

# What to acquire in automatic valence acquisition

Adam Przepiórkowski

Linguistic Engineering Group  
Institute of Computer Science PAS  
<http://nlp.ipipan.waw.pl/>

Regensburg–Warsaw workshop at ICS PAS  
19 April 2006

# Two stages of valence acquisition

# Two stages of valence acquisition

Stage 1: Collecting evidence from corpora.

# Two stages of valence acquisition

Stage 1: Collecting evidence from corpora.

Stage 2: Statistical inference on the collected data.

# Two stages of valence acquisition

Stage 1: Collecting evidence from corpora.

Stage 2: Statistical inference on the collected data.

Evidence collected in Stage 1:

- co-occurrence of verbs and
- phrases, i.e., potential arguments.

# Two stages of valence acquisition

Stage 1: Collecting evidence from corpora.

Stage 2: Statistical inference on the collected data.

Evidence collected in Stage 1:

- co-occurrence of verbs and
- phrases, i.e., potential arguments.

Aim of this talk:

# Two stages of valence acquisition

Stage 1: Collecting evidence from corpora.

Stage 2: Statistical inference on the collected data.

Evidence collected in Stage 1:

- co-occurrence of verbs and
- phrases, i.e., potential arguments.

Aim of this talk: define the Output of Stage 1 = Input of Stage 2.

# Implicit linguistic knowledge

- DAC 'give': <NP[NOM], NP[ACC], NP[DAT]>



# Implicit linguistic knowledge

- DAĆ 'give': <NP[NOM], NP[ACC], NP[DAT]>
- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'

# Implicit linguistic knowledge

- DAĆ 'give': <NP[NOM], NP[ACC], NP[DAT]>
- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'

Implicit:

- NPs are actually NPs, NumPs and (certain kinds of) AdjPs

# Implicit linguistic knowledge

- DAĆ 'give': <NP[NOM], NP[ACC], NP[DAT]>
- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'

Implicit:

- NPs are actually NPs, NumPs and (certain kinds of) AdjPs,
- pro-drop

# Implicit linguistic knowledge

- DAĆ 'give': <NP[NOM], NP[ACC], NP[DAT]>
- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'

Implicit:

- NPs are actually NPs, NumPs and (certain kinds of) AdjPs,
- pro-drop,
- genitive of negation.



# Information difficult to learn from corpora

1. Non-morphosyntactically defined arguments; e.g. Polański 1992:
  - locative NP,
  - temporal-durative NP,
  - conditional NP (*pod warunkiem dotarczenia materiału, przy dobrej pogodzie*), etc.

# Information difficult to learn from corpora

1. Non-morphosyntactically defined arguments; e.g. Polański 1992:

- locative NP,
- temporal-durative NP,
- conditional NP (*pod warunkiem dotarczenia materiału, przy dobrej pogodzie*), etc.

2. Non-exclusive (conjoinable) alternative realisations of an argument, e.g. (Kallas, 1993, p. 103):

- Udawał naiwnego i że nic nie pretended-SG.M123 naïve-SG.M12 and that nothing-ACC not rozumie.

understand-SG.TER

'He pretended to be naïve and to understand nothing.'

- UDAWAĆ 'pretend': <NP[NOM], {NP[ACC]; S[ŻE]}>

# Information difficult to learn from corpora (contd.)

## 3. Embedded questions vs. free relatives, e.g.:

- Wygrał, kto przyszedł pierwszy.  
won-M123 who came first  
'Whoever came first, won.'
- Pytał, kto przyszedł pierwszy.  
asked-M123 who came first  
'He asked who came first.'

# Information difficult to learn from corpora (contd.)

## 4. Object control vs. subject control vs. raising:

- Janek kazał (Marysi) przyjść.  
Janek ordered Marysia come  
'Janek ordered Marysia to come.'
- Janek obiecał (Marysi) przyjść.  
Janek promised Marysia come  
'Janek promised Marysia to come.'
- Janek mógł przyjść.  
Janek could come  
'Janek could come.'



# Information difficult to learn from corpora (contd.)

## 5. Other difficulties:

- oratio recta
- agreement between two arguments
- non-subject nominative, etc.

# Information difficult to learn from corpora (contd.)

## 5. Other difficulties:

- oratio recta
- agreement between two arguments
- non-subject nominative, etc.

The question: what output of Stage 1 will preserve (just about) enough morphosyntactic information from corpora to make it possible to learn such valence information?

# Information difficult to learn from corpora (contd.)

## 5. Other difficulties:

- oratio recta
- agreement between two arguments
- non-subject nominative, etc.

The question: what output of Stage 1 will preserve (just about) enough morphosyntactic information from corpora to make it possible to learn such valence information?

