Syntactic Approximation of Semantic Roles

Wojciech Jaworski and Adam Przepiórkowski

INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ul. Jana Kazimierza 5, 01-248 Warsaw

PolTAL 2014
Warsaw, 18 September 2014

Context

Context:

- the **CLARIN-PL project** (`http://clarin-pl.eu/`; until end of 2015),
- with tasks such as:
  - development of large **valence dictionary** of Polish,
  - development of a **robust LFG grammar** of Polish,
- leading to building a parser constructing **semantic representations** of sentences,
- to be used in **textual entailment** and related tasks.

Hence, neo-Davidsonian representations, with **semantic roles**.

Context

Context:

- the **CLARIN-PL project** (`http://clarin-pl.eu/`; until end of 2015),
- with tasks such as:
  - development of large **valence dictionary** of Polish,
  - development of a **robust LFG grammar** of Polish,
- leading to building a parser constructing **semantic representations** of sentences,
- to be used in **textual entailment** and related tasks.

Hence, neo-Davidsonian representations, with **semantic roles**.

Context

Context:

- the **CLARIN-PL project** (`http://clarin-pl.eu/`; until end of 2015),
- with tasks such as:
  - development of large **valence dictionary** of Polish,
  - development of a **robust LFG grammar** of Polish,
- leading to building a parser constructing **semantic representations** of sentences,
- to be used in **textual entailment** and related tasks.

Hence, neo-Davidsonian representations, with **semantic roles**.

Context

Context:

- the **CLARIN-PL project** (`http://clarin-pl.eu/`; until end of 2015),
- with tasks such as:
  - development of large **valence dictionary** of Polish,
  - development of a **robust LFG grammar** of Polish,
- leading to building a parser constructing **semantic representations** of sentences,
- to be used in **textual entailment** and related tasks.

Hence, neo-Davidsonian representations, with **semantic roles**.

Context

Context:

- the **CLARIN-PL project** (`http://clarin-pl.eu/`; until end of 2015),
- with tasks such as:
  - development of large **valence dictionary** of Polish,
  - development of a **robust LFG grammar** of Polish,
- leading to building a parser constructing **semantic representations** of sentences,
- to be used in **textual entailment** and related tasks.

Hence, neo-Davidsonian representations, with **semantic roles**.

Why semantic roles?

They **facilitate entailment**, e.g.:

- 'Somebody wrote a paper for PolTAL 2014.' $\rightarrow$
- 'Somebody wrote a paper.'

Entailment **not immediate without** semantic roles: (different arities of *write*):

- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a, poltal2014)$ $\rightarrow$
- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a)$

Entailment **immediate with** semantic roles:

- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a) \wedge destination(e, poltal2014)$ $\rightarrow$
- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a)$

Why semantic roles?

They **facilitate entailment**, e.g.:

- 'Somebody wrote a paper for PolTAL 2014.' $\rightarrow$
- 'Somebody wrote a paper.'

Entailment **not immediate without** semantic roles: (different arities of *write*):

- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a, poltal2014)$ $\rightarrow$
- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a)$

Entailment **immediate with** semantic roles:

- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a) \wedge destination(e, poltal2014)$ $\rightarrow$
- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a)$

Why semantic roles?

They **facilitate entailment**, e.g.:

- 'Somebody wrote a paper for PolTAL 2014.' $\rightarrow$
- 'Somebody wrote a paper.'

Entailment **not immediate without** semantic roles: (different arities of *write*):

- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a, poltal2014) \rightarrow$
- $\exists a \exists p\, article(a) \wedge person(p) \wedge write(p, a)$

Entailment **immediate with** semantic roles:

- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a) \wedge destination(e, poltal2014) \rightarrow$
- $\exists e \exists a \exists p\, article(a) \wedge person(p) \wedge write(e) \wedge agent(e, p) \wedge patient(e, a)$

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Issue and outline

The **issue** to be discussed here:

- what semantic roles (FrameNet, VerbNet, Sowa, etc.)...
- ...for a language with no existing semantic role resources...
- ...such as Polish?

Outline:

- **experiments with adopting** VerbNet and Sowa's roles to Polish (briefly; see the *SEM 2014 paper),
- solution proposed instead – to approximate 'semantic roles' via **rich morphosyntax**.

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Experiment

Experiment:

- 37 Polish verbs (selected at random),
- 393 occurrences of verbs,
- with the total of **843 arguments**,
- **annotated with semantic roles**,
- by (the same) **7 annotators** each;
- two repertoires of semantic roles used:
  - **VerbNet** (30 roles; Kipper *et al.* 2000),
  - **Sowa**'s (18 roles; Sowa 2000).

