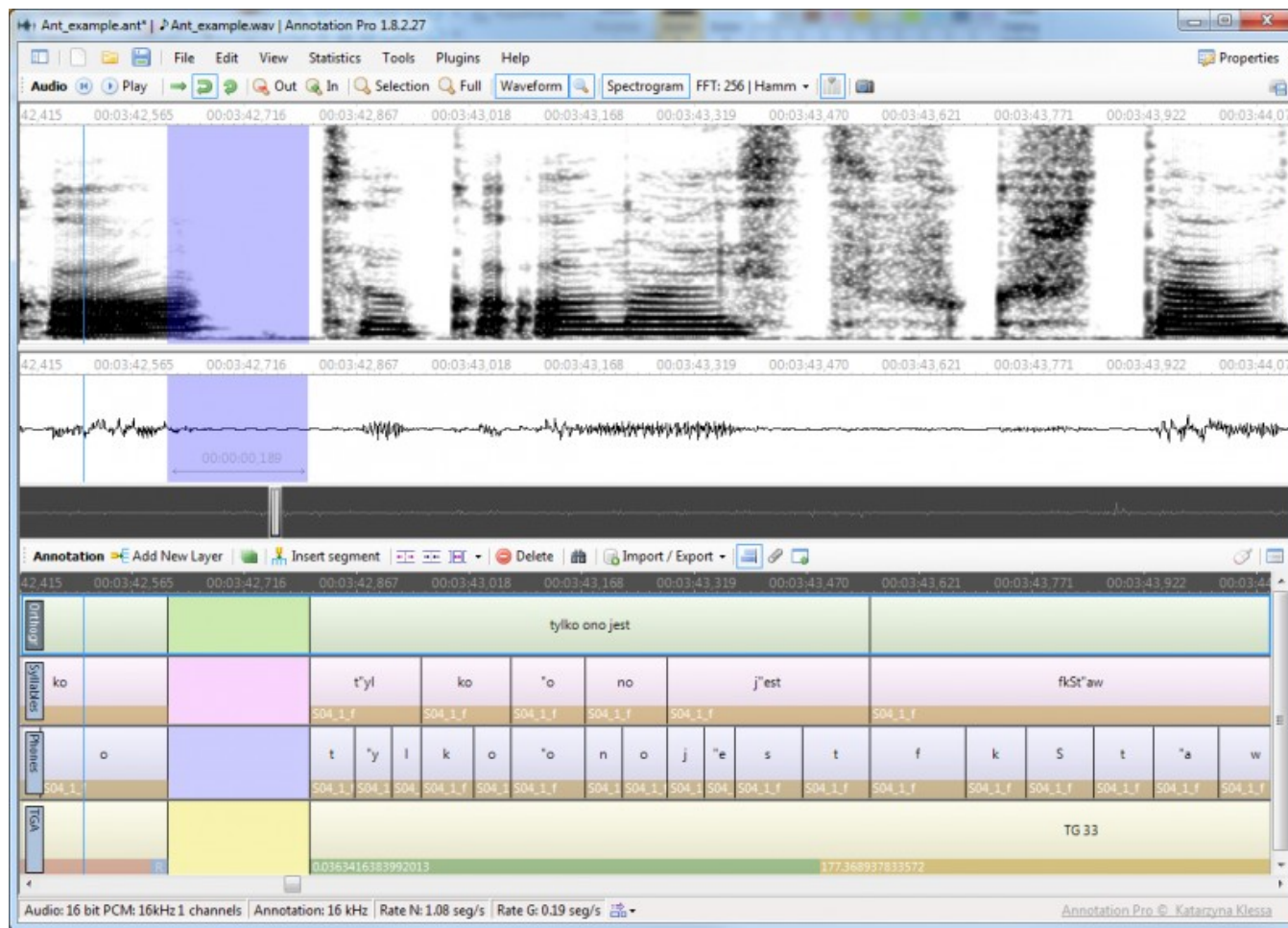
Transkrypcja, segmentacja, anotacja

pe`vnego ,razu || puw`notsni °vjatr i,swontse || spje>tfaliçe || `ktoznix
jest çil,nejfi || `vwaçe pje°xodziw °drogoŵ jaciç >tfwovjek || ovi`penti
°ftçepwi ,pwaftf || `umu ,viliçe °vjents || zetenznix °kturi `pjerfi zmuçi
pfexodzon,tsego || abi`zdjow o`kritçe || `benðze uva°zani zaçilnej,sego ||
puw`notsni °vjatr °zatfow od`razu °doptç s`tsawej ,çiwi || aleim`motspej
>dow || tim`çilnej po°druzni o°tulawçe ,fpwaftf || `vrestçe puw°notsni
°vjatr daw„spokuj || `ftedi °swontse za`tfewo pfi,gzevaç, || af'xfile
'puznej po'druzni zdjow,pwaftf || 'ften >sposup || puw`notsni °vjatr `muçaw
>pfiznatç || ze`swontse jestçil°nejfe od„nego

Transkrypcja, **segmentacja**, anotacja



Transkrypcja, segmentacja, anotacja





Zarządzanie danymi

Wybór technologii – zależny od rodzaju i przeznaczenia korpusu.

