Predicate-argument structure and thematic roles

Iza Škrjanec
Hot and odd topics in semantics
UdS, 2018
Outline

- Events and predicate-argument structure
- Thematic roles: traditional roles and their problems
- Proto roles: Proto-Agent and Proto-Patient (Dowty 1991)
  - Properties of proto-roles
- Large-scale annotation of proto-roles for English (Reisinger et al. 2015)
- Universal Decompositional Semantics (White et al. 2016, http://decomp.net/)
Motivating examples

1) a) John bought that car.
   b) That car was bought by John.

2) a) Mary unlocked the door with a key.
   b) A key unlocked the door.
   c) Mary ate spaghetti with a fork.
   d) *A fork ate spaghetti.
Understanding events and their participants

- *Who did what to whom? (+ Where and when?)*
- Surface and deep structure
- Higher-level representation: **thematic roles** (or *deep cases* or *theta roles*)
  - They express the abstract role the participants have in an event
  - Live in the syntax-semantic interface
  - Verb: **the predicate**
  - Noun phrases around the verb: **the arguments**
  - Predicate-argument structure
    - \[_{\text{ARG0}} \text{The group agreed}_{\text{ARG1}} \text{it wouldn’t make an offer}.\]
Thematic roles: two views

1. Individual thematic roles: very specific
   a. No assumption that there is one thematic role common to the following:
      i. The workers *built* a wall. → the “Builder” role
      ii. The robber *killed* the security guard. → the “Killer” role

2. Argument-indexing view: aims to generalize with the following constraints (Theta-Criterion)
   a. Each NP argument is assigned exactly one thematic role.
   b. The same thematic role is not assigned to two NP arguments of the same predicate.

Strong empirical claims: the arguments of a verb must always be assigned to some official thematic role, and two arguments are distinct enough, so they do not fall under the same role!
Some traditional thematic roles

- **Agent**: volitional causer of an event. *The waiter brought the soup.*
- **Experiencer**: experiences an event. *John has a headache.*
- **Theme**: participant most directly affected by an event. *He stopped the car.*
- **Result**: end product of an event. *The city built a new tennis court.*
- **Beneficiary**: benefits from an event. *He made reservations for his boss.*
- **Source**: origin of the object of a transfer event. *I flew from Frankfurt.*
- **Goal**: destination of an object of a transfer event. *I drove to France.*
- **Locative**: adjunct; *She was sleeping on the couch.*
- **Temporal**: adjunct; *I woke up at 9 a.m.*
Some problems of thematic roles (Dowty 1991)

Lack of consensus between linguists

1. **Role fragmentation** and unclear boundaries between roles
   a. Too few or too many roles?
   b. Greater distinction between arguments, less generalization
      i. *He drove the car 50 m.p.h. (EXTENT)* vs *He drove the car too fast (MANNER?)*. Setting very specific boundaries

2. Cases with no motivation for **distinguishing** between two arguments
   a. *This is similar to that.*
   b. *This resembles that.*
   c. No apparent asymmetry between the arguments: what is the Agent and what is the Theme?
   d. Surface differences → not necessarily role differences
Prototypical roles: Proto-Agent and Proto-Patient (Dowty 1991)

- Roles as “prototypes”, not discrete categories
- Discrete feature decomposition:
  - Works in phonology, morphology, and syntax
  - Why not semantics? No evidence that cognitive interpretation of events is limited to discrete types!
- Dowty’s suggestion: roles are “conceptual clusters of properties”
- Proto-Agent and Proto-Patient
  - Not exclusive
  - An argument does not have to have all properties of a role
  - Principle for deciding between roles
- Adjuncts not included
Proto-Agent properties (Dowty 1991)

Only one Proto-Agent property displayed:

- **Volitional involvement** in the event: *John is being polite to Bill.*
- **Sentience and/or perception**: *John sees Mary.*
- **Causation**: *Teenage unemployment causes crime.*
- **Movement**: relative to other participants: *He accidentally fell.*
- **Independent existence**: the referent is not brought into being by the event, but rather existed before and after the event: *John needs a new car.*
Proto-Patient properties (Dowty 1991)

Only one Proto-Patient property displayed:

- **Change of state**: John made a mistake.
- **Incremental theme**: Temporal progress of an event can be measured in terms of parts of a whole (gradual change). John filled the glass with water.
- **Causally affected**: Smoking causes cancer.
- **Stationary relative to another participant**: The bullet entered the target.
- **Existence not independent of event**: Coming into and out of existence. John built a house.
Semantic role labeling (SRL)