Results:

- rather **low inter-annotator agreement** (Fleiss's κ):
- 0.617 for VerbNet roles,
- 0.648 for Sowa's roles,
- **above 0.8 needed** for reasonable quality (Artstein and Poesio, 2008).

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Conclusions

Conclusions from the experiments:

- inter-annotator **agreement low**,
- also due to **inherent problems** with these semantic role systems,
- perhaps **designing a new repertoire** of semantic roles would solve some of the problems,
- but this would be a **high-risk task**,
- requiring **(too) much subsequent effort**:
  - annotating a (sufficiently large) corpus,
  - training a semantic role labeller.

Decision: explore **a more cost-effective possibility**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object…),
- **grammatical case** (dative, instrumental…),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object…),
- **grammatical case** (dative, instrumental…),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object…),
- **grammatical case** (dative, instrumental…),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object...),
- **grammatical case** (dative, instrumental...),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object…),
- **grammatical case** (dative, instrumental…),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object...),
- **grammatical case** (dative, instrumental...),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

Idea

For a language like Polish, better **approximate semantic roles** on the basis of:

- **grammatical functions** (subject, object...),
- **grammatical case** (dative, instrumental...),
- **preposition forms**, etc.

Expected **advantages**:

- **deterministic assignment** of ‘semantic roles’,
- **high uniqueness** of ‘semantic roles’ among arguments of a single verb,
- should be sufficient to handle **many cases of entailment**.

The meaning of morphosyntax?

A strong **tradition in Slavic linguistics**:

- **morphological cases have meanings**: Roman Jakobson 1971a,b,
- "cases have meanings and that **this meaning can be stated in a precise and illuminating way**" (Wierzbicka, 1986, p. 386): Anna Wierzbicka 1980, 1981, 1983, 1986.

In practice not necessarily very precise:

- "the dative noun refers to an individual affected by a process or state which obtains **in some part of his personal sphere**, be it the sphere of potency, the sphere of empathy, the sphere of awareness, or the private sphere" (Dąbrowska, 1997, p. 68).

The meaning of morphosyntax?

A strong **tradition in Slavic linguistics**:

- **morphological cases have meanings**: Roman Jakobson 1971a,b,
- "cases have meanings and that **this meaning can be stated in a precise and illuminating way**" (Wierzbicka, 1986, p. 386): Anna Wierzbicka 1980, 1981, 1983, 1986.

In practice not necessarily very precise:

- "the dative noun refers to an individual affected by a process or state which obtains **in some part of his personal sphere**, be it the sphere of potency, the sphere of empathy, the sphere of awareness, or the private sphere" (Dąbrowska, 1997, p. 68).

Proposed solution – example

Janek pomógł Marii.
Janek.NOM.SUBJ helped.ACTIVE Maria.DAT.OBJ–TH

‘Janek helped Maria.’

‘Roles’ assigned:

- R0 (‘**actor**’):
  - **subjects** of active verbs,
  - **agentive PPs** of passive verbs headed by PRZEZ ‘**by**’,
- R2 (‘**recipient**’):
  - objective arguments in the **dative**,
  - **PPs** headed by DLA ‘**for**’.

Proposed solution – example

Janek pomógł Marii.
Janek.NOM.SUBJ helped.ACTIVE Maria.DAT.OBJ–TH

‘Janek helped Maria.’

‘Roles’ assigned:

- R0 (‘**actor**’):
  - **subjects** of active verbs,
  - **agentive PPs** of passive verbs headed by PRZEZ ‘**by**’,
- R2 (‘**recipient**’):
  - objective arguments in the **dative**,
  - **PPs** headed by DLA ‘**for**’.

Proposed solution – ‘semantic roles’

Role	Approximate description
R0	Actor of an action (Agent, Effector)
R1	Undergoer of an action (Patient, Theme, Product)
R2	Dative argument (Beneficiary, Recipient)
R3	Instrumental argument (Instrument)
R4	Adlative argument in both physical and abstract (functional, purposive) meaning (Destination, Recipient, Theme)
R5	Ablative argument in both physical and abstract (causal) meaning (Source)
R6	Locative argument in both physical and abstract meaning
R7	Perlative argument
R8	Topic of communication
R9	Temporal argument (point in time)
R10	Manner argument

Proposed solution – grammatical functions and cases

Usually, ‘semantic roles’ are assigned on the basis of the **LFG grammatical function** of the argument (as well as the voice of the verb; below for active voice):

