



Inżynieria języka z językiem migowym w tle

Nina Suszczańska

Przemysław Szmaj

Politechnika Śląska w Gliwicach

Nina.Suszczańska@polsl.pl

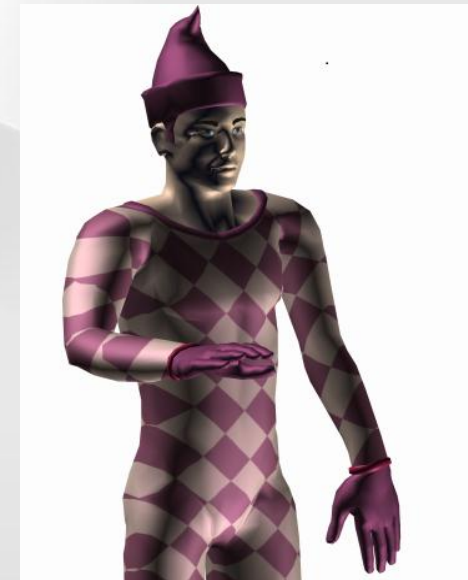
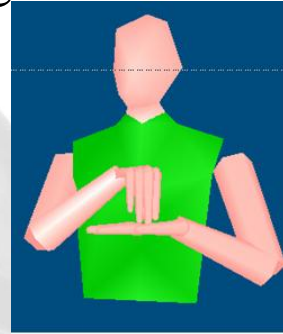
Przemyslaw.Szmal@polsl.pl

Plan

- Inżynieria języka w Gliwicach
 - Trochę historii
 - Thetos
 - LAS
 - Prace badawcze: ANKA, Morf-2, Polmorf-2, Polgener...
- Formalizm SGS# a modelowanie składni i semantyki
 - Cele i założenia badań
 - Formalizm SGS Gładkiego
 - Modyfikacja – SGS#
 - Modelowanie składni
 - Modelowanie semantyki
- Podsumowanie

Trochę historii ●○○○

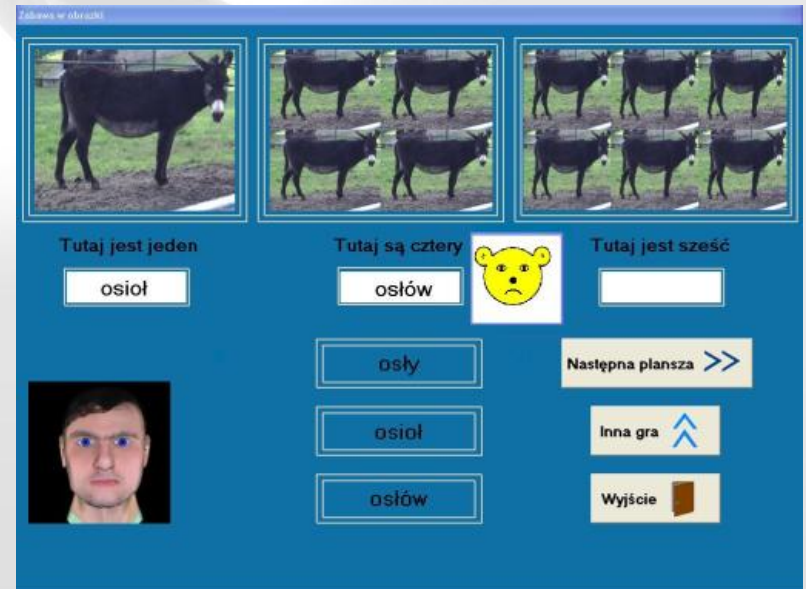
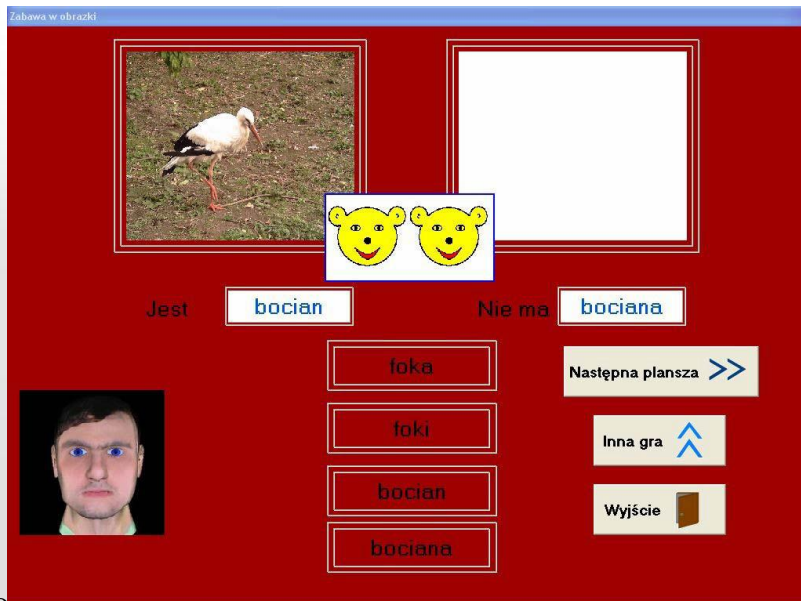
- 1997: Pomysł na system wspomaganie dzieci niesłyszących
 - 1998 – Projekt TGT-1
 - 1999 –2001 Wykonanie



