Doe wat je niet laten kan: A usage-based analysis of Dutch causative constructions

Natalia Levshina
Supervisors: Dirk Geeraerts
            Dirk Speelman

University of Leuven
RU Quantitative Lexicology and Variational Linguistics
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of *doen* and *laten*
   - contextual factors
6. Conclusions and future research
Empirical Cognitive Semantics

• Meaning is a concept

• Meaning reveals itself in usage

• Meaning is subject to social variation
Empirical Cognitive Semantics

- Meaning is a concept
  - categorization and comprehension
categories and exemplars
- Meaning reveals itself in usage
- Meaning is subject to social variation
Empirical Cognitive Semantics

• Meaning is a concept
  → categorization and comprehension
categories and exemplars
• Meaning reveals itself in usage
  → distributional bottom-up approach
• Meaning is subject to social variation
Empirical Cognitive Semantics

- Meaning is a concept
  - categorization and comprehension
  - categories and exemplars
- Meaning reveals itself in usage
  - distributional bottom-up approach
- Meaning is subject to social variation
  - lectally diverse data
State of the Art

- bias towards ‘alternations’ in studies of constructions (cf. Goldberg 2002: each construction should [also] be studied on its own right)
- main focus on generalizations and prototypes, although exemplar effects in human categorization are proved to be powerful
- insufficient attention paid to contextual (social and genre-related) variation; some voices that it is too problematic to be studied as a factor (Gries & Divjak 2010)
- lack of theoretically interpretable yet robust statistical techniques that can integrate different sources of variation
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of *doen* and *laten*
   - contextual factors
6. Conclusions and future research
Dutch Causative Constructions

De tovenaar *deed* zijn dienaars een kasteel bouwen.

*liet*

The magician **made**/**let**

his servants **cause**

a castle **affect**

build **effect**

slot names: Kemmer & Verhagen (1994)
<table>
<thead>
<tr>
<th>doen</th>
<th>laten</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct causation</td>
<td>Indirect causation</td>
</tr>
<tr>
<td>«The initiator produces the effected event directly; there is no intervening energy source ‘downstream’»</td>
<td>«Some other force besides the initiator is the most immediate source of energy in the effected event»</td>
</tr>
</tbody>
</table>

Verhagen & Kemmer (1997)
Other dimensions of variation

Speelman & Geeraerts 2009:

• *doen* is used more frequently in formal registers
• *doen* is used more frequently in Belgian Dutch
• *doen* is more prone to collocational fixation

“*doen* is an obsolescent form with a tendency towards semantic and lexical specialization”
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of *doen* and *laten*
   - contextual factors
6. Conclusions and future research
Data

• 5768 observations with causative *doen* and *laten*

• corpus

  - formal written genre: articles from Dutch and Belgian quality newspapers in TwNC and LeNC
  - informal written genre: posts from Usenet.nl and Usenet.be in Dutch, spam-checked
  - informal spoken genre: Dutch and Flemish spontaneous face-to-face conversations from CGN

• retrieval: XML parser (newspapers), regular expressions, manual cleaning-up (e.g. *laten we zeggen*)
37 Variables, 731 features

- **General context**: sentence type, clause type, clause mood, clause tense, Cx syntactic function
- **Causer**: semantic class, syntactic expression, POS, person, number, definiteness
- **Causee**: see Causer; intentionality, thematic proto-role
- **Affectee**: see Causer
- **Effected Predicate**: semantic class, semantic source and target domain, prefix, ‘root’, extended transitivity, prepositional complements
- **Other**: coreferentiality, possession, negation, modals, adverbials
## Some corpus methods in CogSem

<table>
<thead>
<tr>
<th></th>
<th>Polysemy</th>
<th>Synonymy</th>
<th>Exemplars</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-and multinomial</td>
<td></td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Regression</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>predef. senses</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Collostructional</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Profiles</td>
<td></td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>predef. senses</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Exemplar Spaces</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Exemplar Spaces Method