Argument	**Role**
SUBJ	R0
OBJ	R1
OBJ-TH	R2
OBL-INST	R3
OBL-GEN	R1
OBL-STR	R1
OBL	(see next slide)
XCOMP	R8
COMP	R8
XCOMP-PRED	R8

Proposed solution – prepositions and cases (for obl)

Preposition / morphological case	Role
dla[gen], przeciw[dat], wobec[gen]	R2
do[gen], ku[dat], między[acc], na[acc], nad[acc], po[acc], pod[acc], pomiędzy[acc], ponad[acc], poza[acc], przed[acc], w[acc], za[acc]	R4
dzięki[dat], od[gen], spod[gen], spośród[gen], wskutek[gen], z[gen], zza[gen]	R5
koło[gen], między[inst], na[loc], nad[inst], po[loc], pod[inst], pomiędzy[inst], ponad[inst], poniżej[gen], poza[loc], przed[inst], przy[loc], u[gen], w[loc], wokół[gen], wśród[gen], za[inst]	R6
bez[gen], poprzez[acc], przez[acc], z[inst]	R7
jako[nom], o[acc], o[loc]	R8
podczas[gen]	R9
według[gen]	R10

Semantics after syntax

This solution adopts the **description-by-analysis** approach to doing semantics in LFG:

- **semantic representation** is obtained by analysing f-structures,
- i.e., **non-tree-configurational syntactic representations** containing information about predicates, grammatical functions and morphosyntactic features,
- this is a **common approach** in LFG.

This is opposed to **co-description** (usual in other frameworks), where semantic analysis proceeds together with syntactic analysis.

Semantics after syntax

This solution adopts the **description-by-analysis** approach to doing semantics in LFG:

- **semantic representation** is obtained by analysing f-structures,
- i.e., **non-tree-configurational syntactic representations** containing information about predicates, grammatical functions and morphosyntactic features,
- this is a **common approach** in LFG.

This is opposed to **co-description** (usual in other frameworks), where semantic analysis proceeds together with syntactic analysis.

Semantics after syntax

This solution adopts the **description-by-analysis** approach to doing semantics in LFG:

- **semantic representation** is obtained by analysing f-structures,
- i.e., **non-tree-configurational syntactic representations** containing information about predicates, grammatical functions and morphosyntactic features,
- this is a **common approach** in LFG.

This is opposed to **co-description** (usual in other frameworks), where semantic analysis proceeds together with syntactic analysis.

Semantics after syntax

This solution adopts the **description-by-analysis** approach to doing semantics in LFG:

- **semantic representation** is obtained by analysing f-structures,
- i.e., **non-tree-configurational syntactic representations** containing information about predicates, grammatical functions and morphosyntactic features,
- this is a **common approach** in LFG.

This is opposed to **co-description** (usual in other frameworks), where semantic analysis proceeds together with syntactic analysis.

Advantages

Advantages of this approach:

- **cheap**:
  - a few days of defining the mapping from morphosyntax to 'semantic roles'
  - *instead of* years of building a training resource like VerbNet;
- **deterministic**: for a given argument, role is assigned deterministically;
- high **uniqeness**: for a given verb, different arguments get different roles:
  - only 1.7% verbs get non-unique assignment of 'semantic roles',
  - compared to 4.4% for Sowa's roles and 2.5% for VerbNet roles;
- captures many **entailments** (no quantitative evaluation available now).

Advantages

Advantages of this approach:

- **cheap**:
  - a few days of defining the mapping from morphosyntax to 'semantic roles'
  - *instead of* years of building a training resource like VerbNet;
- **deterministic**: for a given argument, role is assigned deterministically;
- high **uniqeness**: for a given verb, different arguments get different roles:
  - only 1.7% verbs get non-unique assignment of 'semantic roles',
  - compared to 4.4% for Sowa's roles and 2.5% for VerbNet roles;
- captures many **entailments** (no quantitative evaluation available now).

Advantages

Advantages of this approach:

- **cheap**:
  - a few days of defining the mapping from morphosyntax to 'semantic roles'
  - *instead of* years of building a training resource like VerbNet;
- **deterministic**: for a given argument, role is assigned deterministically;
- high **uniqeness**: for a given verb, different arguments get different roles:
  - only 1.7% verbs get non-unique assignment of 'semantic roles',
  - compared to 4.4% for Sowa's roles and 2.5% for VerbNet roles;
- captures many **entailments** (no quantitative evaluation available now).

