

Korpusowe badania składniowe

Magdalena Derwojedowa
derwojed@uw.edu.pl

Sesja korpusowa
16 lipca 2009 r.

- 1 Wstęp: korpus tekstów
- 2 A. Wierzbicka, *System składniowy prozy polskiego renesansu*
- 3 M. Świdziński, *Własności składniowe wypowiedników polskich*
- 4 M. Derwojedowa, *Porządek linearny składników zdania elementarnego*
- 5 W. Kieraś: Eksperyment z ramami interpretacyjnymi
- 6 RAMKI — zależności głębinowoskładniowe



Wstęp: korpus tekstów

Definicja

Korpus tekstów to zbiór książek, czasopism, artykułów itp. przeznaczony do jakichś prac lub badań. Termin specjalistyczny.

(ISJP)

A. Wierzbicka, *System składniowy prozy polskiego renesansu*

- gramatyka zależnościowa (Tesnière 1959);
- badania konotacyjnych relacji między wyrazami:
 - stopień konotacji;
 - kierunek konotacji;
- badania szyku:
 - konfiguracyjność (projekcyjność; por. Lecerf 1960, Fitiatow 1962 
 - ciągłość składników bezpośrednich;
 - szyk progresywny i antycypacyjny (por. Bally 1965, §327–328);
- głębokość konstrukcji 
- opis morfologiczny (adaptowany z Mel'čuk 1974).

M. Świdziński, *Własności składniowe wypowiedników polskich*

- Grant 1 P104 030 04 *Ukierunkowana gramatycznie tekstowa baza danych: korpus wypowiedzeń współczesnej polszczyzny pisanej 1993–1996* IJP UW;
- empiryczna weryfikacja *Gramatyki formalnej języka polskiego*;
- korpus *Słownika frekwencyjnego polszczyzny współczesnej (1963–1967/1974–1977/1991; por. SFPW)*;
- opis 1000 próbek korpusu, 50 000 słów, 9000 wypowiedników.

Korpus *SFPW*

- 5 stylów (popularnonaukowy, publicystyczny, dramat artystyczny, proza artystyczna, drobne wiadomości prasowe);
- próbki gronowe (50 słów);
- fiszki dostępne na stronie <http://www.mimuw.edu.pl/polszczyzna/pl196x/fiszki/index.htm>;
- przykładowa fiszka:

A0001 ŻL 1963/5 str. 3 koL. 4

Sztuka utraciła swoją²⁴¹ moc¹⁴¹ pobudzającą²⁴¹; przykrym²⁵¹ widowiskiem staje⁵⁰¹ się koncert¹¹¹ wybitnej²²¹ niegdyś śpiewaczki¹²¹ i nie usłwienią go⁴² nawet nigdy dotąd nie wykonywane²¹² utwory¹¹² genialnego²²¹ kompozytora¹²¹. Słuchacze myślą⁵ o⁶⁶ czym⁴⁶ innym²⁶¹, rozmawiają o⁶⁶ czym⁴⁶ innym²⁶¹ i śpiewaczka, postarzawszy⁵⁰¹ się, głos¹⁴¹ straciła. Nie było dotąd takich²²² koncertów u⁶² Iwaszkiewicza¹²¹.


Gramatyka formalna Świdzińskiego

- gramatyka składników bezpośrednich;
- skracająca;
- w zasadzie bezkontekstowa;
- wyrażona w Prologu.

Baza wypowiedników

```
STYL: DR
TEKST: *121*, 3-1 (W tobie) <jest> /teraz/ \już\ [tylko jedno uczucie].
TYP: Z
CEN: jest
HAS: być
KL: V
ASP: i
CHAR: 3pe,te
SCH: 15
OPIS: PS
DL: 8
VF: 1
SU: 3
TPSU:
OB1: 2
OB2: 0
LU1: 1
LU2: 0
IN: 1
TPI: juz
SZYK: OVS
```


M. Derwojedowa, *Porządek linearny składników zdania elementarnego*

- badanie porządku linearnego składników zdania;
- badanie nieciągłości konstrukcji składniowej ;
- baza składniowa z badania M. Świdzińskiego.

W. Kieraś, Eksperyment z ramami interpretacyjnymi (wymagania czasowników związanych z piłką nożną)

- koncepcja ram interpretacyjnych (Fillmore 1982);
- korpus ze źródeł internetowych o objętości 500 tys. segmentów otagowany za pomocą taggera regułowego Ł. Dębowskiego;
- zestaw scenariuszy, ról i ram z *Kicktionary* (http://www.kicktionary.de/index_de.html);
- znakowanie za pomocą ról semantycznych i powierzchniowskładniowe;
- możliwy do przeglądania w przeglądarce Poliqarp (<http://korpus.pl/index.php?page=poliqarp>).