Inżynieria języka w Gliwicach

Trochę historii ○●○○

- 2001: Thetos
- Animacja twarzy i mowy, rozwiązanie prototypowe



Trochę historii



- 2001: Thetos
 - 2003 – 2005 – Thetos-2
 - 2009 – 2011 – Thetos 3



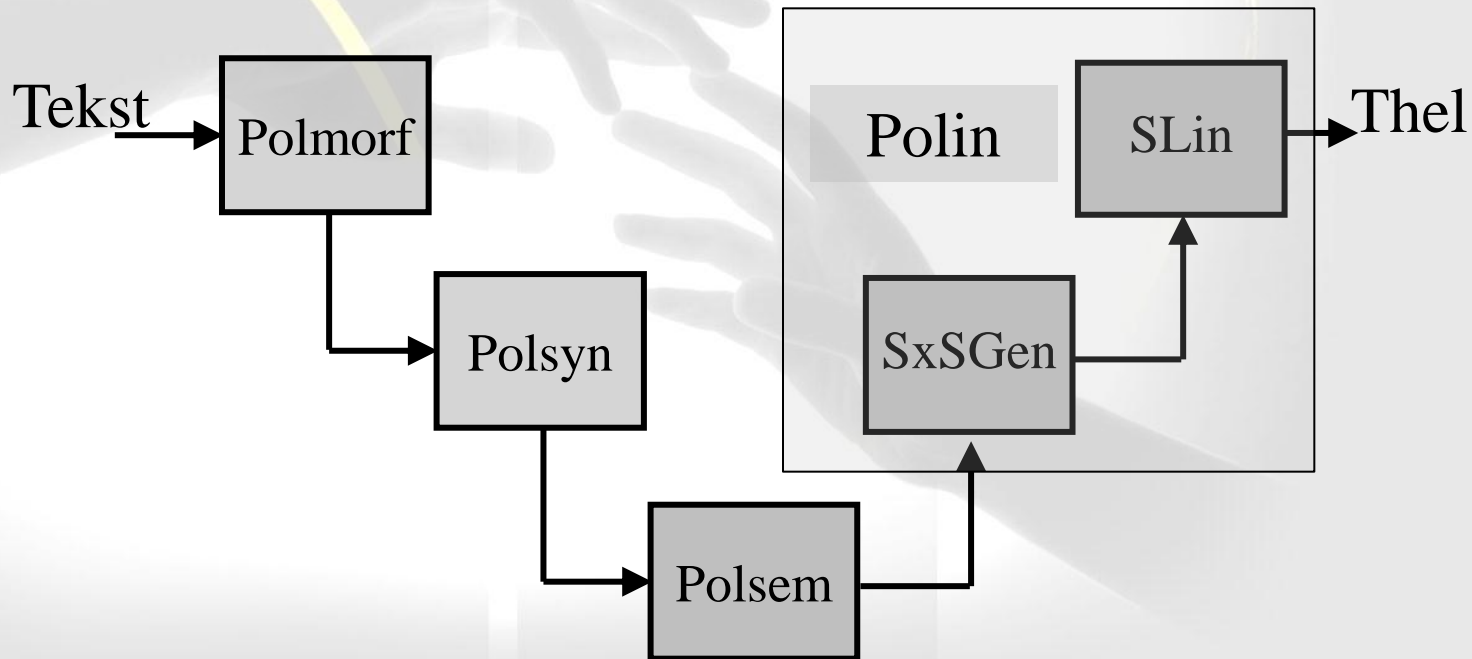
Trochę historii ○○○●

- 1997 – c. d. – Słownik Morf (ponad 90 tys. haseł)
- 1999 – 2003 – Słownik Semsyn
- 2001 – 2009 – Liana
- 2001 – 2009 – Polsumm

Prace badawcze – opracowanie SG-modelu języka:

- 1996 – c. d. – modelowanie składni
- modelowanie aspektów semantyki

● **The Polish Text into Sign Language Translator**



- Polmorf – analizator morfologiczny
- Polsyn – deep parser
- Polsem – analizator semantyczno-syntaktyczny
- Polin – linearyzator wypowiedzi w Thel

● Tłumaczenie w czasie rzeczywistym

● Języki docelowe

➤ SJM

➤ PJM

➤ Język przestrzenny

➤ Z atrybutami

➤ Zapożyczenia

➤ Złożenia

➤ itp.