Example of Data

<table>
<thead>
<tr>
<th></th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplar 1</td>
<td>A</td>
<td>yes</td>
<td>X</td>
</tr>
<tr>
<td>Exemplar 2</td>
<td>A</td>
<td>yes</td>
<td>Y</td>
</tr>
<tr>
<td>Exemplar 3</td>
<td>B</td>
<td>no</td>
<td>X</td>
</tr>
</tbody>
</table>

Distances (Gower)

<table>
<thead>
<tr>
<th></th>
<th>Ex. 1</th>
<th>Ex. 2</th>
<th>Ex. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplar 1</td>
<td>0</td>
<td>0.33</td>
<td>0.66</td>
</tr>
<tr>
<td>Exemplar 2</td>
<td>0.33</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exemplar 3</td>
<td>0.66</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Partitioning into k Clusters

<table>
<thead>
<tr>
<th></th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplar 1</td>
<td>1</td>
</tr>
<tr>
<td>Exemplar 2</td>
<td>1</td>
</tr>
<tr>
<td>Exemplar 3</td>
<td>2</td>
</tr>
</tbody>
</table>

Interpretation: Cluster profiles

<table>
<thead>
<tr>
<th>Cluster</th>
<th>A</th>
<th>B</th>
<th>yes</th>
<th>no</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

distinctive features
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of doen and laten
   - contextual factors
6. Conclusions and future research
Exemplar Space of *doen*

SMACOF MDS, *doen*

stress: 0.07
2 Clusters of *doen*
Clusters of *doen*: Top distinctive features

<table>
<thead>
<tr>
<th>Red Cluster</th>
<th>Blue Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psych. Caused Event (fig.)</td>
<td>3\textsuperscript{rd} Person Causee</td>
</tr>
<tr>
<td>Psych. Caused Event (lit)</td>
<td>Social Caused Event (fig.)</td>
</tr>
<tr>
<td>Human Causee</td>
<td>Nominal Causee</td>
</tr>
<tr>
<td>Pronominal Causee</td>
<td>Social Caused Event (lit.)</td>
</tr>
<tr>
<td>1\textsuperscript{st} Person Causee</td>
<td>Causee undergoes change</td>
</tr>
<tr>
<td>No change occurs</td>
<td>No prepositional complements</td>
</tr>
<tr>
<td>Verb class Mental Processes</td>
<td>Explicit Causee</td>
</tr>
<tr>
<td>Root <em>denk</em>-</td>
<td>Explicit NP Affectee</td>
</tr>
<tr>
<td>Preposition <em>aan</em></td>
<td>Verb class Motion</td>
</tr>
<tr>
<td>Implicit Causee</td>
<td>Indefinite NP Affectee</td>
</tr>
</tbody>
</table>
Causative *doen*: A dictionary entry

1. Cause a change in the world, especially a social change conceptualized as an abstract objective event.

   *Betere booklets zijn er geregeld (maar dat doet de prijs stijgen).*

2. Trigger a psychological reaction.

   *Dacht ook dat ik het wist, tot je vraag me deed twijfelen, en terecht...*

   *doen denken aan* “remind of”

   "*Groovin’ on E* " doet met zijn fris jengelende gitaren en neuzelende fluisterzang denken aan *The Kinks.*
Contextual variation of *doen*

*doen* is more frequent
• in Belgium
• in formal genres

(Chi-test, $\alpha = 0.05$)
doen: Interaction Context*Semantics

Belgium
- newspapers
- Usenet
- dialogues

Netherlands
- newspapers
- Usenet
- dialogues
Belgium, newspapers

doen, Flemish newspapers
The Netherlands, newspapers

doen, Dutch newspapers
Belgium, Usenet

doen, Flemish Usenet
The Netherlands, Usenet
doen, Dutch Usenet
Belgium, conversations

doen, Flemish conversations
The Netherlands, conversations
doen, Dutch conversations
**doen: Lectally enriched entry**

1. *esp. Belg. and formal* Cause a change in the world, especially a social change conceptualized as an abstract objective event.

   Betere booklets zijn er geregeld (maar dat doet de prijs stijgen).