Optymalnie jest, gdy sposób systematyzacji danych dobrany jest **już na etapie wstępnego projektowania** tworzonego zasobu.

Różne sposoby systematyzacji korpusów językowych: np.

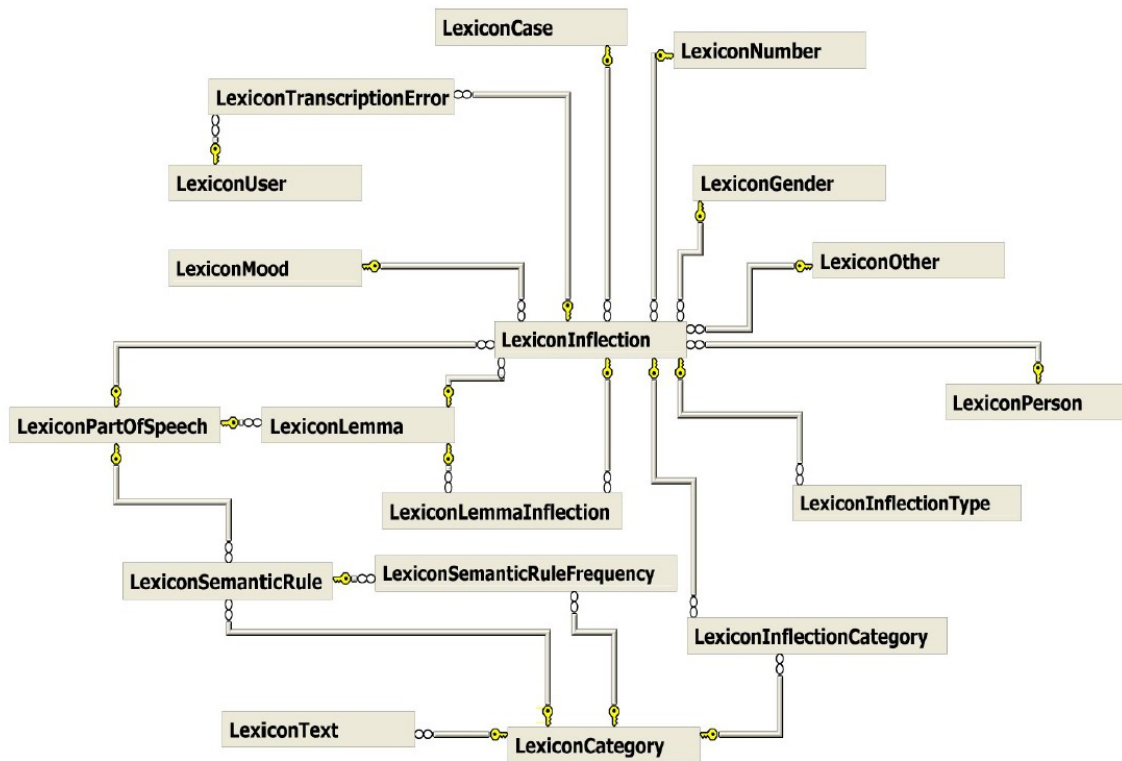
- bazy **plikowe**
- **arkusze** kalkulacyjne
- **relacyjne** bazy danych

» Relacyjna baza danych – zbiór danych zorganizowanych w postaci tabel powiązanych różnego rodzaju relacjami