● Analiza porównawcza prof. Świdzińskiego:

Polski	SJM	PJM
linearny	przestrzenny	przestrzenny
statyczny	statyczny (?)	dynamiczny
wyrażenie: ciąg	wyrażenie: ciąg	wyrażenie: scena
fleksja	fleksja	nie ma fleksji
brak klasyfikatorów	brak klasyfikatorów	są klasyfikatory
idiomatyka	idiomatyka	brak idiomów

● 2009 – 2010 – Thetos w turystyce



● 2009 – 2010 – Thetos w turystyce



- 2010 – 2011 – Thetos w urzędach
 - Urzędy miast Katowice i Chorzów
 - Kioski Informat-e
informat-e.pl



- <https://thetos.aei.polsl.pl>

- Demo



Zastosowania (1) : w kioskach informacyjnych *infomat-e*

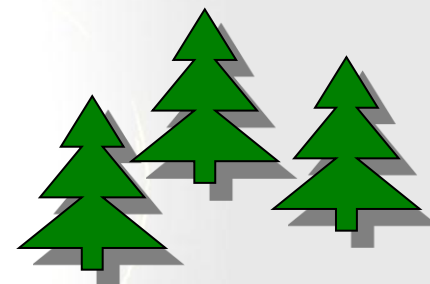


Miejsce instalacji: Urzędy Miasta w Katowicach i w Chorzowie

Automatyczny tłumacz tekstów
na język migowy

Serwer usług lingwistycznych LAS

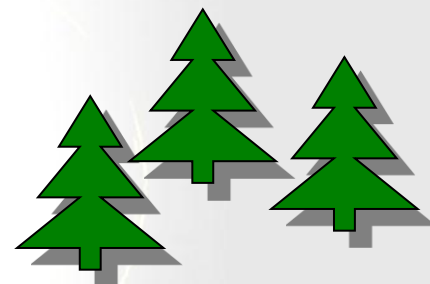
- Linguistic Analysis Server **LAS**
- Usługi bezpłatne
- Użytkownik
 - Anonimowy – ograniczenia
 - Zarejestrowany – pełny dostęp
- Dostęp
 - <http://las.aei.polsl.pl/las2>
 - Webservice, SOAP



Serwer usług lingwistycznych LAS

- Usługi i moduły

- Morfologia – Polmorf – XML, reprezentacja graficzna („Dla użytkownika”)
- Syntaktyka – Polsyn – XML, reprezentacja 3D w opracowaniu
- Semantyka – Polsem – XML, raport przebiegu przetwarzania
- Thetos – Polin – tekst w Thel
- Anafory – Analizator anafor – Demo



- Koło studenckie IPIJ (ipij.aei.polsl.pl), prace dyplomowe
 - ANKA
 - Baza danych Morf-2
 - Analiza i generowanie (Polmorf-2 i Polgener-2)
 - Semantyczna interpretacja SG
 - ◆ NG i VG
 - Wyszukiwanie pojęć podstawowych dla modelu logicznego (projektowanie obiektowe SI): *aktor, metoda, klasa, obiekt.*
 - ◆ PG i PS
 - Określenie *miejsca, czasu oraz sposobu wykonania* czynności
 - Ujednoznacznienie semantyczne (analiza kontekstu)

Cele i założenia badań

- Cel badań – opracowanie informatycznego modelu języka polskiego (SG-modelu) i jego implementacja
- Założenia:
 - Model języka buduje się w celu automatycznego rozpoznawania i przedstawienia treści tekstu
 - Osiągnąć wymagany poziom modelowania można tylko przez powiązanie aspektów semantyki i składni
 - Wymaga opracowania odpowiednich modeli, a dalej – analizatorów
 - Podstawa sukcesu modelowania – trafny wybór formalizmów dla reprezentacji składni, semantyki i znaczenia tekstu
 - Centralne ogniwo – modelowanie składni
 - ◆ Dokładność i elastyczność
 - ◆ Podstawa dla formalizmów dalszych etapów przetwarzania, przede wszystkim analizy semantycznej

Cele i założenia badań

- Słowo *informatyczny* w nazwie modelu:
 - model teoretyczny powinien nadawać się do implementacji
 - model zaimplementowany ma stanowić system oprogramowania
 - implementacja modelu przy przestrzeganiu przyjętych w informatyce praw tworzenia złożonych systemów
 - Modułowość
 - Elastyczność i niezawodność
 - Złożoność pamięciowa i czasowa
 - Sposób organizacji i przepływu danych
 - Przepływ sterowania
- Dwa wątki badań:
 - *informatyka dla inżynierii języka*
 - *inżynieria języka dla informatyki*

