Let the movies begin:
A typology of LET-verbs in European languages based on a parallel corpus of film subtitles

Natalia Levshina
Philipps University of Marburg

WS Parallel Corpora and Linguistic Theory
46th SLE meeting, Split 2013
Outline

• Verbs of letting in European languages: semantic peculiarities

• Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles
  - why film subtitles?
  - data and method
  - results: language space and exemplar space of LET

• Conclusions
What is LETTING?

- Letting can be defined as 'negative causation'.
- Causation proper: the Antagonist overrides the intrinsic tendency of the Agonist (Talmy 2000).
- Letting: the Antagonist fails, deliberately or not, to override the Agonist's intrinsic tendency.

**Antagonist**  
Causation: *The detective* made *the thief* confess where he had hidden the stolen diamonds.

**Letting:**  
*The detective* let *the thief* escape.
Semantic variation of LETTING

- letting is a very complex concept, which varies along many dimensions, most importantly:
  - force dynamic patterns:
    - active or passive Permitter (*Let my people go* vs. *Let it be*)
    - dynamic or static Permittee (*Let my people go* vs. *Let it be*)
  - physical or psychosocial domain:
    *The plug's coming loose let the water drain from the tank.*
    *The mother doesn't let the child eat junk food.*
A bouquet of functions

- Let's drink!
- Let no one be dead today!
- Let's go away
- Let's leave, get away
- Let's release, let go
- Let's allow, permit
- Let's cause
- Optative
- Cohortative
- 'Middle'
- Stop, quit
The aim of the study

- to explore semantic differences and similarities of etymologically related *let*-verbs in 4 Germanic and 4 Romance languages

  *let, laten, lassen, låta*

  *laisser, dejar, deixar, lasciare*

- historical prototypes (Soares da Silva 2007):
  - Romance: Lat. *laxare* 'let go, release'
  - Germanic: Goth. *letan* 'leave, let'

- what about now?
A methodological problem

• how to compare highly polysemous constructions in different languages?
  – e.g. German lassen and Dutch laten. Both can express letting go, leaving, abandonment, indirect causation, hortativity, etc.

• Often the differences are a matter of degree, or 'prototypicality' of a specific sense. These differences, however, may suggest ongoing language change.
  – E.g. French laisser is losing its 'let go' function, as a manifestation of its more advanced grammaticalization in comparison with other Romance let-verbs (Soares da Silva 2012)
Methodological proposal

- use a multilingual parallel corpus to compare the LET-verbs
- compare the distributions of LET-verbs across many aligned contexts
- compare the semantic scope of each verb by looking at the region occupied by its exemplars in a common semantic space
- compare the prominence of each sense by examining the density of the LET-exemplars in the corresponding region of the common semantic space
Outline

• Verbs of letting in European languages: semantic peculiarities

• Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles
  - why film subtitles?
  - data and method
  - results: language space and exemplar space of LET

• Conclusions
Why film subtitles?

- freely available online for many languages
- timing information helps to align sentences
- free from censorship and normativity, although quality is to some extent controlled by the online community and fans
- closer to 'normal' spoken language than any other existing parallel texts (in comparison to Europarl and Bible translations)

Monolingual subtitles serve as a better approximation of entrenchment in lexical decision experiments than standard 'balanced' corpora!
Is translationese an issue?

- a sample of 15 randomly selected English-source and 15 English-target film subtitles (source languages: German, French, Spanish, Italian, Swedish, Danish, Dutch, Brasilian Portuguese, Russian, Japanese, Chinese), various genres

- search for verb 'let' (per 100 000 words)

- medians are identical, no significant difference in normalized frequencies of *let* (Mann-Whitney test: $p = 0.461$)

- more variation in translated English, although this is not statistically significant (Flinger's test: $p = 0.97$)
Data