Przykład: znakowanie składniowe (STRZELAĆ I)

STRZELIĆ — STRZELAĆ II

- (207) {Piłkarze Górnika}_{sub} od razu zaatakowali i (w dwie minuty)_{Temp}
strzelili_{sub} [dwa gole]_{acc}.
- (208) W pierwszej części pierwszej połowy dłużej przy piłce utrzymywała się {Legia}_{sub} i dzięki temu **strzelila**_{sub} [bramkę]_{acc}.
- (209) [Korona]_{sub} **strzelila** [szczęśliwą bramkę]_{acc} (zaraz na początku meczu)_{Temp} i (tuż przed przerwą)_{Temp}, (z rzutu karnego)_{Akc}.

STRZELIĆ/STRZELAĆ II oznacza tyle, co 'oddać strzał (kopnąć piłkę) w kierunku bramki'.

Schemat: V-0 (Mod, Adl, Abl, Temp)

Scenariusz: 5.4

Rama: 5.4.3

Przykład: znakowanie semantyczne (STRZELAĆ II)

5.4.3. Rama: STRZAŁ

W skład ramy **strzał** wchodzi następujące czasowniki: STRZELIĆ/STRZELAĆ II, LOBOWAĆ/PRZELOBOWAĆ, GLÓWKOWAĆ.

STRZELIĆ/STRZELAĆ II

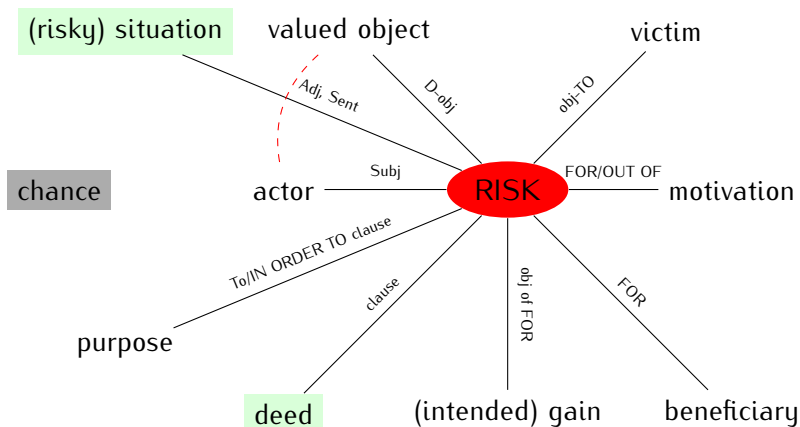
- (86) Piłka spadła pod nogi [Griszczenki]_{STRZELAJĄCY}, który **strzelił** mocno [z piętnastu metrów]_{ŹRÓDŁO}, ale niecelnie.
- (87) A [Szymkowiak]_{STRZELAJĄCY} prawie **strzelił** [na bramkę]_{CEL}.
- (88) *Napastnik Korony po przebiegnięciu kilkunastu metrów wrzucił futbolówkę w pole karne, ale [Hermes]_{STRZELAJĄCY} **strzelił** [obok słupka]_{CEL}.*
- (89) Po rzucie różnym [Karwan]_{STRZELAJĄCY} niecelnie **strzelił** [głowa]_{CZĘŚĆ_CIAŁA}.
- (90) Edi wyklada piłkę [Williamowi]_{STRZELAJĄCY}, a ten **strzelił** zbyt lekko, by zaskoczyć Pilarza.
- (91) [Bilski]_{STRZELAJĄCY} **strzelał** mocno [z dwudziestu metrów]_{ŹRÓDŁO}.
- (92) [Mendez]_{STRZELAJĄCY} **strzela** [na bramkę]_{CEL} [z bliskiej odległości]_{ODLEGŁOŚĆ}.

frazy wymagane	frazy luźne
SUB ← STRZELAJĄCY	Abl ← ŹRÓDŁO
	Abl ← ODLEGŁOŚĆ
	Adl ← CEL
	Mod ← CZĘŚĆ_CIAŁA

RAMKI – zależności głębinowskładniowe

- grant MNiSW N104 024 32/1840;
- 200 jednostek czasownikowych;
- klasyfikacja czasowników wg własności składniowych, semantycznych (i pragmatycznych):
 - powierzchniowe własności walencyjne;
 - spis ról realizowanych przez poszczególne elementy ramy;
 - schematyczny opis sytuacji (scenariusz);
 - oznakowane przykład z korpusu;
- <http://www.ramki.uw.edu.pl/>.