Formalizm SGS Formalizm SGS (Systemów Grup Syntaktycznych) Gładkiego

- Formalizm SGS – 1975, prof. Gładki – uogólnienie dwóch formalizmów: drzew zależności i składników bezpośrednich
- Zaletą SGS jest jawne wskazywanie zbiorów słów składających się na grupy składniowe oraz możliwość określenia relacji składniowych między grupami
 - SGS pozwala hierarchizować składowe zdania jak to jest w formalizmie drzew frazowych, a także ustalać związki między składowymi, które są zbliżone do naturalnych, przyjętych w gramatyce tradycyjnej
- Nacisk na naturalność i semantyczność w metodzie wyznaczania
 - Relacji składniowych
 - Struktur składowych, zarówno prostych, jak i złożonych
- SGS jest wręcz idealny dla języków mających swobodny szyk wyrazów w zdaniu

Formalizm SGS Gładkiego

- W pierwotnym formalizmie SGS struktura zdania reprezentowana jest przez graf

$$S = \langle G, r \rangle,$$

gdzie G to niepusty zbiór grup składniowych (SG), a r – relacja zależności wyznaczona na G . Badania właściwości języka są prowadzone przez nałożenie pewnych ograniczeń na topologię grafu

- Niech Z będzie skończonym, uporządkowanym w sposób naturalny zbiorem numerowanych wyrazów w zdaniu i niech będzie zdefiniowana relacja składniowa r na zbiorze Z
 - Symbole $\in, \cup, \cap, \subset, \subseteq$ oznaczają operacje zwykle używane w teorii zbiorów

Formalizm SGS Gładkiego

- *Definicja 1.* Zbiór grup składniowych to zbiór niepustych podzbiorów G
 - w tym Z i wszystkie jednoelementowe podzbiory Z
- *Definicja 2.* Grupy SG jednoelementowe oraz Z nazywane są *grupami trywialnymi*
- *Definicja 3.* Grupa G_1 jest *bezpośrednią* składową grupy G_2 , jeśli i nie istnieje grupa G_3 , dla której byłby spełniony warunek

$$G_1 \subset G_3 \subset G_2$$

- Oczywiście jest fakt, że każda SG inna niż Z jest bezpośrednią składową tylko jednej SG
- *Definicja 4.* Niech grupa G_2 jest zależna składniowo od grupy G_1 . Zależność jest notowana jako $r(G_2, G_1)$ i traktowana jako krawędź grafu S , której początkiem jest G_1 , a końcem – G_2

Formalizm SGS Gładkiego

- *Definicja 5.* Graf $S = \langle G, r \rangle$ jest systemem grup składniowych, jeśli są dla niego spełnione następujące aksjomaty:

Pierwsze dwa ograniczenia dotyczą relacji, następne – elementów zbioru

- *Aksjomat 1.* Jeśli istnieje relacja $r(G_2, G_1)$, to G_1 i G_2 są bezpośrednimi składowymi tej samej SG, wspólnej dla G_1 i G_2
 - Stąd niemożność zależności SG od siebie. To oznacza, że S jest grafem bez pętli
- *Aksjomat 2.* Każda SG ze zbioru G może mieć kilka (zero lub więcej) SG podrzędnych, ale sama może zależeć tylko od jednej SG. Innymi słowy, jeśli występują relacje $r(G_2, G_1)$ oraz $r(G_2, G_3)$, to $G_1 = G_3$
- *Aksjomat 3.* Jeśli $G_1, G_2 \in G$, to $G_1 \cap G_2 = 0$ lub $G_1 \subseteq G_2$ lub $G_2 \subseteq G_1$. Przy tym, jeżeli $G_1 \cap G_2 = 0$, to nie istnieją takie pary elementów $g_{11}, g_{12} \in G_1$ i $g_{21}, g_{22} \in G_2$, że jeden z elementów którejś pary mieści się, a drugi element nie mieści się między elementami pary innej
- *Aksjomat 4.* Jeśli dla $G_1, G_2 \in G$ istnieje relacja $r(G_2, G_1)$, a $G_3 \in G$, to $G_1 \cup G_2$ i G_3 nie zazębiają się
- *Aksjomat 5.* Jeśli $G_1, G_2, G_3, G_4 \in G$ i istnieją relacje $r(G_2, G_1)$ oraz $r(G_4, G_3)$, to $G_1 \cup G_2$ i $G_3 \cup G_4$ nie zazębiają się