2. Trigger a psychological reaction.

   Dacht ook dat ik het wist, tot je vraag me deed twijfelen, en terecht...

   doen denken aan “remind of”

   "Groovin’ on E" doet met zijn fris jengelende gitaren en neuzelende fluisterzang denken aan The Kinks.
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of *doen* and *laten*
   - contextual factors
6. Conclusions and future research
Exemplar Space of *laten*

SMACOF MDS, *laten*

stress: 0.09
3 Clusters of *laten*

SMACOF MDS, laten
## Clusters of *laten*: Top distinctive features

<table>
<thead>
<tr>
<th>Red Cluster</th>
<th>Blue Cluster</th>
<th>Green Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psych. Caused Event (lit)</td>
<td>Explicit NP Causee</td>
<td>Human Undefined Causee</td>
</tr>
<tr>
<td>Psych. Caused Event (fig)</td>
<td>Intransitive EP</td>
<td>Intentional Causee</td>
</tr>
<tr>
<td>Human Causee</td>
<td>V. Motion</td>
<td>Causee causes change</td>
</tr>
<tr>
<td>V. Perception</td>
<td>Abstract Causee</td>
<td>V Creation</td>
</tr>
<tr>
<td>Root zie-</td>
<td>Causee is changed</td>
<td>V. Phys Manipulation</td>
</tr>
<tr>
<td>Pronominal Causee</td>
<td>Indefinite NP Affectee</td>
<td>V. Social Interaction</td>
</tr>
<tr>
<td>V. Factive</td>
<td>V. Existence</td>
<td>V. Change of Possession</td>
</tr>
<tr>
<td>Root weet-</td>
<td>Plural Affectee</td>
<td>Material Object Affectee</td>
</tr>
<tr>
<td>Affectee Clause</td>
<td>No coreferentiality</td>
<td>Sing. Number Causee</td>
</tr>
<tr>
<td>1(^{st}) Person Causee</td>
<td>Nominal Affectee</td>
<td>Explicit NP Affectee</td>
</tr>
</tbody>
</table>
Causative *laten*: An entry

1. Stop or fail to carry out impingement, abandon, leave.
   
   *Ook Rombouts laat Van Hecke vallen.*

2. Suffer or benefit from someone’s actions, e.g. services.
   
   *En als ik het geschreven heb ga ik het eerst laten registreren bij ISBN.*
   
   *Ref Allaerts laat zich vangen.*

3. Provide information, show.
   
   *Nou dat heeft ie laten zien.*
Contextual variation of *laten*

- NL > BE newspapers
- overall written > spoken
  
  (Chi-test, $\alpha = 0.05$)
*laten*: Interaction Context*Semantics*
The Netherlands, newspapers
Belgium, Usenet

latten, Flemish Usenet
The Netherlands, Usenet
Belgium, conversations

luten, Flemish conversations
The Netherlands, conversations

 laten, Dutch conversations
**laten:** Lectally enriched entry

1. Stop or fail to carry out impingement, abandon, leave.
   
   *Ook Rombouts laat Van Hecke vallen.*

2. Suffer or benefit from someone’s actions, e.g. services.
   
   *En als ik het geschreven heb ga ik het eerst laten registreren bij ISBN.*

3. esp. Neth. Dutch Provide information, show.
   
   *Nou dat heeft ie laten zien.*
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of doen
   - semantic and contextual variation of laten
5. Onomasiological variation (near-synonymy)
   - variation of doen and laten
   - contextual factors
6. Conclusions and future research
“Het is moeilijk een scherp onderscheid te maken tussen het gebruik van *doen* en dat van *laten*, al zijn er wel enkele tendensen aan te wijzen. Bij de keuze voor één van beide werkwoorden kunnen niet alleen betekenisverschillen, maar ook geografische en stilistische verschillen een rol spelen. In ieder geval zijn *doen* en *laten* niet altijd zonder meer uitwisselbaar.” ANS 1997
doen vs laten: Onomasiological Variation