Przykład: RISK (Fillmore i Atkins 1992)



Lista ram

[<< Poprzednia rama](#) [Następna rama >>](#)

Rama: Awareness

Jednostki w ramie:

[myśleć](#)
[oczekiwać](#)
[rozumieć](#)
[rozumieć](#)
[rozumieć](#)
[wiedzieć](#)
[wierzyć](#)

Przykładowa analiza

rozumieć

rama: [Awareness](#)

głosa: znamy przyczynę/itotę czegoś, np. mnie rozumiem twojej decyzji

schematy:

1.1 (Acc)

1.7 (że)

Przykłady:

dofinansowanie helikoptera dla zakopiańskiego TOPR. (WAK) * **RozumiemVerb**, że_{Sent} **Content** helikopter jest sprawą niezwyklej wagi, niemniej jeśli na wrzesień – mówi Wojciech Bloch wiceburmistrz Olkusza. – **RozumiemVerb** nauczycieli **Obj0** **Content** i popieram ich oburzenie. Ale to nie gmina ciężarem. Wokulski_{Subj} **Cognizer** widząc to sam spotniał. "Już_{Lu0} **Time** **rozumiećVerb** – pomyślał – tajemnicę **Obj0** **Content** Geista. On mnie zamagnetyzował.

Scenariusz: COGNIZER ma fragment CONTENT w swoim modelu świata. CONTENT nie powstaje konieczniej w wyniku bezpośredniej percepcji, ale zwykle poprzez dedukcję na postawie postrzeżeń. W niektórych przypadkach dedukcja CONTENT nie wprost jest oparta na źródłach informacji, w innych na logice, a w innych – źródło wiedzy nie jest profilowane.

[<< Poprzednia rama](#) [Następna rama >>](#) [<< Poprzedni czasownik](#) [Następny czasownik >>](#)

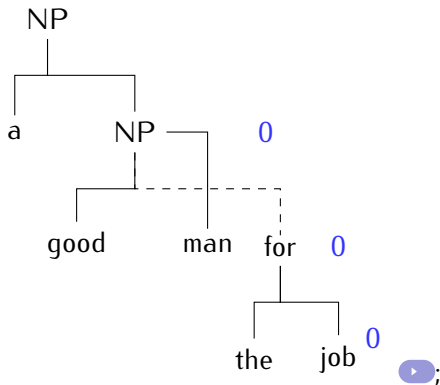
Literatura I

- [Bally 1965] Ch. Bally, *Linguistique générale et linguistique française*, wyd. IV, Berne 1965.
- [Derwojedowa 2000] M. Derwojedowa, *Porządek linearny składników zdania elementarnego*, Elipsa 2000.
- [Fillmore 1982] Ch. J. Fillmore, *Frame semantics*, [w:] *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing Co., s. 111–137.
- [Fillmore i Atkins 1992] Ch. J. Fillmore i B. T. S. Atkins, *Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors*, [w:] A. Lehrer i E. Kittay (red.) *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*, Hillsdale: Lawrence Erlbaum Associates, s. 75–102.
- [Fitiatow 1962] S. J. Fitiatow, *О моделировании синтакса в структурной лингвистике*, Проблемы структурной лингвистики, Moskwa, 1962.
- [Lecerf 1960] Y. Lecerf, *Programme des conflits, modèle des conflits*, „La Traduction Automatique”, 4–5.

Literatura II

- [Mel'čuk 1974] I. A. Mel'čuk *Опыт теории лингвистических моделей смысл⇔ текст*, Moskwa 1974.
- [SFPW] I. Kurcz, A. M. Lewicki, J. Sambor, K. Szafran i J. Woronczak, *Słownik frekwencyjny polszczyzny współczesnej*, Kraków 1990.
- [Świdziński 1996] M. Świdziński, *Gramatyka formalna języka polskiego*, Warszawa.
- [Świdziński 1996] M. Świdziński, *Własności składniowe wypowiedników polskich*, Warszawa.
- [Tesnière 1959] L. Tesnière, *Éléments de syntaxe structurale*, Paris 1959.
- [Wierzbicka 1972] A. Wierzbicka, *System składniowy prozy polskiego renesansu*, t. 5 serii *Historia i teoria literatury. Studia. Teoria literatury*, Warszawa, 1972.

Głębokość konstrukcji



Konfiguracyjność (projekcyjność)

Definicja

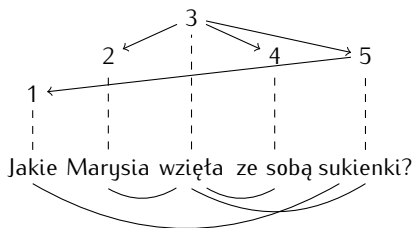
Konfiguracyjność (projekcyjność) drzewo zależności

$T = (V, \rightarrow, \preceq)$

jest projekcyjne, jeżeli

$$j_1, j_2 \in \text{Subtree}_i \ \& \ \nu \in (j_1, j_2) \rightarrow \nu \in \text{Subtree}_i$$

Konfiguracyjność (projekcyjność), cd.



Nieciągłość