Advantages

Advantages of this approach:

- **cheap**:
  - a few days of defining the mapping from morphosyntax to 'semantic roles'
  - *instead of* years of building a training resource like VerbNet;
- **deterministic**: for a given argument, role is assigned deterministically;
- high **uniqeness**: for a given verb, different arguments get different roles:
  - only 1.7% verbs get non-unique assignment of 'semantic roles',
  - compared to 4.4% for Sowa's roles and 2.5% for VerbNet roles;
- captures many **entailments** (no quantitative evaluation available now).

Possible entailments – examples

Examples of some **possible entailments**:

- *Janek pobił Tomka.* ‘Janek beat Tomek up.’ →
- *Tomek został pobity.* ‘Tomek was beaten up.’

- *Janek przesłał do Tomka książkę.*
  ‘Janek sent a book to Tomek.’
  (lit. ‘Janek sent to Tomek (a/the) book.ACC.’) →
- *Janek przekazał Tomkowi książkę.*
  ‘Janek transferred a book to Tomek.’
  (lit. ‘Janek transferred Tomek.DAT (a/the) book.ACC.’)

- *Janek powiedział, że Tomek wygrał.*
  ‘Janek said that Tomek had won.’ →
- *Janek mówił o Tomku.*
  ‘Janek was talking about Tomek.’

Possible entailments – examples

Examples of some **possible entailments**:

- - *Janek pobił Tomka.* ‘Janek beat Tomek up.’ →
  - *Tomek został pobity.* ‘Tomek was beaten up.’
- - *Janek przesłał do Tomka książkę.*
    ‘Janek sent a book to Tomek.’
    (lit. ‘Janek sent to Tomek (a/the) book.ACC.’) →
  - *Janek przekazał Tomkowi książkę.*
    ‘Janek transferred a book to Tomek.’
    (lit. ‘Janek transferred Tomek.DAT (a/the) book.ACC.’)
- - *Janek powiedział, że Tomek wygrał.*
    ‘Janek said that Tomek had won.’ →
  - *Janek mówił o Tomku.*
    ‘Janek was talking about Tomek.’

Possible entailments – examples

Examples of some **possible entailments**:

- - *Janek pobił Tomka.* ‘Janek beat Tomek up.’ →
  - *Tomek został pobity.* ‘Tomek was beaten up.’

- - *Janek przesłał do Tomka książkę.*
    ‘Janek sent a book to Tomek.’
    (lit. ‘Janek sent to Tomek (a/the) book.ACC.’) →
  - *Janek przekazał Tomkowi książkę.*
    ‘Janek transferred a book to Tomek.’
    (lit. ‘Janek transferred Tomek.DAT (a/the) book.ACC.’)

- - *Janek powiedział, że Tomek wygrał.*
    ‘Janek said that Tomek had won.’ →
  - *Janek mówił o Tomku.*
    ‘Janek was talking about Tomek.’

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Disadvantages

Main **disadvantage**:

- same morphosyntax may express different ‘semantic roles’,
- e.g. **oblique PPs** headed by z ‘**with**’ + **instrumental** case:
  - perlative (R7),
  - thematic (R1),
  - co-agentive (R0);
- the **mapping always selects only one** of these, for all verbs.

Possible solutions:

- consider **constellations of arguments**, not just single argument,
- consider **lexical semantics** of arguments of prepositions.

Thank you for your attention!

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, **34**(4), 555–596.

Dąbrowska, E. (1997). *Cognitive Semantics and the Polish Dative*, volume 9 of *Cognitive Linguistics Research*. Mouton de Gruyter, Berlin.

Jakobson, R. O. (1971a). Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus. In *Selected Writings II*, pages 23–71. Mouton, The Hague.

Jakobson, R. O. (1971b). Morfologičeskie nabljudenija nad slavjanskim skloneniem. In *Selected Writings II*, pages 154–183. Mouton, The Hague.

Kipper, K., Dang, H. T., Schuler, W., and Palmer, M. (2000). Building a class-based verb lexicon using TAGs. In *Proceedings of TAG+5 Fifth International Workshop on Tree Adjoining Grammars and Related Formalisms*.

Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA.

Wierzbicka, A. (1980). *The Case for Surface Case*. Karoma, Ann Arbor, MI.

Wierzbicka, A. (1981). Case marking and human nature. *Australian Journal of Linguistics*, **1**, 43–80.

Wierzbicka, A. (1983). The semantics of case marking. *Studies in Language*, **7**, 247–275.

Wierzbicka, A. (1986). The meaning of a case: A study of the Polish dative. In R. D. Brecht and J. S. Levine, editors, *Case in Slavic*, pages 386–426. Slavica Publishers, Columbus, OH.