Formalizm SGS Gładkiego

- *Aksjomat 6.* Jeśli $G_1, G_2 \in G$ i istnieje relacja $r(G_2, G_1)$ to G_1 nie dzieli G_2 . To znaczy, że w G_1 nie istnieje taki element, który mieściłby się między dwoma dowolnymi elementami G_2
- *Aksjomat 7.* Jeśli $G_1, G_2, G_3 \in G$ i istnieją relacje $r(G_3, G_2)$ oraz $r(G_2, G_1)$ to G_1 nie dzieli $G_2 \cup G_3$
- *Definicja 6.* Graf S jest SGS, jeśli dla niego spełnione są aksjomaty 1-5, jeśli dodatkowo spełnione są aksjomaty 6 i 7, to SGS nazywany *silnym*
- **Arytmetyka na SGS (operacje nie mogą zmieniać formalizmu)**
 - dodać nową SG do SGS
 - usunąć SG z SGS
 - dodać do SGS taką SG, której można przekazać wszystkie zewnętrzne relacje składowych SGS
 - skasować wszystkie zewnętrzne relacje nie zmieniając zbioru
 - wykluczyć z SGS nietrywialną SG razem z jej relacjami itp.

Formalizm SGS Gładkiego

- Podsumowanie
 - SGS przedstawia strukturę syntaktyczną zdania jako skierowany etykietowany graf z krotnymi krawędziami bez pętli, dla którego spełnione są aksjomaty 1-5 lub 1-7
 - Węzłami grafu są grupy składniowe, krawędziami – skierowane związki między nimi
 - Węzły SGS są podgrafami i również tworzą SGS
- W SGS sformułowano cztery ogólne kryteria wyodrębnienia SG spośród elementów zbioru Z:
 - Leksykalny
 - Spójności
 - Operatorowo-kontekstowy
 - Termowo-kontekstowy
- Kryteria te, ze słów Autora SGS, apelują do intuicji Czytelnika, i nie są przeznaczone do konkretnego użycia przy przetwarzaniu komputerowym

Modyfikacja – SGS#

- Udowodniono, że SGS jest formalizmem i bazując na nim może być napisana formalna gramatyka
 - Przykładowa gramatyka napisana nad alfabetem (a,b)
 - Udowodniono także, że graf reprezentacji składniowej bazującej na SGS jest drzewem drzew, czyli *krzakiem*
 - Teoretycznie SGS jest przydatny także do języka naturalnego, w tym języka ze swobodnym szykiem zdania
 - Jednak pierwotna postać SGS okazała się niewystarczająca, żeby wyrazić całą złożoność składni języka naturalnego
- Na potrzeby opracowania gramatyki formalnej dla języka naturalnego formalizm SGS został zmieniony, zmodyfikowany ukierunkowany na praktyczne wykorzystanie
- Aksjomaty Gładkiego, które stanowią bazę formalizmu pozostały bez zmian
- Rozszerzony i uzupełniony formalizm dostał nazwę *SGS#*

Modyfikacja – SGS#

- SGS# jest formalizmem przede wszystkim silnym, inaczej byłby nieprzydatny do formalizacji języka naturalnego
- Graf struktury składniowej SGS# jest etykietowany, jego elementom przypisywane są pewne etykiety odzwierciedlające ich cechy morfologiczne albo składniowe
- SGS# może być reprezentowany przez szóstkę

$$\langle G, r, E_G, E_r, \varphi, \psi \rangle,$$

gdzie $\langle G, r \rangle$ jest systemem grup składniowych w sensie SGS, E_G oraz E_r to skończone zbiory etykiet węzłów i krawędzi, φ i ψ są funkcjami odwzorowania

- Składnia języka polskiego jest na tyle skomplikowana, że jej modelowanie należy podzielić na kilka poziomów
 - Dla każdego poziomu należy opracować odrębny SGS#

Modyfikacja – SGS#

- Cztery podstawowe kryteria wydzielenia grup SGS – uściślone, rozszerzone, a niektórych przypadkach zmienione
 - Kryterium leksykalny – trzy kryteria, przy czym pierwotna zasada wyboru leksykalnego pozostała tylko dla przypadku, gdy słowa w grupie nie zmieniają ani formy, ani swojej kolejności
 - Kryterium spójności – także rozdzielono na trzy, odpowiadające osobnym przypadkom wyboru
 - Operatorowo-kontekstowy pozostał bez zmian
 - Termowo-kontekstowy przeistoczył się w trzy odrębne kryteria
- Dziesięć nowych kryteriów tworzy podstawę opracowania szczegółowych reguł wyodrębnienia grup składniowych
- Kryteria te odegrały zasadniczą rolę także przy sporządzeniu listy typów grup

