Grammaticalization paths and clines: Induction and visualization

Natalia Levshina
Leipzig University

Workshop “Parallel text analysis in diachronic research”
Marburg, 23 February 2018
Grammaticalization paths, scales, clines, etc.

- Global ones, e.g. grammaticalization scales

Lehmann 2015: 15
Cycle of language change

Dixon 1994: 183-184
Local scales

- Germanic sandwich (English > Dutch > German)
- Romance sandwich (French > Italian > Spanish)

van Haeringen 1956, Lamirov 2012
Outline of the talk

• Local cline:
  • Grammaticalization of causative auxiliaries MAKE in Romance languages

• Global cline:
  • Grammaticalization of case markers (based on English, Finnish, German, Russian and Turkish)
ParTy corpus

• Corpus of subtitles (films and TED talks)
• Minimum 10 languages per film/unit
• Sentences in English versions aligned with sentences in other languages, pairwise
• Partly available from https://github.com/levshina/ParTy-1.0
Why subtitles?

Cluster Dendrogram

Based on the frequencies of 3-grams (Levshina 2017)
Short: 8 oz.
Tall: 12 oz.
Grande: 16 oz.
Venti: 20 oz.
A local cline:
Grammaticalization of causal auxiliaries MAKE in Romance
Analytic causatives as comparative concepts

FUNCTION: An AC designates a causative event, which involves a causing event (or state) and a caused event (or state), and their participants, most importantly, the causer and the causee. The causer initiates or is responsible for the causing event, whereas the causee is the entity that brings about the caused event (state). There can be also other participants involved (such as the affectee, i.e. the final affected entity). The causing event is underspecified.

FORM: An AC is a construction that consists of two VERBS and their arguments. One VERB (V1) represents in an abstract way the causing event, whereas the other VERB (V2) represents the caused event. The order of the predicates may vary. The clauses should be closely integrated: at least some arguments of V2 should be grammatically dependent on V1.
Examples of ACs

• Don’t *make* me cry.
• *Let* my people go.
• You’re *forcing* me to be the voice of reason.
• 6 careers that *allow* to you to travel around the world.
Films
Dataset

• Translations in 18 European languages (15 Indo-European and 3 Finno-Ugrian languages)
• All ACs extracted manually from each doculect.
• 392 contexts with at least one language having an AC

For more details, see Levshina 2015
Method

• Distance matrix based on weighted features: auxiliary (the greatest weight), form of effected predicate, transitivity, reflexivity, active/passive.

• Multi-Dimensional Scaling (cf. Wälchli & Cysouw 2012) with smacof

• An interactive plot with googleVis: http://www.natalialevshina.com/presentations.html

For more details, see Levshina 2015
Zooming in on Romance ACs

- ita: fare + Vinf
- fra: faire + Vinf
- spa: hacer + (NP) + Vinf
- por: fazer + (NP) + Vinf/Vinf_inflected
- rom: a face + să + Vsubj
Examples

• French, Amélie
  Amandine Poulain aime: (...) *Faire briller le parquet avec des patins...*
  Amandine Poulain likes: (...) polishing the parquet with slippers...

• Italian, Avatar
  *Stronzate, fammi vedere!*
  Bullshit  make.me see
  Bullshit, let me see that!
Romance sandwich(es)

- Lamiroy 2011:
  - French > Italian > Spanish
  - Based on a large number of diverse phenomena

- Soares da Silva 2012:
  - Italian > French > Spanish > Portuguese
  - Based on causative auxiliaries only
Spanish
Italian
Summary

Romanian < Portuguese < Spanish < French < Italian

• Semantics:

  Coercion  Abstract causation
  Lexical   Grammatical

• Interesting: an iconic correlation in syntax:

  $V \ s\ddot{\text{a}} \ V_{\text{subj}} \quad V \ NP \ V_{\text{inf}}/V_{\text{infl}} \quad V \ (NP) \ V_{\text{inf}} \quad VV_{\text{inf}} \quad VV_{\text{inf}}$

  Autonomy  Integration
On semantic maps and diachrony

• “It should be stressed that the MDS method has been criticized because it cannot take into account diachronic information, if available (van der Auwera, 2008, 2013; Narrog, 2010). For example, there is no way to infer any directionality from Figure ...” (Georgakopoulos & Polis 2018)
On semantic maps and diachrony

• “It should be stressed that the MDS method has been criticized because it cannot take into account diachronic information, if available (van der Auwera, 2008, 2013; Narrog, 2010). For example, there is no way to infer any directionality from Figure ...” (Georgakopoulos & Polis 2018)

