Construction Grammar meets Semantic Vector Spaces: A radically data-driven approach to semantic classification of slot fillers

Natalia Levshina†
Kris Heylen‡

†University of Marburg, RC Deutscher Sprachatlas
‡University of Leuven, RU Quantitative Lexicology and Variational Linguistics
Acknowledgments

- Most of the presented work was presented when the first author worked at the University of Leuven
- This research was partly funded by an FWO grant awarded to Dirk Geeraerts and Dirk Speelman
Outline

• quantitative approaches to constructional semantics: the problem of semantic classes
• distributional semantic models as a method of semantic classification
• experiments with nominal and verbal classes in Dutch doen and laten CCx
• discussion and future research
Quantitative Models of Syntactic Variation

<table>
<thead>
<tr>
<th>semasiological perspective</th>
<th>onomasiological perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>collexeme-based</td>
<td>collexeme-based</td>
</tr>
<tr>
<td>feature-based</td>
<td>feature-based</td>
</tr>
</tbody>
</table>

- Collostructional analysis (Stefanowitsch & Gries 2003)
- Cluster analysis (Levshina 2012)
- Distinctive collexeme analysis (Gries & Stefanowitsch 2004)
- Logistic regression (Bresnan et al. 2007), classification trees (Baayen 2008), random forests (Levshina 2011)
- Mixed-effect models (Baayen 2008)

NWASV Nijmegen 2012
Quantitative Models of Syntactic Variation

ALL approaches involve semantic classes, either at the coding stage (feature-based analyses), or at the interpretation stage (collexemes)

NWASV Nijmegen 2012
Why do we need semantic classes?

- Theoretically, learning constructions involves learning generalizations, such as semantic classes (cf. Goldberg 2006).
- Epistemologically, we are interested in the most parsimonious explanation.
- This might be a rare case when the interests of both speakers and linguists converge.
Semantic Classes: State of the Art

- as a rule, intuitive and subjective
- ‘standard’ classifications (e.g. Levin 1993, Garretson 2004):
  - not many
  - for English
  - incomplete
  - not tested empirically
- Gries and Stefanowitsch 2010: corpus-driven verb classes, but
  - limited contextual features (18 prepositions)
  - subjective evaluation

NWASV Nijmegen 2012
Semantic Classes: Desiderata

- data-driven, (potentially) entire vocabulary
- objective validation
- semantic relationships are multidimensional
  - different criteria of similarity
- varying schematicity of semantic relationships
  - different levels of granularity

Instead of working with one *a priori* classification, let’s compare different ones and see which works the best.
Outline

• quantitative approaches to constructional semantics: the problem of semantic classes
• our proposal: distributional semantic models as a method of semantic classification
• experiments with nominal and verbal classes in Dutch doen and laten CCx
• discussion and future research
Our proposal

- a bottom-up quantitative approach based on distributional Semantic Vector Space models
- task-specific:
  - adjustable criteria of similarity
  - adjustable granularity
- validation in a real data set for near-synonymous doen and laten CCx (onomasiological perspective)
Semantic Vector Space Models

Standard technique in Computational Linguistics:

- corpus based, bottom-up clustering of semantically related words into semantic classes (Turney & Pantel 2010)

Based on the Distributional Hypothesis (Firth 1957):

- *You shall know a word by the company it keeps*
- Words appearing in similar contexts tend to have similar meanings

Method

- each word is assigned a vector stating the word's co-occurrence frequencies with a range of possible contexts
- words with similar context vectors have similar meanings
Semantic Vector Space Models

|      | gun | psychopath | knife | cruelly | lovingly | mother | lovers | toilet | ...
|------|-----|------------|-------|---------|----------|--------|--------|--------|------
| kiss | 2   | 2          | 0     | 1       | 89       | 56     | 98     | 3      |      
| hug  | 3   | 1          | 2     | 5       | 77       | 49     | 88     | 0      |      
| kill | 10  | 59         | 67    | 69      | 0        | 8      | 12     | 1      |      
| murder | 97 | 65         | 58    | 81      | 1        | 9      | 9      | 2      |      
| ....  |     |            |       |         |          |        |        |        |      