Modyfikacja – SGS#

- Zmianom uległo także pojęcie zależności składniowej
- Zależność jest traktowana w szerokim zakresie i zgodnie z realiami języka przedstawia zbiór relacji składniowych, niekoniecznie zależnościowych
- Wydzielono pięć poziomów modelowania składni: od zerowego do czwartego
 - Tyleż poziomów mają SG
 - Poziom zerowy odpowiada definicji grupy trywialnej
 - Poziom czwarty jest poziomem funkcyjnym, grupy na tym poziomie noszą miano spójnych
 - Na każdym z poziomów 1 – 4 maksymalną jednostką składniową jest SG tego poziomu, jednostką minimalną jest SG poziomu poprzedniego

Modyfikacja – SGS#

● Podsumowanie

- Zgodnie z formalizmem SGS# tekst (niekoniecznie zdanie) jest przedstawiany jako zbiór etykietowanych (*otagowanych*) *tokenów*
- Jedną z etykiet jest numer porządkowy (*indeks*) tokenu w tekście
- Każde zdanie można przedstawić jako skończony zbiór Z elementów (*token + indeks*), gdzie *indeks* jest liczbą naturalną
- Niepusty podzbiór SG zbioru Z nazywa się grupą składniową, jeżeli spełnia on aksjomaty Gładkiego
 - Aksjomaty definiują strukturę SG wykorzystując szeroko rozumiane pojęcie zależności między elementami zbioru Z i określają ogólne zasady grupowania
- Kryteria \longrightarrow Ogólne reguły \longrightarrow Gramatyka

Modelowanie składni

● Model języka – alfabet i gramatyka

- W przypadku języków naturalnych gramatyka ta powinna być sensowna z punktu widzenia lingwistyki
- Język naturalny jest zbyt skomplikowany, aby można było przedstawić jego składnię w postaci monolitowego modelu

● SG-model składa się z kilku powiązanych ze sobą modeli różnych warstw składni

● SG-model

- Alfabet V – dwie składowe: V_G – grup składniowych i V_R – alfabet relacji

$$V^i = \langle V_G^i, V_R^i \rangle$$

- Gramatyka oparta o SGS# (SG-gramatyka, SGGP)

- Specyfikacja warunków grupowania tokenów zbioru Z w SG
- Wyszczególnienie relacji składniowych powstających między elementami SG

Modelowanie składni

- Rozumienie pojęcia *grupa składniowa* przyjęte w danej pracy mieści się w ramach definicji prof. Świdzińskiego
 - O ile SG ma spełniać wymogi SGS#, to
 - nie może być zbiorem pustym
 - zdanie jest SG
 - nie może mieć wspólnych elementów z inną grupą
 - na zbiorze grup wprowadza się operacje, w tym naturalnego uporządkowania
 - grupa nie może mieć relacji składniowej ze sobą itd. (patrz aksjomaty Gładkiego)
- Na każdym z poziomów modelowania występuje kilka typów SG
 - Dla każdego typu na każdym poziomie istnieją osobne reguły powiązania elementów SG
 - Sposób wiązania elementów w SG, jak i powiązania SG w zdaniu odzwierciedlany w relacjach składniowych. (W SGGP – 38)

Modelowanie składni

- Gramatyka SGGP

- Skończony zbiór produkcji postaci

$$SG_L \rightarrow_{\Omega, \omega, Rel} SG_P$$

gdzie SG_L i SG_P są grupami składniowymi

- Każda produkcja oznacza dość rozległe polecenie

**Podstaw SG_P zamiast SG_L ,
jeżeli zostały spełnione wymogi Ω i warunki ω ,
właściwe danej produkcji,
oraz utwórz podzbiór zbioru Rel tych relacji między składowymi,
które odpowiadają tej produkcji**

- Produkcje SGGP opracowano z myślą o ich algorytmizacji

- Implementacja reguł powinna służyć rzeczywistemu wyodrębnieniu grup składniowych w autentycznych tekstach języka polskiego

Modelowanie składni

- SGGP operuje na pięciu skończonych zbiorach: ***G***, ***I***, ***K***, ***A***, ***R***
 - ***G*** jest zbiorem SG z wprowadzoną na nim relacją częściowego uporządkowania
 - Podział SG na 5 poziomów odpowiada stopniowi złożoności struktury SG oraz typom relacji składniowych zachodzących pomiędzy składowymi SG
 - Grupy, za wyjątkiem poziomu zerowego, składają się rekursywnie z SG różnych poziomów
 - Każdy element zbioru ***G*** jest opisany przez skończony zbiór etykiet (zbiór interpretacji syntaktycznych)
 - Wśród etykiet – identyfikator SG, typ i cechy morfosyntaktyczne, (dziedziczone lub generowane). Reguły atrybowania SG są częścią składową SGGP
 - ***I*** jest zbiorem identyfikatorów SG
 - Każdej nowo zbudowanej SG jest przypisywany identyfikator (ID), unikatowy w obrębie analizowanego tekstu