For example, more than case value is necessary to discover agreement or non-subject nominative.

# Verbal information

It is not sufficient to record the lemma.

# Verbal information

It is not sufficient to record the lemma. Also needed:

- polarity information:
  - negation may change valence,
  - inherently negative / affirmative verbs

# Verbal information

It is not sufficient to record the lemma. Also needed:

- polarity information:
  - negation may change valence,
  - inherently negative / affirmative verbs,
- personal / impersonal:
  - any nominative phrase with an impersonal form indicates a non-subject nominative argument

# Verbal information

It is not sufficient to record the lemma. Also needed:

- polarity information:
  - negation may change valence,
  - inherently negative / affirmative verbs,
- personal / impersonal:
  - any nominative phrase with an impersonal form indicates a non-subject nominative argument,
- agreement features of personal verbs:
  - detection of non-subject nominative arguments,
  - detection of agreement between (unrealised) subject and another argument.

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP            nominal phrase



# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase
PrepNP	prepositional-nominal phrase
PrepNumP	prepositional-numeral phrase
PrepAdjP	prepositional-adjectival phrase

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase
PrepNP	prepositional-nominal phrase
PrepNumP	prepositional-numeral phrase
PrepAdjP	prepositional-adjectival phrase
AdvP	adverbial phrase

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase
PrepNP	prepositional-nominal phrase
PrepNumP	prepositional-numeral phrase
PrepAdjP	prepositional-adjectival phrase
AdvP	adverbial phrase
InfP	infinitival phrase

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase
PrepNP	prepositional-nominal phrase
PrepNumP	prepositional-numeral phrase
PrepAdjP	prepositional-adjectival phrase
AdvP	adverbial phrase
InfP	infinitival phrase
CP	sentential phrase introduced by a complementiser

# Basic argument types

On the basis of Świdziński 1998, Saloni and Świdziński 1998 and Polański 1992:

NP	nominal phrase
NumP	numeral phrase
AdjP	adjectival phrase
PrepNP	prepositional-nominal phrase
PrepNumP	prepositional-numeral phrase
PrepAdjP	prepositional-adjectival phrase
AdvP	adverbial phrase
InfP	infinitival phrase
CP	sentential phrase introduced by a complementiser
KP	sentential phrase introduced by an interrogative or relative pronoun



# Possible extensions of basic argument types



# Possible extensions of basic argument types

- Elective adjectival phrases

# Possible extensions of basic argument types

- Elective adjectival phrases,
- syntactic numerals (i.e., those occurring with the non-agreeing SG.N.TER verbal form when in subject position)

# Possible extensions of basic argument types

- Elective adjectival phrases,
- syntactic numerals (i.e., those occurring with the non-agreeing SG.N.TER verbal form when in subject position),
- oratio recta

# Possible extensions of basic argument types

- Elective adjectival phrases,
- syntactic numerals (i.e., those occurring with the non-agreeing SG.N.TER verbal form when in subject position),
- oratio recta,
- the reflexive marker *się*.

# Morphosyntactic information

case

NP, NumP, AdjP,  
PrepNP, PrepNumP, PrepAdjP

# Morphosyntactic information

case	NP, NumP, AdjP, PrepNP, PrepNumP, PrepAdjP
number, gender	NP, NumP, AdjP, PrepAdjP

# Morphosyntactic information

case	NP, NumP, AdjP, PrepNP, PrepNumP, PrepAdjP
number, gender	NP, NumP, AdjP, PrepAdjP
person	NP[NOM]

# Morphosyntactic information

case	NP, NumP, AdjP, PrepNP, PrepNumP, PrepAdjP
number, gender	NP, NumP, AdjP, PrepAdjP
person	NP[NOM]
aspect	InfP



# Lexical information

preposition	PrepNP, PrepNumP, PrepAdjP
main noun	NP, NumP, PrepNP, PrepNumP
main adjective	AdjP, PrepAdjP
main adverb	AdvP
complementiser	CP
wh-word	KP

# Lexical information

preposition	PrepNP, PrepNumP, PrepAdjP
main noun	NP, NumP, PrepNP, PrepNumP
main adjective	AdjP, PrepAdjP
main adverb	AdvP
complementiser	CP
wh-word	KP

Wh-words instead of just a single wh-word?

# Lexical information

preposition	PrepNP, PrepNumP, PrepAdjP
main noun	NP, NumP, PrepNP, PrepNumP
main adjective	AdjP, PrepAdjP
main adverb	AdvP
complementiser	CP
wh-word	KP

Wh-words instead of just a single wh-word?