- co-occurrence frequency of target words (rows) with context words (columns)
- High dimensional matrix (only small subset shown)
Semantic Vector Space Models

|     | gun | psychopath | knife | cruelly | lovingly | mother | lovers | toilet | ...
|-----|-----|------------|-------|---------|----------|--------|--------|--------|-----
| kiss| 2   | 2          | 0     | 1       | 89       | 56     | 98     | 3      |     
| hug | 3   | 1          | 2     | 5       | 77       | 49     | 88     | 0      |     
| kill| 10  | 59         | 67    | 69      | 0        | 8      | 12     | 1      |     
| murder| 97 | 65         | 58    | 81      | 1        | 9      | 9      | 2      |     
| ....|    |            |       |         |          |        |        |        |     

- general overview of collactional behaviour and distributional properties of a language's vocabulary
- \(\approx\) behavioural profiles (Divjak etal 2006) but many more features

NWASV Nijmegen 2012
Semantic Vector Space Models

- weighting of frequencies (pointwise mutual information)
- projection of vectors into a semantic "word space"
- measure proximity of vectors in space (cosine)
- cluster words based on vector proximity
Semantic Vector Space Models

SVS come in many different flavours

- **Technical parameters** (frequency weighting scheme, similarity measure, dimensionality reduction technique, clustering technique,...)
- **Number of clusters** → granularity of semantic distinctions
  - dependent on specific application (cf. infra)
- **Definition of 'context'** → type of semantics captured
  - Document-based context features: topical relations
  - Window-based, bag-of-words context features: loose associations
  - Dependency-based context features: tight relations (synonymy)
  - Subcategorisation Frame features (verbs only): Levin-like classes

NWASV Nijmegen 2012
Bag of words model

- context feature = co-occurring word in window left and right of target word
- looser, associative semantic relations: e.g. *doctor*-*hospital*

|      | gun | psychopath | knife | cruelly | lovingly | mother | lovers | toilet | ...
|------|-----|------------|-------|---------|----------|--------|--------|--------|------
| kiss | 2   | 2          | 0     | 1       | 89       | 56     | 98     | 3      |      
| hug  | 3   | 1          | 2     | 5       | 77       | 49     | 88     | 0      |      
| kill | 10  | 59         | 67    | 69      | 0        | 8      | 12     | 1      |      
| murder | 97 | 65         | 58    | 81      | 1        | 9      | 9      | 2      |      
| .... |     |            |       |         |          |        |        |        |      

N WASV Nijmegen 2012
Bag of words model

- context feature = co-occurring word in window left and right of target word
- looser, associative semantic relations: e.g. *doctor-hospital*

<table>
<thead>
<tr>
<th></th>
<th>gun</th>
<th>psychopath</th>
<th>knife</th>
<th>cruelly</th>
<th>lovingly</th>
<th>mother</th>
<th>lovers</th>
<th>toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td>kiss</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td>56</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>hug</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>77</td>
<td>49</td>
<td>88</td>
<td>0</td>
</tr>
</tbody>
</table>

The psychopath _killed_ his victims with a blunt knife. ...
## Dependency based model

<table>
<thead>
<tr>
<th></th>
<th>+PP with gun</th>
<th>+SU psychopat</th>
<th>+OBJ psychopat</th>
<th>+PP with knife</th>
<th>+ADV cruelly</th>
<th>+ADV lovingly</th>
<th>+SU mother</th>
<th>+SU Lovers</th>
<th>+PP on toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td>kiss</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td>56</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>hug</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>77</td>
<td>49</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>kill</td>
<td>10</td>
<td>59</td>
<td>2</td>
<td>67</td>
<td>69</td>
<td>0</td>
<td>8</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>murder</td>
<td>97</td>
<td>65</td>
<td>1</td>
<td>58</td>
<td>81</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- context feature = word in specific syntactic dependency relation with target word
- tight semantic relations: *hospital - clinic*