Modelowanie składni

- SGGP operuje na pięciu skończonych zbiorach: G, I, K, A, R
 - K jest zbiorem typów SG
 - Na zbiór K składają się podzbiory typów SG każdego z poziomów SG
 - Dla każdego typu SG na każdym poziomie istnieje własna SG-gramatyka
 - A jest zbiorem morfosyntaktycznych atrybutów SG
 - Atrybuty, tak samo jak typ i nazwa, należą do interpretacji syntaktycznej SG
 - W regułach wyznaczania cech morfosyntaktycznych każdego typu SG na każdym poziomie grupowania zastosowano elementy teorii gramatyk unifikacyjnych
 - R jest zbiorem potencjalnych relacji składniowych
 - Na każdym poziomie przy grupowaniu SG różnych typów powstają różne relacje syntaktyczne
 - Zbiór R obecnie zawiera trzydzieści osiem relacji

Modelowanie składni

- Na iloczynie kartezjańskim $G \times I \times K \times A \times R$ wprowadza się operacje, za których pomocą odbywa się grupowanie elementów G i ich etykietowanie
 - Żadna z operacji nie może zmienić głównego warunku istnienia SG jako elementu SGS#
- Operatory grupowania
 - Zdefiniowane z góry
 - Niektóre SG, określone przez gramatykę
 - modyfikator *mod*
 - łącznik \oplus
 - łącznik \otimes
- Operacje grupowania na G – rola K i A
 - Ważne są interpretacje morfosyntaktyczne operandów
 - Wykonanie produkcji gramatyki wprost zależy od typów SG_L i SG_p oraz wartości ich atrybutów morfosyntaktycznych

Modelowanie semantyki

- Model = słownik + gramatyka
 - Uwypukła istotne cechy modelowanego obiektu
- Semantyka – bada się znaczenie
 - Słownik – jednostki języka poddawane badaniu
 - Gramatyka – formalizm stosowany do reprezentacji znaczenia tych jednostek
- Ogólne miarkowania
 - W ramach wybranego formalizmu
 - ◆ relacje rozpatrywane między tymi jednostkami
 - ◆ sposób reprezentacji znaczenia
 - ◆ mechanizmy dedukcji – zasady wnioskowania
 - ◆ ograniczenia na badane obiekty
 - nakładane przez zakres modelowania
 - przez pragmatykę badanych tekstów

Modelowanie semantyki

● Założenia

Treść tekstu – zbiór *pojęć* powiązanych ze sobą za pomocą *relacji* pragmatycznych

- Pojęcia mają odpowiadać realnie istniejącym bytom w dziedzinie problemu
 - Relacje są elementami skończonego zbioru relacji zadanych na elementach tejże dziedziny problemu
- Struktura składniowa zdania zależy od jego treści i jest swoistym zapisem algorytmu jej rozpoznawania
- Wniosek: struktura SG i relacje między nimi – mechanizm semantyki języka
- **Relacje składniowe nie odpowiadają relacjom treściowym powstającym między bytami, o które chodzi w zdaniu**
- ## ● W jaki sposób są powiązane te same SG na poziomie składniowym i na poziomie semantycznym?

Modelowanie semantyki

- Hipotezy

- SG to jest składniowa jednostka języka, która służy do przekazywania treści na poziomie semantycznym
- Różne typy SG odgrywają różne role w mechanizmie przekazywania treści
- Różne typy relacji składniowych odgrywają różne role w mechanizmie przekazywania treści

- Poszukiwano

- reguły przejścia od relacji składniowych do relacji semantycznych między elementami zdania (między SG³)
- sposób kodowania semantyki w jednostkach syntaktycznych

Modelowanie semantyki

- SG-model semantyki w stanie raczkującym, ale
 - Wnioski z badań i eksperymentów
 - Nie można zamodelować semantykę za pomocą jakiegoś jednego formalizmu
 - Różnym aspektom semantyki powinny odpowiadać osobne warstwy modelowania
 - Zasady formalizmu SGS# przydatne do modelowania nie tylko składni, ale także semantyki zdania, a ewentualnie – tekstu
- Obecny stan badań
 - Poziom pierwszy
 - Kształtowanie struktury predykatywno-argumentowej zdania
 - Poziom drugi
 - Interpretacja semantyczna SG-struktury predykatywno-argumentowej