SMACOF MDS, doen and laten

stress: 0.09

dim 1

dim 2

red: doen
blue: laten
doen vs laten: Top distinctive features

<table>
<thead>
<tr>
<th>doen</th>
<th>laten</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract Causer</td>
<td>human Causer</td>
</tr>
<tr>
<td>intransitive predicates</td>
<td>transitive predicates</td>
</tr>
<tr>
<td>3rd person Causer</td>
<td>pronominal Affectee</td>
</tr>
<tr>
<td>V. denoting Mental Processes</td>
<td>no prepositional complements</td>
</tr>
<tr>
<td>root <em>denk</em>-</td>
<td>intentional Causee</td>
</tr>
<tr>
<td>nominal Affectee</td>
<td>3rd person Causee</td>
</tr>
<tr>
<td>preposition <em>aan</em></td>
<td>explicit NP Affectee</td>
</tr>
<tr>
<td>explicit NP Causee</td>
<td>V. of Perception</td>
</tr>
<tr>
<td>abstract Affectee</td>
<td>implicit Causee</td>
</tr>
<tr>
<td>1st person Causee</td>
<td>causing (agentive) Causee</td>
</tr>
</tbody>
</table>
Confirmatory Regression: Design

• Response: *doen* or *laten*
• General semantic predictors:
  - human/abstract/other Causer
  - transitive/intransitive effected predicate
  - Causee causes or undergoes a change
  - intentional or unintentional Causee
  - implicit or explicit Causee
• Contextual variables: Country and Genre
• Exemplar effects?
Modelling Exemplar Effects

• Speelman & Geeraerts (2009): lexical fixation operationalized with the help of collostructional analysis (Stefanowitsch & Gries 2003)

• Bybee (2010): rare collexemes can be semantically similar to frequent collexemes and thus typical of Cx; see also Schmid (2010)

• solution (cf. exemplar model by Nosofsky 1986):
  
  similarity to neighbours = $e^{-d^2}$
  
  where $d$ is distance between exemplars of the same category (doen or laten)
proxDoen and proxLaten
# Main Effects

|            | Estimate | Std. Error | Pr(>|z|)    |
|------------|----------|------------|------------|
| (Intercept)| -4.02    | 0.31       | < 2e-16 ***|
| CrSem=Abstr| 1.86     | 0.17       | < 2e-16 ***|
| CrSem=Oth  | 2.1      | 0.24       | < 2e-16 ***|
| proxDoen   | 0.06     | 0.004      | < 2e-16 ***|
| proxLaten  | -0.01    | 0.001      | < 2e-16 ***|
| Genre=spont_dial | -0.91 | 0.24       | <0.001***   |
| Genre=Usenet | -0.16  | 0.13       | 0.23        |
| CeEnergy=Change | 1.16  | 0.32       | <0.001***   |
| CeEnergy=Oth | 1.26    | 0.29       | <0.001***   |
| CeIntent=No | 0.62    | 0.18       | <0.001***   |
| CeIntent=Undef | -0.04 | 0.24       | 0.85        |
| Country=BE  | 0.33     | 0.12       | 0.008**     |
| CeSynt=Expl | 0.37     | 0.17       | 0.029*      |
| EPTrans = Intr | 0.01  | 0.17       | 0.964       |

full model with interactions: C=0.96 Gamma=0.91 pseudo-$R^2=0.68$
Interpretation of the Model

- *doen* involves direct causation, *laten* indirect?
  - abstract Causers (*doen*) can be only direct ‘causes’, whereas human Causers (*laten*) can affect the situation indirectly
  - implicit Causees (*laten*) are often intermediaries
  - agent-like (intentionally causing) Causees (*laten*) allow the Causer to influence the world indirectly
- *doen* is more popular in formal genres and Belgium
Interaction Country*Genre

Belgium
- newspapers
- Usenet
- dialogues

Netherlands
- newspapers
- Usenet
- dialogues

Legend:
- red: doen
- blue: laten
Average probability of *doen*

**clusters of doen in common space**

- Dimension 1: 77%
- Dimension 2: 53%

Stress: 0.09
doen: Onomasiologically enriched entry

1. esp. Belg. and formal  Cause a change in the world, especially a social change conceptualized as an abstract objective event. Syn. laten

Betere booklets zijn er geregeld (maar dat doet de prijs stijgen).