• Okay, but what about this?
See also a different representation of diachronic information on MDS-based semantic maps in Cysouw & Forker 2009.
A global cline:
Grammaticalization of case markers
Grammaticalization parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Paradigmatic</th>
<th>Syntagmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>Integrity</td>
<td>Structural scope</td>
</tr>
<tr>
<td>Cohesion</td>
<td>Paradigmaticity</td>
<td>Bondedness</td>
</tr>
<tr>
<td>Variability</td>
<td>Paradigmatic variability</td>
<td>Syntagmatic variability</td>
</tr>
</tbody>
</table>

Lehmann 2015: 15
Case hierarchy

• A hierarchy by Blake (2001):

  nominative < accusative / ergative < genitive < dative < locative / prepositional < ablative / instrumental < others

• If a language has a case listed on the hierarchy, it will usually have at least one case from each position to the left. Thus if a language has a dative case it will have a genitive, an accusative or ergative or both, and a nominative.
  • E.g. German: nom < acc < gen < dat/obl
Case study

• Marking of different semantic roles of nouns
• Which roles are more grammaticalized than others?
• Onomasiological approach: from meaning to form (we can’t rely on language-specific categories!)
Data

- eng, deu, fin, rus, tur
- ParTy corpus, 96 NPs, only SG
Examples of roles

• connect_TO_AVATAR
• come_out_FROM_LINK
• write_BOOK
• be_back_FROM_BALL
• train_FOR_MISSION
• dissect_FROG
• go_TO_SELFRIDGE
• love_PUTTER

• get_back_ON_GURNEY
• transduction_from_ROOT
• transduction_TO_ROOT
• contaminate_WITH_SALIVA
• COLONEL_S_orders
• hang_on_TILL_MORNING
• seeds_OF_TREE
• ELSA_build
Ordinal MDS

Cf. Hartmann et al. 2014
Roles
Roles

T and typical P
Roles

Atypical P  T and typical P

S/A
Roles

Atypical P

T and typical P

Possessor
Roles

Atypical P

Partitive

Location

S/A

T and typical P

Possessor
Roles

- Roles: S/A, Atypical P, T and typical P
- Concepts: Location, Partitive, Possessor, Goal

Diagram showing the distribution of roles and concepts in a vector space.
Roles

Atypical P

Typical P

Possessor

Partitive

Location

Goal

Recipient

S/A

T and typical P

Possessor
Roles

- Atypical P
- Typical P
- Possessor
- Partitive
- Location
- Goal
- Recipient
- Purpose
- S/A
- T and typical P
Roles

- Atypical P
- Typical P
- Possessor
- Partitive
- Location
- Goal
- Recipient
- Purpose
- Source
- S/A

- Transduction
- From
- Root
- Back
- From
- Ball

- Atypical Partitives
- Typical Partitives

- Transduction to
- Roger
- Freedom
Roles

- Atypical P
- Typical P
- Possessor
- Partitive
- Location
- Goal
- Recipient
- Purpose
- S/A
- Companion

Diagram with various categories and terms related to roles.
Operationalization of parameters

• Integrity: length of role marker in segments
• Syntactic scope: the entire NP with determiners, adjectives, etc. or only the noun. NP = 1, Noun = 0
• Bondedness:
  • Separability: Possible insertions between the nominal stem and the marker of the role. Yes = 1, No = 0
  • Phonological stability (e.g. vowel harmony in TUR, merge with article in GER, variants of ‘s in ENG). Yes = 1, No = 0
Example of coding: eng

• I just need you to give a message to your boss.
  • Length = 2
  • Scope = 1 (NP)
  • Separability = 1 (Yes)
  • Phonological stability = 1 (Yes)
Example of coding: rus

- Ja ... xoču poprosit’ vas peredat’ poslanije
- I want ask you give message
- vaš-emu   načalnik-u.
  your-DAT.SG boss-DAT.SG

- Length = 1
- Scope = 0 (Noun)
- Separability = 0 (No)
- Phonological stability = 0 (No)
Example of coding: deu

• Ich will nur, dass Sie Ihrem Chef-∅ etwas ausrichten.
  • Length = 0
  • Scope = 0 (Noun)
  • Separability = 0 (No)
  • Phonological stability = 0 (No)
Roles