- Automatically finding thematic roles for each argument of the predicate
- Supervised machine learning task: labeled data
- Resources with role labels:
  - PropBank (Palmer et al. 2005): individual verb senses
  - FrameNet (Ruppenhofer et al. 2006): frames of the same event (buy, sell, purchase)
  - ...
- Semantic Proto-Role Labeling: Dowty (1991) → but not yet tested on large data sets
Proto-roles: large scale and corpus based (Reisinger et al. 2015)

- Proto-role hypothesis: evidence from two rating experiments
- **Experiment 1**: based on a psycholinguistic study (Kako 2006), rating Proto-Agent and Proto-Patient properties of arguments
- **Experiment 2**: corpus-based, same questions
- 12 properties tested: are they more like Proto-A or Proto-P?

<table>
<thead>
<tr>
<th>Role property</th>
<th>Q: How likely or unlikely is it that...</th>
</tr>
</thead>
<tbody>
<tr>
<td>instigated</td>
<td>Arg caused the Pred to happen?</td>
</tr>
<tr>
<td>volitional</td>
<td>Arg chose to be involved in the Pred?</td>
</tr>
<tr>
<td>awareness</td>
<td>Arg was/were aware of being involved in the Pred?</td>
</tr>
<tr>
<td>sentient</td>
<td>Arg was sentient?</td>
</tr>
<tr>
<td>moved</td>
<td>Arg changes location during the Pred?</td>
</tr>
<tr>
<td>phys.existed</td>
<td>Arg existed as a physical object?</td>
</tr>
<tr>
<td>existed_before</td>
<td>Arg existed before the Pred began?</td>
</tr>
<tr>
<td>existed.during</td>
<td>Arg existed during the Pred?</td>
</tr>
<tr>
<td>existed_after</td>
<td>Arg existed after the Pred stopped?</td>
</tr>
<tr>
<td>changed.poss</td>
<td>Arg changed possession during the Pred?</td>
</tr>
<tr>
<td>changed_state</td>
<td>The Arg was/were altered or somehow changed during or by the end of the Pred?</td>
</tr>
<tr>
<td>stationary</td>
<td>Arg was stationary during the Pred?</td>
</tr>
</tbody>
</table>

Table 2: Questions posed to annotators.
Proto-roles: large scale and corpus based (Reisinger et al. 2015)

- **Experiment 1**: sentences with transitive verbs and non-word NPs
  - Example: “The neeglur killed the bogrub.”
  - Question: How likely or unlikely is it that the bogrub was altered or somehow changed during or by the end of the killing?

- **Experiment 2**: sentences from the PropBank (no modification with non-words)
  - Greater number of verb lemmas
Proto-roles: large scale and corpus based (Reisinger et al. 2015):

- Confirmed the psycholinguistic results (crowdsourcing is cheaper than lab)
- Large-scale support to the proto-role hypothesis
Universal Decompositional Semantics on Universal Dependencies (White et al. 2016)

- Thematic roles dwell in syntax too!
- The Universal Dependencies Project (UD)
  - Framework for cross-linguistically consistent grammatical annotation
  - Syntactic dependency annotation
  - Many languages, one standard
  - http://universaldependencies.org/
- White et al. 2016: semantic annotation atop of syntax
Example from Universal Dependencies

English, Bulgarian, Czech, Swedish
Universal Decompositional Semantics on Universal Dependencies (White et al. 2016)

- The Universal Decompositional Semantics Project (Decomp, http://decomp.net/)
  - decomposition of lexical meanings into component parts
- PredPatt software: identifies the predicate-argument structure from Universal Dependencies
  - Tested for several languages, focused on the English UD
- White et al. 2016: improvements of Reisinger et al. 2015
- The output of PredPatt used for annotation on 3 levels (crowdsourcing):
  - Semantic roles
  - Event decomposition
  - Word sense decomposition
- Good inter-annotator agreement; potential for cross-lingual comparisons of proto roles
Conclusions

- Thematic roles: abstract representations of the roles arguments take in the event described by the predicate
- Proto-role hypothesis
- From theory to computation: SRL
- First empirical and large-scale evidence for the proto-role hypothesis: for English, but may expect other languages (Universal Dependencies)
References

- Decompositional Semantics Initiative (Decomp): [http://decomp.net/](http://decomp.net/)
- Universal Dependencies: [http://universaldependencies.org/](http://universaldependencies.org/)