NWASV Nijmegen 2012
Dependency based model

<table>
<thead>
<tr>
<th></th>
<th>+PP with gun</th>
<th>+SU psychopath</th>
<th>+OBJ psychopath</th>
<th>+PP with knife</th>
<th>+ADV cruelly</th>
<th>+ADV lovingly</th>
<th>+ SU mother</th>
<th>+ SU Lovers</th>
<th>+ PP on toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td>kiss</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td>56</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>hug</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>77</td>
<td>49</td>
<td>88</td>
<td>0</td>
</tr>
</tbody>
</table>

- context feature = word in specific syntactic dependency relation with target word
- tight semantic relations: *hospital - clinic*

NWASV Nijmegen 2012
Subcat frame model

| Action | SU | SU/OBJ | SU/OBJ/ADV | SU/PP | SU/OBJ/PP | SU/OBJ/OBJ | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>kiss</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>hug</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>kill</td>
<td>10</td>
<td>59</td>
<td>2</td>
<td>67</td>
<td>69</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>murder</td>
<td>97</td>
<td>65</td>
<td>1</td>
<td>58</td>
<td>81</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- context feature = subcategorization frame co-occurring with target verb (only used for verbs!) (Schulte i.Walde 2006)
- Levin-like verb classes: e.g. *lie, stand, sit, lean*

NWASV Nijmegen 2012
The psychopath killed his victims with a blunt knife.

---

**Subcat frame model**

| SU | SU/OBJ | SU/OBJ/ADV | SU/PP | SU/OBJ/PP | SU/OBJ/IOBJ | ...
|----|--------|------------|-------|-----------|-------------|------
| kiss | 2     | 2          | 0     | 0         | 1           | 89   |

- **context feature** = subcategorization frame co-occurring with target verb (only used for verbs!) (Schulte i.Walde 2006)
- **Levin-like verb classes**: e.g. *lie, stand, sit, lean*

*NWASV Nijmegen 2012*
Semantic Vector Space Models

3 models form a continuum lexical to syntactic
purely lexical distributional information

\[ \downarrow \]

lexical and syntactic (dependency) information

\[ \downarrow \]

purely syntactic (subcat.) distributional properties

more intermediate forms, depending on

• number of dependency relations (e.g. arguments only)
• inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)
Semantic Vector Space Models

3 models form a continuum lexical to syntactic
purely lexical distributional information
↓
lexical and syntactic (dependency) information

more intermediate forms, depending on
• number of dependency relations (e.g. arguments only)
• inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

The psychopath killed his victims with a blunt knife. ...
Semantic Vector Space Models

3 models form a continuum lexical to syntactic
purely lexical distributional information
\[ \Downarrow \]
lexical and syntactic (dependency) information
\[ \Downarrow \]
purely syntactic (subcat.) distributional properties
more intermediate forms, depending on
• number of dependency relations (e.g. arguments only)
• inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

NWASV Nijmegen 2012
Semantic Vector Space Models

3 models form a continuum lexical to syntactic

purely lexical distributional information

↓

lexical and syntactic (dependency) information

- number of dependency relations (e.g. arguments only)
- inclusion of some "lexical" info in subcat frames (e.g. prepositions or semantic noun classes)

The psychopath killed his victims with a blunt knife. ...

SU_anim / OBJ_anim / PP_with

NWASV Nijmegen 2012
Semantic Vector Space Models

- 16 different models for verbs
- 2 different models for nouns

<table>
<thead>
<tr>
<th>LEXICAL</th>
<th>SYNTACTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag of words</td>
<td>Dependently based</td>
</tr>
<tr>
<td>BOW5 BOW4</td>
<td>DEP8 DEP7 DEP3</td>
</tr>
<tr>
<td>SFr23 SFp23 SFc23</td>
<td>SFr9 SFp9 SFc9 SFr5 SFc5</td>
</tr>
<tr>
<td>SFs23</td>
<td>SFs9 SFs5</td>
</tr>
</tbody>
</table>
Outline