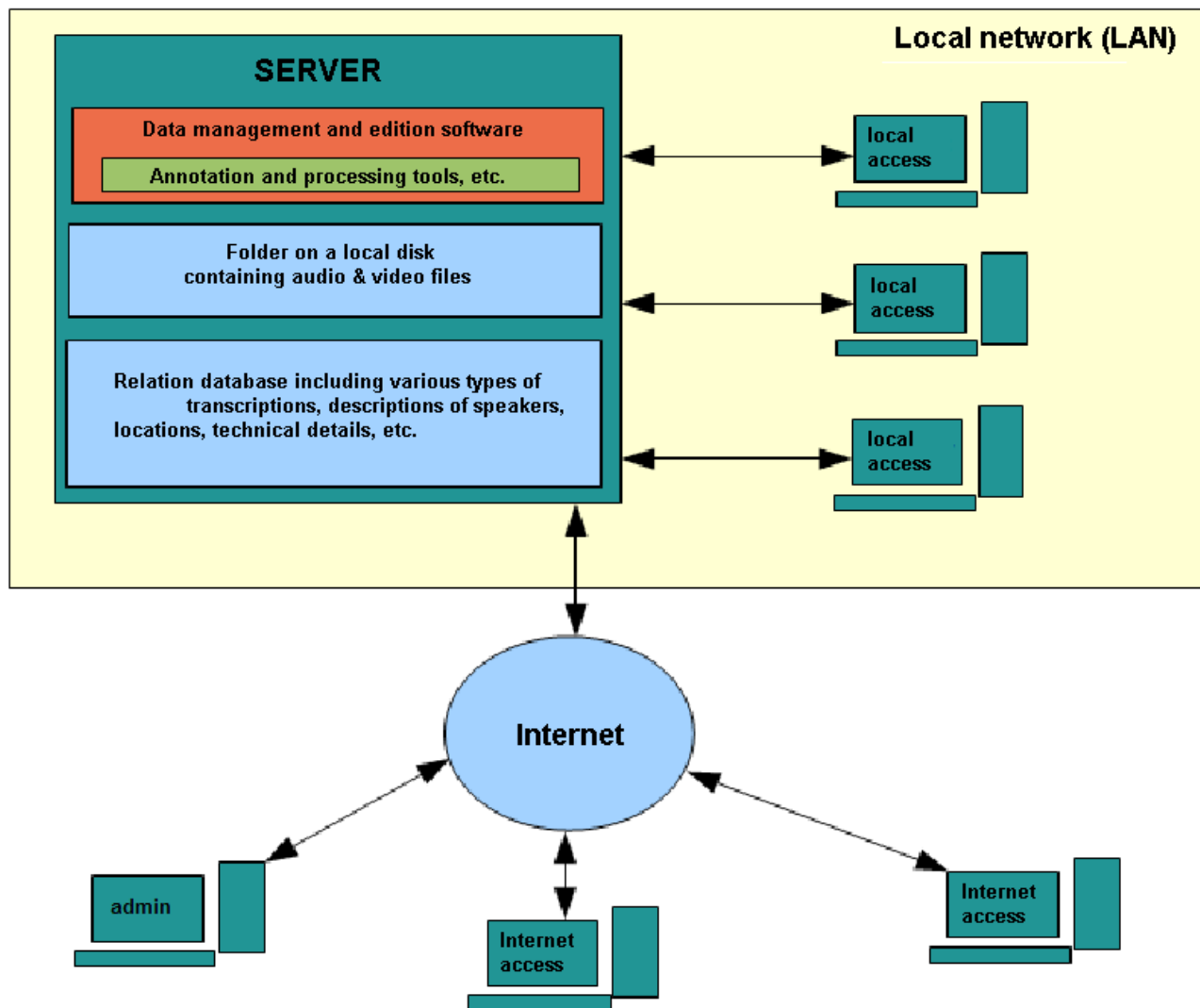
Relacyjna baza danych

- i relacyjny model rzeczywistości



Relacyjna baza danych oparta jest na tzw relacyjnym modelu rzeczywistości: ma odzwierciedlać pewien fragment rzeczywistości, jej elementów oraz powiązań między nimi.

Architektura klient - serwer

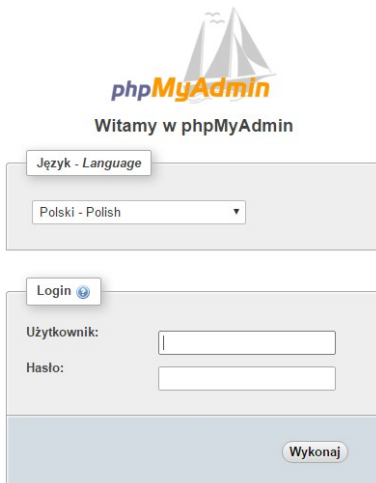




Jak nauczyć się “rozmawiać” z relacyjną bazą danych?

1. Przede wszystkim **trzeba ją mieć** – instalacja *(różne platformy, Microsoft SQL Server, MySQL, Oracle).
2. Warto, a czasem trzeba poznać **język obcy: SQL**
3. Z bazą relacyjną można komunikować się **różnymi kanałami:**

- bezpośrednio zapytania do bazy (w języku SQL)
- za pośrednictwem program(ów) do zarządzania / edycji danych (które wykonają polecenia SQL w naszym imieniu, wówczas nie musimy znać składni tego języka zapytań)
- lokalnie lub przez Internet





Warto - nie warto - trzeba?

Czy / kiedy warto szukać czegoś więcej niż zwykły zbiór plików czy arkusz kalkulacyjny?

- kiedy mamy 1 osobę pracującą nad 1 plikiem?
- kiedy mamy 1 osobę i wiele danych?
- kiedy mamy dużo osób i średnią ilość danych?
- kiedy mamy dużo osób i bardzo dużo danych?
- kiedy różne osoby korzystają z danych z różnych miejsc, zdalnie?