Modelowanie semantyki

- Semantyczna struktura zdania w SG-modelu – struktura predykatywno-argumentowa przy wykorzystaniu aparatu formalnego relacji binarnych
- Referencyjnymi argumentami są grupy typu NG, GIM i częściowo PG; argumenty atrybutowe to: PS, MOD oraz reszta PG
- Liczba argumentów i ich typy zależą od własności predykatu i jest to określone przez gramatykę semantyki
 - Opis jednej z gramatyk, wyznaczających wymagania predykatu co do semantycznych cech jego argumentów znajduje się w słowniku syntaktyczno-generatywnym czasowników polskich
 - Jeśli na poziomie składniowym gramatyka ta była nadmiarowa, to na poziomie semantycznym jest po prostu niezbędna
- Pomysł – podobieństwo reprezentacji składniowej i semantycznej: wystarczy nałożyć semantyczny graf na składniowy
 - ◆ określić nowe, semantyczne role SG w szaniu
 - ◆ odnotować nowe, semantyczne relacje

Modelowanie semantyki

- Pospolita konwersja reprezentacji składniowej do reprezentacji semantycznej nie jest wystarczająca do rozumienia treści zdania, tym bardziej tekstu
- Należy opracować gramatykę interpretacji semantycznej
 - Wpływ naturalności i semantyczności formalizmu SGS
- W SG-modelu w reprezentacji predykatywno-argumentowej odzwierciedla się role semantyczne SG *silnie spójnych*
 - Poziom pierwszy – bada się przede wszystkim SG⁴ i SG³ (SG⁰)
 - Na poziomie drugim, w części uzupełniającej strukturę predykatywno-argumentową bada się SG czterech poziomów: od zerowego do trzeciego
- Stosujemy dodatkowo gramatykę stanowiącą adaptację gramatyki funkcjonalnych przypadków Fillmore'a, (ang. *case grammar*)

Modelowanie semantyki

- W celach badawczych gramatyka ta została nieco zmodyfikowana
 - Teoretycznie pod uwagę bierze się 17 relacji predykatu (AKCJI) z innymi składowymi: AKTOR, CO-AKTOR, COUNT-AKTOR, OBIEKT, INSTR, ADRESAT, MIEJSCE, PPOCZ, PKON, CZAS, POCZ-CZAS, KON-CZAS, WYNIK, PRZYCZ, CEL, ŹRÓDŁO, SPOSÓB
 - Realizowano 4 relacji: AKTOR, OBIEKT, INSTR, ADRESAT
 - Pierwsze eksperymenty, próby uzupełnienia: MIEJSCE, CZAS, SPOSÓB
- Opracowano pierwszą wersję gramatyki semantycznej interpretacji
 - Reguły konwersji wybranych relacji w strukturze predykatywno-argumentowej do relacji sytuacyjnych
 - Zasady interpretacji SG różnych typów na wszystkich poziomach składni
 - Zasady interpretacji „wewnętrznych”, składniowych w grupach

Przykład: interpretacja relacji „wewnętrznej” #atr

- Trzy NG: *pelerynka z jedwabiu, pelerynka z kapturkiem, pelerynka z Marszałkowskiej*
- Relacja składniowa #atr(PG,NG_i)
 - Po trywialnym przekształceniu #atr(PG+NG_i,NG) => #atr(NG_i,NG)
 - #atr(jedwab, pelerynka)
 - #atr(kapturek, pelerynka)
 - #atr(Marszałkowska, pelerynka)
- Słowniki (np. ontologia) – cechy semantyczne obiektów, np. *czym jest*
 - pelerynka – *czym jest* – ubranie
 - jedwab – *czym jest* – materiał
 - Marszałkowska – *czym jest* – ulica
 - ulica – *czym jest* – część miasta
 - miasto – *czym jest* – miejsce
 - kaptur – *czym jest* – część ubrania
- **Inne przykłady i Demo – w kularach**

Zamiast podsumowania – Część trzecia

- Ostatni krok modelowania – opracowanie całościowego modelu *spójnego tekstu*
 - ▶ *spójny tekst* – tekst, zdania którego stanowią powiązaną z sobą semantycznie i strukturalnie całość, a nie luźny zbiór, i służą do przekazywania wspólnej treści
- Bada się sposób łączenia zdań w spójnym tekście
- Modelowanie struktury tekstu i relacji między zdaniami wymaga opracowania między innymi algorytmów analizy i formalnej reprezentacji aspektów kompresji językowych, takich jak anafory i elipsy

A hand with a yellow measuring tape is positioned at the top left, and another hand is reaching out from the right. The word 'Kornico' is written in large, 3D, white letters with black outlines and shadows. The letters are set against a background of a hand with a measuring tape and another hand reaching out.

Kornico