• quantitative approaches to constructional semantics: the problem of semantic classes
• distributional semantic models as a method of semantic classification
• experiments with nominal and verbal classes in Dutch doen and laten CCx
• discussion and future research

NWASV Nijmegen 2012
Dutch causative constructions

Haar stem deed het glas barsten.
her voice did the glass break

Harry liet het glas barsten.
Harry made the glass break
Dutch causative constructions

Haar stem deed het glas barsten.
her voice did the glass break

Harry liet het glas barsten.
Harry made the glass break
Dutch causative constructions

Haar stem deed het glas barsten.
her voice did the glass break

Harry liet het glas barsten.
Harry made the glass break
Dutch causative constructions

Haar stem deed het glas barsten.
her voice did the glass break

Harry liet het glas barsten.
Harry made the glass break

NWASV Nijmegen 2012
Dutch causative constructions

Haar stem deed het glas barsten.
her voice did the glass break

Harry liet het glas barsten.
Harry made the glass break
Classes of Cr and Ce

• data: Twente News Corpus
• 2 models:
  - purely lexical distributional information (bag of words)
  - lexical and syntactic dependency information (e.g. noun N as a subject of a verb V)
• granularity: from 2 to 100 classes emerging from hierarchical cluster analysis
Classes of Effected Predicates

- 16 models, a continuum from purely lexical information to purely syntactic information (subcat frames)
- different granularity (number of clusters in hierarchical cluster analysis): from 5 to 100
Evaluation

• test set: 6800 obs. with causative doen and laten from Dutch newspaper corpora
• objects: all explicit non-pronominal fillers of Causer, Causee and Effected Predicate slots
• criterion: prediction of doen or laten in the observations
• method: logistic regression model, several indicators $R^2$, $C$, Somer’s $D_{xy}$, Gamma, Tau, AIC
Prediction by Causer Classes

Causer's classes, Nagelkerke $R^2$

- BOW
- DepRel8

$n$ of clusters

N WASV Nijmegen 2012
## 6 Clusters of Causer

<table>
<thead>
<tr>
<th>No</th>
<th>Top freq nouns</th>
<th>doen or laten</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cd, cijfer, plaat, herstel, stem, aanslag, afwezigheid, rentree, resultaat, aanpak</td>
<td>doen</td>
</tr>
<tr>
<td>2</td>
<td>Feyenoord, PSV, dirigent, speler, beurs, orkest, doelman, zanger, Van Gaal, componist</td>
<td>laten</td>
</tr>
<tr>
<td>3</td>
<td>Gergiev, Van Hecke, gemeente bestuur, Harnoncourt, Morissette, Pollini, secretaris generaal, AH To Go, alleskunner, Ax</td>
<td>laten</td>
</tr>
<tr>
<td>4</td>
<td>Verenigde Staten, VS, Amerika, Europa, Washington, geheel, tempo, Duitsland, Engels, India</td>
<td>laten</td>
</tr>
<tr>
<td>5</td>
<td>regering, minister, bedrijf, trainer, president, belegger, muziek, ploeg, premier, man</td>
<td>laten</td>
</tr>
<tr>
<td>6</td>
<td>Mahlerstem, zaal technicus, beroepskader, Neal Evans, Bastuba en contrabasclarinet, Tuomarila, Wanderlied</td>
<td>NA</td>
</tr>
</tbody>
</table>
Prediction by Causee Classes

Causee classes, Nagelkerke $R^2$

R^2

BOW  DepRel8

n of clusters

NWASV Nijmegen 2012
Prediction by Effected Predicate

EP classes, Nagelkerke $R^2$

- BOW15
- BOW15 rVrel
- BOW15 Varel
- BOW15 Varel
- rVrel
- rVrel
- Vrel
- Vrel
- Varel
- Varel
- 23richsubcat
- 23richsubcat
- 5relrichsubcat
- 5relrichsubcat
- 23relprep
- 23relprep
- 23sclass
- 23sclass
- 9sclass
- 9sclass
- 5sclass
- 5sclass
- 23syn
- 23syn
- 9syn
- 9syn
- 5syn
- 5syn

