Distributional data for constraints-based semantic parsing

Emmanuele Chersoni

Aix-Marseille University
University of Pisa
Research funded by A*Midex Foundation
Grammar as a set of descriptions

- Integration of the notion of construction in linguistic theories (Fillmore et al., 1988)
- A pairing of form and meaning, describing a specific linguistic phenomenon
- **Ditransitive construction** Subj V Obj1 Obj2
  *She gave him a kiss*
- **Idiomatic construction** *kick the bucket, beating a dead horse*
- No separation between lexicon and syntax: constructions can be lexemes, phrases, sentences...
Properties of a construction

- Constructions result from the convergence of different (syntactic, semantic...) properties
- Property Grammars (Blache: 2000): a framework for description of linguistic phenomena in terms of interacting properties
- Syntactic properties of different types (linearity, co-occurrence, exclusion, uniqueness, dependency)
- Each property represented independently from the others
- Construction described in terms of sets of satisfied / violated properties: we can describe ungrammatical input
PG and semantics

- PG describe all syntactic properties by means of constraints
- Semantic side of the constructions has to be represented too
- What kind of “semantic constraints” should be included in a construction? How do they come into play in semantic composition?
Constructions include info on the frames they evoke (Sag: 2012).

Empty slots to be filled with lexical material matching some properties listed in a feature structure.

Symbolic features, properties of objects and events.

Our proposal: to integrate another dimension of meaning, i.e. word usage.
We argue that there at least two major types of information from which we learn word meanings. The first is what we call experiential information... the second type of information is language-based. In particular, it is derived from the general linguistic context in which words appear. (Vigliocco: 2009)
In **Distributional Semantics** word meaning represented by vectors

The value of each dimension is (a function of) the frequency of occurrence of a target word within a context (data extracted from large corpora)

Words having similar vectors will also have similar meanings (cosine the most popular measure of vector similarity)

In **structured Distributional Semantic models**, co-occurrence statistics in the form of triples: word pairs and the syntactic relation linking them (Baroni, Lenci: 2010)
# Distributional semantic models

<table>
<thead>
<tr>
<th>target</th>
<th>lecture-n</th>
<th>gun-n</th>
<th>speech-n</th>
<th>shoot-v</th>
<th>correct-v</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher-n</td>
<td>28</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>professor-n</td>
<td>31</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>soldier-n</td>
<td>0</td>
<td>20</td>
<td>2</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>politician-n</td>
<td>2</td>
<td>0</td>
<td>27</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>target</td>
<td>depType</td>
<td>context</td>
<td>freq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>-----------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>professor-n</td>
<td>subj</td>
<td>teach-v</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>professor-n</td>
<td>subj</td>
<td>claim-v</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>deer-n</td>
<td>obj</td>
<td>shoot-v</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>soldier-n</td>
<td>subj</td>
<td>fight-v</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>good-j</td>
<td>attr</td>
<td>student-n</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clock-n</td>
<td>obj</td>
<td>beat-v</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Distributional semantic models

- Dependency as the generic syntactic-semantic relation linking a dependent to its head

- Dependency relation as a bridge between Construction Grammars and Structured Distributional Semantic Models

- Symbolic features still used in the unification process...

- ... but we can also use the distributional vectors of prototypical fillers to measure how much a given word fits a slot
Constructions and distributional data

[\[
\begin{array}{c}
\text{BOY}\\
\text{FORM} \quad \text{BOY} \\
\text{SYN} \quad \begin{bmatrix}
\text{CAT} & \text{N} \\
\text{VAL} & \text{-} \\
\text{PROPERTIES} & \text{-}
\end{bmatrix} \\
\text{SEM} \quad \begin{bmatrix}
\text{INDEX} & 1 \\
\text{FRAMES} & \text{-} \\
\text{SYM}\_\text{FEATURES} & \text{LIVING\_BEING, HUMAN, MALE…} \\
\text{SEM}\_\text{VECTOR} & \text{PLAY-V 36.3, SCHOOL-N 11.2, YOUNG-J 29.7…}
\end{bmatrix}
\end{array}
\]
Constructions and distributational data

<table>
<thead>
<tr>
<th>VP FORM</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CAT</strong></td>
<td><strong>VP</strong></td>
</tr>
<tr>
<td><strong>VAL</strong></td>
<td><strong>NP¹</strong></td>
</tr>
<tr>
<td><strong>PROPERTIES</strong></td>
<td>[V \prec NP, PP, NP \prec PP, V[\text{TRANS}]\text{[ACTIVE]} \rightarrow NP[\text{OBJ}]\ldots]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INDEX</strong></td>
</tr>
<tr>
<td><strong>FRAMES</strong></td>
</tr>
<tr>
<td><strong>SYM_FEATURES</strong></td>
</tr>
<tr>
<td><strong>SEM_VECTOR</strong></td>
</tr>
<tr>
<td><strong>ARG_VECTORS</strong></td>
</tr>
</tbody>
</table>
Constructions and distributional data
Constructions and distributional data

- Both symbolic features and distributional vectors to characterize the slots.

- Each slot gets assigned the vectors of words typically satisfying the dependency constraint represented by the slot.

- The more a word resembles the typical fillers, the more it will be fitting the slot.

- Methods for the comparison: exemplar-based, or prototype-based (Lenci: 2011). E.g. to get the fitting score of 'apple' as obj of 'to eat', compare the 'apple' vector with the avg vector of (food, meat, pasta...).
Constructions and distributional data

- The result of evaluation of symbolic constraints is a boolean value.
- The evaluation of a 'distributional constraint' returns a scalar value.
- Even in case of violation, the input still gets a description.
Constructions and distributional data

TO_EAT+AN_APPLE
FORM

TO_EAT_AN_APPLE
CAT VP
VAL NP¹
PROPERTIES

INHERITS_FROM VP
VIOLATIONS {...}

INDEX V

INHERITS_FROM: TO_EAT

FRAMES
ARG¹: eater
LIVING_BEING
INDEX_REF A
CONST_SAT 1

SYM_FEATURES
TO_EAT: ... , APPLE: ...

SEM_VECTOR
BREAKFAST-N 16.0, KITCHEN-N 10.5, HUNGRY-N 2.7...

ARG_VECTORS
ARG¹: [MAN-N, DOG-N...], ARG²: [SEM_SIM 0.7]
Constructions and distributional data

THE_BOY+TO_EAT+AN_APPLE
FORM

THE_BOY_EATS_AN_APPLE

CAT S
VAL -

SYN

PROPERTIES

INHERITS_FROM TransConstruction

VIOATIONS { ... }
Constructions and distributional data: the advantages

- Possibility of taking into account another ‘dimension’ of meaning: word usage
- Concepts have internal structure: different degrees of prototypicality
- Easy to automatically derive word representations from corpora
Constructions and distributional data: future work

- A method for modeling semantic compositionality with vectors
- Model of activation integrating both PG and semantic constraints


C. Fillmore et al., Regularity and idiomaticity in grammatical constructions, in Language, vol. 64, no. 3, 1988, pp. 501-538;