Atypical P

Partitive

T and typical P

Companion

Recipient

Purpose

Possessor

Source

D1

D2

Location
Average normalized length
Average scope/separability
Average phonological stability
### Ranking of roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Length</th>
<th>Scope/ Separable</th>
<th>Phon. Stability</th>
<th>Mean for 3 parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/A</td>
<td>0.037</td>
<td>0.08</td>
<td>0</td>
<td>0.034</td>
</tr>
<tr>
<td>Atypical Object</td>
<td>0.071</td>
<td>0.113</td>
<td>0</td>
<td>0.061</td>
</tr>
<tr>
<td>Patient</td>
<td>0.07</td>
<td>0.205</td>
<td>0</td>
<td>0.092</td>
</tr>
<tr>
<td>Theme (transfer)</td>
<td>0.083</td>
<td>0.225</td>
<td>0</td>
<td>0.103</td>
</tr>
<tr>
<td>Partitive</td>
<td>0.143</td>
<td>0.225</td>
<td>0</td>
<td>0.127</td>
</tr>
<tr>
<td>Possessor</td>
<td>0.182</td>
<td>0.46</td>
<td>0.06</td>
<td>0.234</td>
</tr>
<tr>
<td>Recipient</td>
<td>0.223</td>
<td>0.5</td>
<td>0.3</td>
<td>0.341</td>
</tr>
<tr>
<td>Location</td>
<td>0.326</td>
<td>0.775</td>
<td>0.491</td>
<td>0.531</td>
</tr>
<tr>
<td>Source</td>
<td>0.419</td>
<td>0.78</td>
<td>0.391</td>
<td>0.531</td>
</tr>
<tr>
<td>Goal</td>
<td>0.327</td>
<td>0.838</td>
<td>0.641</td>
<td>0.602</td>
</tr>
<tr>
<td>Topic (about)</td>
<td>0.512</td>
<td>0.8</td>
<td>0.6</td>
<td>0.637</td>
</tr>
<tr>
<td>Companion</td>
<td>0.502</td>
<td>1</td>
<td>0.7</td>
<td>0.734</td>
</tr>
<tr>
<td>Purpose</td>
<td>0.502</td>
<td>1</td>
<td>0.925</td>
<td>0.809</td>
</tr>
</tbody>
</table>
Zero, Suffix or Adposition?

<table>
<thead>
<tr>
<th>Role</th>
<th>ENG</th>
<th>DEU</th>
<th>FIN</th>
<th>RUS</th>
<th>TUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/A</td>
<td>Zero</td>
<td>Zero</td>
<td>Zero (Suffix?)</td>
<td>Zero/Suffix</td>
<td>Zero/Suffix</td>
</tr>
<tr>
<td>Atypical Object</td>
<td>Zero</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
<td>Zero/Suffix</td>
<td>Zero/Suffix</td>
</tr>
<tr>
<td>Patient</td>
<td>Zero</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Theme</td>
<td>Zero</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Partitive</td>
<td>Prep</td>
<td>Zero</td>
<td>Zero/Suffix</td>
<td>Suffix</td>
<td>Zero</td>
</tr>
<tr>
<td>Possessor</td>
<td>Suffix/Prep</td>
<td>Suffix/Prep</td>
<td>Suffix</td>
<td>Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Recipient</td>
<td>Zero/Prep</td>
<td>Zero/Suffix</td>
<td>Suffix (+ Post)</td>
<td>Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Location</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix</td>
<td>Prep + Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Source</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix</td>
<td>Prep + Suffix</td>
<td>Suffix</td>
</tr>
<tr>
<td>Goal</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix (+ Post)</td>
<td>Prep + Suffix</td>
<td>Suffix (+ Post)</td>
</tr>
<tr>
<td>Topic (about)</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix</td>
<td>Prep + Suffix</td>
<td>Suffix + Post</td>
</tr>
<tr>
<td>Companion</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix + Post</td>
<td>Prep + Suffix</td>
<td>Suffix + Suffix/Post</td>
</tr>
<tr>
<td>Purpose</td>
<td>Prep</td>
<td>Prep (+ Suffix)</td>
<td>Suffix + Post</td>
<td>Prep + Suffix</td>
<td>Suffix + Post</td>
</tr>
</tbody>
</table>
What we’ve learnt

• Blake’s categorical hierarchy can also be interpreted as a continuum of grammaticalization of particular semantic roles.


• Partitives: probably due to high predictability given the ‘classifiers’, e.g. *a cup of tea* → *cuppa tea*.

• More research on the link between predictability and grammaticalization is needed.
So, what can parallel corpora do for (quantitative) diachronic research?

• Induction and visualization of local grammaticalization clines on probabilistic semantic maps
  • Methods: MDS, Multiple Correspondence Analysis (for construction-based maps) – this talk

• Induction and testing of global grammaticalization clines
  • Method: MDS (auxiliary) – this talk

• Detection and testing of changes in multiple probabilistic constraints

• Investigation of diachronic development of language types, computation of transitional probabilities
  • Method: phylogenetic trees (Verkerk 2015)

• …?
References

Thanks!

The slides are available at
http://www.natalialevshina.com/presentations.html

The ParTy corpus is available (partly) at
https://github.com/levshina/ParTy-1.0