NWASV Nijmegen 2012
# 35 Clusters for EP (Examples)

<table>
<thead>
<tr>
<th>No</th>
<th>Top freq verbs</th>
<th>doen or laten</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>herleven, kantelen, versmelten, beven, samensmelten, verslappen, sidderen, smelten, instorten, vervagen</td>
<td>doen</td>
</tr>
<tr>
<td>24</td>
<td>stijgen, zakken, dalen, groeien, overlopen, rijzen, terugzakken, wankelen, daveren, overvloeien</td>
<td>doen</td>
</tr>
<tr>
<td>25</td>
<td>denken, vermoeden, geloven, besluiten, vrezen, hopen, verzuchten, toegeven, betogen, concluderen</td>
<td>doen</td>
</tr>
<tr>
<td>23</td>
<td>zien, weten, leiden, maken, doen, blijken, voelen, kennen, worden, schijnen</td>
<td>laten</td>
</tr>
<tr>
<td>12</td>
<td>horen, verstaan, betrappen, afleiden, merken, delen, rijmen, verkiezen, verwachten, associëren</td>
<td>laten</td>
</tr>
<tr>
<td>34</td>
<td>liggen, vallen, gaan, komen, lopen, staan, zitten</td>
<td>laten</td>
</tr>
</tbody>
</table>
The best models for 3 Slots

Best models with Causer, Causee and EP classes, Nagelkerke $R^2$
The best models for 3 Slots

Best models with Causer, Causee and EP classes, Nagelkerke $R^2$

Causers give more data reduction than verbs!
All three slots in one model

3 slots, Nagelkerke $R^2$

NWASV Nijmegen 2012
All three slots in one model

3 slots, Nagelkerke $R^2$

slots interact in conveying constructional meaning

NWASV Nijmegen 2012
Outline

• quantitative approaches to constructional semantics: the problem of semantic classes
• distributional semantic models as a method of semantic classification
• experiments with nominal and verbal classes in Dutch doen and laten CCx
• discussion and future research
Desiderata Revisited

• data-driven, (potentially) entire vocabulary

• objective validation

• semantic relationships are multidimensional
  ➔ different criteria of similarity

• varying schematicity of semantic relationships
  ➔ different levels of granularity

NWASV Nijmegen 2012
Desiderata Revisited

- data-driven, (potentially) entire vocabulary
  bottom-up classes work!
- objective validation
- semantic relationships are multidimensional
  \[\implies\] different criteria of similarity
- varying schematicity of semantic relationships
  \[\implies\] different levels of granularity
Desiderata Revisited

• data-driven, (potentially) entire vocabulary
  bottom-up classes work!

• objective validation
  Cr and EP classes perform better than Ce

• semantic relationships are multidimensional
  \[\Rightarrow\text{different criteria of similarity}\]

• varying schematicity of semantic relationships
  \[\Rightarrow\text{different levels of granularity}\]
Desiderata Revisited

- data-driven, (potentially) entire vocabulary
  bottom-up classes work!
- objective validation
  Cr and EP classes perform better than Ce
- semantic relationships are multidimensional
  \(\Leftrightarrow\) different criteria of similarity
  syntax-sensitive models perform the best
- varying schematicity of semantic relationships
  \(\Leftrightarrow\) different levels of granularity
Desiderata Revisited

- data-driven, (potentially) entire vocabulary
  bottom-up classes work!
- objective validation
  Cr and EP classes perform better than Ce
- semantic relationships are multidimensional
  ➞ different criteria of similarity
  syntax-sensitive models perform the best
- varying schematicity of semantic relationships
  ➞ different levels of granularity
  nouns ‘need’ less classes than verbs

NWASV Nijmegen 2012
Future research

- the bottom-up construction-specific classes are similar to the classes found in the literature. Are semantic classes cross-constructionally (cross-linguistically) stable?
  - compare with other constructions
    - compute validity measures for different clustering solutions (e.g. silhouette widths)
- a solution for semasiological studies
Thank you!

for further information:

natalevs@gmail.com
kris.heylen@arts.kuleuven.be