Also degree in case of AdjP and PrepAdjP?

# Examples

- Nie daliśmy tym siedmiu marynarzom  
 not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
 najlepszej z naszych łodzi.  
 [best-GEN of our boats]  
 'We have not given those seven sailors the best of our boats.'

# Examples

- Nie daliśmy tym siedmiu marynarzom  
 not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
 najlepszej z naszych łodzi.  
 [best-GEN of our boats]  
 'We have not given those seven sailors the best of our boats.'
- DAĆ.NEG.FIN.PL.M1.PRI:  
 <NumP[MARYNARZ.DAT], AdjP[DOBRY.GEN.SG.F]>

# Examples

- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'
- DAĆ.NEG.FIN.PL.M1.PRI:  
<NumP[MARYNARZ.DAT], AdjP[DOBRY.GEN.SG.F]>

Should be like this?

- DAĆ.NEG.FIN.PL.M1.PRI:  
<NumP[MARYNARZ.DAT], AdjP[ŁÓDŹ.GEN.SG.F]>

# Examples

- Nie daliśmy tym siedmiu marynarzom  
not gave-PAST.PL.M1.PRI [these-DAT seven-DAT sailors-DAT]  
najlepszej z naszych łodzi.  
[best-GEN of our boats]  
'We have not given those seven sailors the best of our boats.'
- DAĆ.NEG.FIN.PL.M1.PRI:  
<NumP[MARYNARZ.DAT], AdjP[DOBRY.GEN.SG.F]>

Should be like this?

- DAĆ.NEG.FIN.PL.M1.PRI:  
<NumP[MARYNARZ.DAT], AdjP[ŁÓDŹ.GEN.SG.F]>

Or like this?

- DAĆ.NEG.FIN.PL.M1.PRI:  
<NumP[MARYNARZ.DAT], NP[ŁÓDŹ.GEN.SG.F]>

## Examples (contd.)

- Wołano go Grubas.  
call-IMPS him-ACC Fat-NOM  
'He was called Fat.'
- WOŁAĆ.AFF.IMPS:  
<NP[ON.ACC], NP[GRUBAS.NOM.SG.M1]>



## Examples (contd.)

- Wołano go Grubas.  
call-IMPS him-ACC Fat-NOM  
'He was called Fat.'
- WOŁAĆ.AFF.IMPS:  
<NP[ON.ACC], NP[GRUBAS.NOM.SG.M1]>
- Wołali go Grubas.  
call-PL.M1.TER him-ACC Fat-NOM.SG.M1  
'They called him Fat.'
- WOŁAĆ.AFF.FIN.PL.M1.TER:  
<NP[ON.ACC], NP[GRUBAS.NOM.SG.M1]>

# Control and raising

Subject vs. object control/raising:

- Janek kazał (Marysi) być miłą.  
Janek-NOM ordered Marysia-DAT be nice-INST.FEM  
'Janek ordered Marysia to be nice.'
- Janek obiecał (Marysi) być miły(m).  
Janek promised Marysia be nice-NOM/INST.M123  
'Janek promised Marysia to be nice.'
- Janek mógł być miły(m).  
Janek could be nice-NOM/INST.M123  
'Janek could be nice.'

# Positional equivalence

From Kallas 1993 and Świdziński 1993:

- Udawał naiwnego i że nic nie  
 pretended-SG.M123 naïve-SG.M12 and that nothing-ACC not  
 rozumie.  
 understand-SG.TER  
 'He pretended to be naïve and to understand nothing.'
- Jana dziwi, że Maria woli Piotra, i jej brak  
 John-ACC surprise that Mary prefers Peter and her lack-NOM  
 gustu.  
 good taste-GEN  
 'John is surprised that Mary prefers Peter and by her lack of  
 good taste.'

# References

- Kallas, K. (1993). *Składnia współczesnych polskich konstrukcji współrzędnych*. Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Polański, K., editor (1980–1992). *Słownik syntaktyczno-generatywny czasowników polskich*. Zakład Narodowy im. Ossolińskich / Instytut Języka Polskiego PAN, Wrocław / Kraków.
- Saloni, Z. and Świdziński, M. (1998). *Składnia współczesnego języka polskiego*. Wydawnictwo Naukowe PWN, Warsaw, 4th (changed) edition.
- Świdziński, M. (1993). Dalsze kłopoty z bezokolicznikiem. In J. Sambor, J. Linde-Usiekniewicz, and R. Huszcza, editors, *Językoznawstwo synchroniczne i diachroniczne*, pages 303–314. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.
- Świdziński, M. (1998). Syntactic dictionary of Polish verbs. Unpublished manuscript, Version 3a, Warsaw University.