Warto - nie warto - trzeba?

Czy / kiedy warto szukać czegoś więcej niż zwykły zbiór plików czy arkusz kalkulacyjny?

- kiedy mamy 1 osobę pracującą nad 1 plikiem - **nie warto**
- kiedy mamy 1 osobę i wiele danych - **warto**
- kiedy mamy dużo osób i średnią ilość danych - **warto**
- kiedy mamy dużo osób i bardzo dużo danych - **trzeba**
- kiedy różne osoby korzystają z danych z różnych miejsc, zdalnie - **warto** czasem **trzeba** - zależnie od ilości danych (np. za stroną internetową z danymi stoi zwykle baza danych)



wymagania techniczne

językoznawstwo/lingwistyka korpusowa

korpus

językoznawstwo/lingwistyka komputerowa

baza danych

ochrona danych

dane surowe

free vs. open source

dane metadane

„interoperacyjność” / interoperability

wielokrotny użytek / re-usability

udostępnianie

prawa autorskie

współdzielenie

zasady dostępu

prawa majątkowe

kwestie etyczne
społeczne „tabu”

LANGUAGES IN DANGER

EN | PL | DE | HU | NL

home interactive map book of knowledge

teaching materials what can you do about

Search

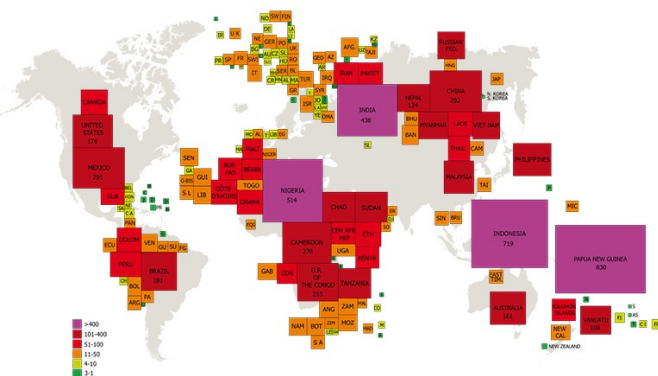
FEATURED LANGUAGE

miyako

Miyako belongs to the Ryukyuan group of languages which together with their "big sister" Japanese make up the Japonic language...

more

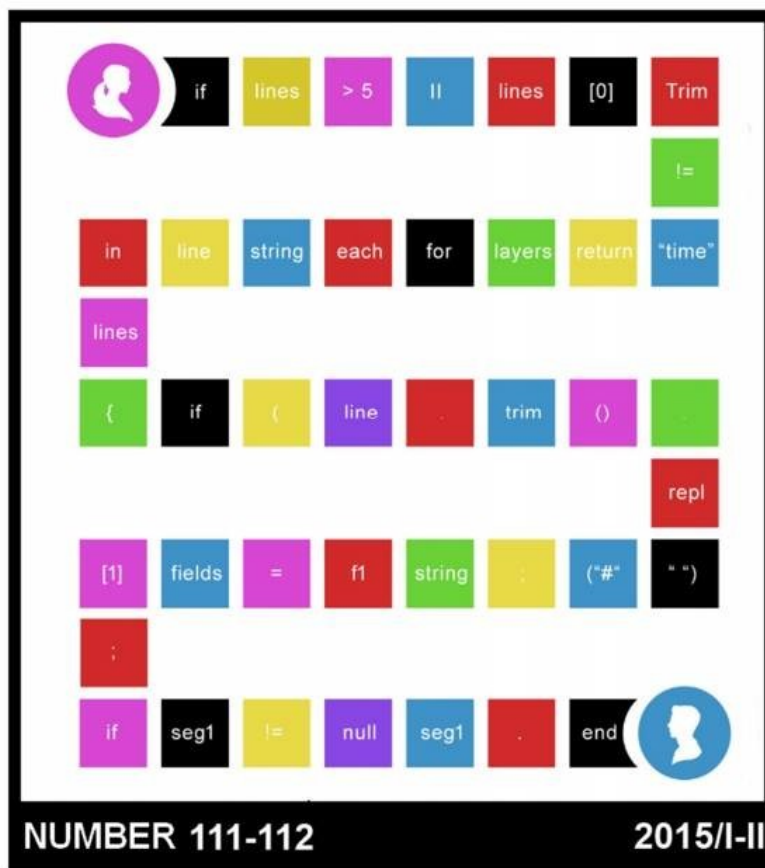
INTERACTIVE MAP





the Phonetician

A Publication of ISPhS/International Society of Phonetic Sciences



Wydanie
poświęcone
narzędziom dla
fonetyki, m.in.
narzędziom do
opisu i analizy
nagrań mowy.



Dziękuję za uwagę!

klessa@amu.edu.pl
annotationpro.org
katarzyna.klessa.pl