2. Trigger a psychological reaction.

Dacht ook dat ik het wist, tot je vraag me deed twijfelen, en terecht... Syn. laten (?)

doen denken aan “remind of”

"Groovin’ on E" doet met zijn fris jengelende gitaren en neuzelende fluisterzang denken aan The Kinks.
doen als oorzakelijk hulpww.
synoniem: laten

*ik doe opmerken dat enz.*
*iets doen of doen doen*
*eens steen deed hem struikelen*
*doen te weten*
Average probability of *laten*

clusters of *laten* in common space
laten (both auxiliary and notional)

... 16. maken dat het object de werking verricht die door de onbep. ww. wordt uitgedrukt. **synoniem**: doen

  *zij liet mij vragen of ik meeging,*
  *iem. laten struikelen*
  *zijn gedachten, zijn oog laten gaan*
  *iem. iets laten weten*
Outline

1. Theoretical and methodological background
2. Dutch causative constructions
3. Data and method
4. Semasiological variation (polysemy)
   - semantic and contextual variation of *doen*
   - semantic and contextual variation of *laten*
5. Onomasiological variation (near-synonymy)
   - variation of *doen* and *laten*
   - contextual factors
6. Conclusions and future research
**doen: Summary**

Global semasiological variation:

- two main senses related to semantic domain of causation (cf. Verhagen & Kemmer 1997, but rather social than ‘physical’ causation)

Contextual semasiological variation:

- *doen* is quantitatively and qualitatively poorer in informal genres and in Netherlandic Dutch (cf. Speelman & Geeraerts 2009), biased towards affective causation, esp. fixed *doen denken aan*
A historical perspective

- Verhagen 2000: *doen* used to have a more diverse semantic palette

<table>
<thead>
<tr>
<th>Sense</th>
<th>18th C.</th>
<th>Millenium</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Belgium</td>
<td>Netherlands</td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>+</td>
<td>??</td>
<td>??</td>
<td></td>
</tr>
<tr>
<td>‘Physical’</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

- Belgian Dutch is ‘lagging behind’ (cf. Geeraerts et al. 1999)
*laten*: Summary

Global semasiological variation:
- three main senses related to different semantic roles of Causer and Causee (cf. Verhagen & Kemmer 1997)

Contextual semasiological variation:
- *sense laten3* (*laten zien/weten/horen*) is unpopular in Belgian newspapers
**doen vs. laten: Summary**

- On the general semantic level, the difference between *doen* and *laten* can be formulated in terms of causation (in)directness (cf. Kemmer & Verhagen).
- Cf. polysemy: in all three senses of *laten* the Causer does not actually affect, and the Causee is not affected (indirectntess); in the two senses of *doen* the Causer is the (inevitable) cause/stimulus (directness).
- Exemplar effects play an important role, too.
doen vs. laten: Summary

- Non-mental intransitive *doen* and *laten* are more prone to synonymy than the other senses. At the same time, *doen* can be readily replaced by *laten*, but not the other way round.

Laat wat je doen kan,
maar doe niet wat je laten kan!
Conclusions

• unsupervised bottom-up exemplar spaces are a way of integrating different interrelated types of variation in one semantic model (complete recontexualization)
• the semasiological and onomasiological perspectives provide specific yet compatible semantic information
• both exemplar effects and more general features are found to be important in categorization
• contextual variation has a large influence on the semasiological structure, but less so on the division of labour between *doen* and *laten*
Future research

• experimental confirmation of findings
• comparison with unsupervised distributional methods (Vector Space Models)
• a diachronic study of exemplar space evolution
Vragen? Suggesties? Laat je maar gaan!

for further information:
http://wwwling.arts.kuleuven.be/qlvl
natalia.levshina@arts.kuleuven.be