- 6 films in different original languages
Data

- subtitles in eight languages from opensubtitles.org
- srt format
- converted to XML
- aligned with the help of timing information (Tiedemann 2011): trimming, Procrustes procedure
- search for verb stems in each language
- 364 multilingual contexts
<table>
<thead>
<tr>
<th>Number</th>
<th>Start time</th>
<th>End time</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00:00:56,080</td>
<td>00:01:00,798</td>
<td>&lt;i&gt;Sooner or later though, you always have to wake up.&lt;/i&gt;</td>
</tr>
<tr>
<td>2</td>
<td>00:01:12,792</td>
<td>00:01:15,476</td>
<td>&lt;i&gt;In cryo you don't dream at all.&lt;/i&gt;</td>
</tr>
<tr>
<td>3</td>
<td>00:01:15,511</td>
<td>00:01:17,250</td>
<td>&lt;i&gt;Doesn't feel like 6 years.&lt;/i&gt;</td>
</tr>
<tr>
<td>4</td>
<td>00:01:17,285</td>
<td>00:01:20,688</td>
<td>&lt;i&gt;More like after tequila and ass kicking.&lt;/i&gt;</td>
</tr>
<tr>
<td></td>
<td>DE_Cx</td>
<td>EN_Cx</td>
<td>ES_Cx</td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>16</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Quantitative analyses

1. The space of 8 verbs: a matrix of dissimilarities between eight verbs (Gower's distances) represented by MDS.

   The Distributional Principle: the **closer** the verbs in the space, the more frequently they are used in the **same contexts**.

2. The space of exemplars:

   2.1. A matrix of dissimilarities between 364 exemplars represented in a MDS-based semantic space.

   The Iconicity Principle: the **closer** the exemplars in the space, the more frequently they are translated with the **same type of construction** (let-verb or not let). In addition, the more central the exemplars in the space, the more frequently they are translated with a let-verb.

   2.2. Kriging to obtain central tendencies for each language
Distances between 8 *let*-verbs

Distances between cognates

![Graph showing distances between cognates of 8 *let*-verbs in different languages.](image)
Quantitative analyses

1. The space of 8 verbs: a matrix of dissimilarities between eight verbs (Gower's distances) represented by MDS.

   The Distributional Principle: the closer the verbs in the space, the more frequently they are used in the same contexts.

2. The space of exemplars:

   2.1. A matrix of dissimilarities between 364 exemplars represented in a MDS-based semantic space.

       The Iconicity Principle: the closer the exemplars in the space, the more frequently they are translated with the same type of construction (let-verb or not let). In addition, the more central the exemplars in the space, the more frequently they are translated with a let-verb.

   2.2. Kriging to obtain central tendencies for each language.
Exemplar space of LET
Exemplar space of LET

'leave':
Ela é minha. Deixem-na em paz.
„She is mine! Leave her to me!“
Exemplar space of LET

'Let's X'!
'let go'
Würdest du den ausreisen lassen an deren Stelle?
Exemplar space of LET

I can't let you marry.

'allow'
I can't let you marry.
Exemplar space of LET

'cause':
Dein wirklicher Vater ließ uns Portale auf der ganzen Welt öffnen.
Form-meaning mapping: Germanic
Kriging: Germanic

German

English

Dutch

Swedish
Form-meaning mapping: Romance
Kriging: Romance

Spanish

French

Italian

Portuguese
Germanic vs. Romance
Outline

• Verbs of letting in European languages: semantic peculiarities

• Quantitative comparison of eight LET-verbs in a parallel corpus of film subtitles
  - why film subtitles?
  - data and method
  - results: language space and exemplar space of LET

• Conclusions
Conclusions

- Film subtitles are a promising source of data that allows for quantitative comparison of many languages.

- Effects of translationese do not seem to be dramatic, but we need further investigation based on a larger sample, controlled for source languages and genres.

- Germanic and Romance *let*-verbs seem to be developing in the opposite directions, having 'swapped' their most typical meanings in comparison with their historical prototypes. The semantic divergence is higher for the Germanic verbs than for the Romance ones.
The End