Can syntax influence morphological complexity? Evidence from the gender congruency effect

Clara Cohen  Susanne Gahl

UC Berkeley

Cognition and Language Workshop
April 14, 2012
The question

Sometimes it is possible to use either of two roughly synonymous words with differing morphological complexity

(1) I don’t doubt the **truth** of his story
(2) I don’t doubt the **truthfulness** of his story
Sometimes it is possible to use either of two roughly synonymous words with differing morphological complexity

(1) I don’t doubt the **truth** of his story

(2) I don’t doubt the **truthfulness** of his story

What factors can influence this decision?

What are the cumulative consequences of this decision on the morphological structure of the language?
Known constraints

- Usually word-internal (e.g., selectional requirements, relative frequency of stem and affix)
- Can word-external constraints have an effect?
  - Janssen and Caramazza (2003): production latency of complex forms is affected by interaction of context and gender congruency
- Specifically: Can syntactic context constrain not only latency, but also morphological derivation and usage?
Test case requirements

*Syntactic context can constrain morphological derivation and usage*

- Morphological process must be (reasonably) productive
- Must be (in some sense) optional
  - No radical changes in meaning
  - *truth* ~ *truthfulness*, but *fool* ~ *foolishness*
- Morphological process must interact with syntactic context
Test case: Dutch diminutive NPs

- Extremely productive
- Can be optional (i.e., base word is usually also appropriate)

(3) Even mijn schrift / schrift-je pakken
PARTC my notebook / notebook-DIM get.INF
(Just a moment—) I need to get my notebook

- Interacts with syntax through determiner agreement
Test morphological process: diminutivization

- Two genders in Dutch: neuter and common
  
  Neuter nouns always take the determiner *het*
  - *het schaap*, ‘the sheep’;
  - *het vergiet*, ‘the colander’

  Common gender nouns always take the determiner *de*
  - *de beer*, ‘the bear’;
  - *de gieter*, ‘the watering can’
Test morphological process: diminutivization

- Two genders in Dutch: neuter and common
- Diminutives are always neuter

Neuter nouns always take the determiner *het*
- *het schaap,* ‘the sheep’;
  *het vergiet,* ‘the colander’
- *het schaap-je,* ‘the little sheep;
  *het vergiet-je,* ‘the little colander’

Common gender nouns always take the determiner *de*
- *de beer,* ‘the bear’;
  *de gieter,* ‘the watering can’
- *dehet beer-tje,* ‘the little bear’;
  *dehet gieter-tje,* ‘the little watering can’
Diminutives formed from neuter nouns will be called **congruent**

- **het schaap-je**
  ‘the little sheep’
- **het vergiet-je**
  ‘the little colander’

Diminutives formed from common nouns will be called **incongruent**

- **dehet beer-tje**
  ‘the little bear’
- **dehet gieter-tje**
  ‘the little watering can’
Test syntactic context: presence of agreeing determiner

- Janssen and Caramazza (2003): Dutch speakers are slower to produce incongruent diminutives *but only after a definite determiner*
- This effect is absent when the diminutive is produced as a bare noun or with a non-agreeing determiner
- Conclusion: gender information from the base noun interferes with retrieving the neuter determiner *het*, hence delaying speech onset.
If speakers are consistently delayed in producing incongruent diminutives in this particular context, they might be less likely to use the diminutive, and fall back on the (roughly) synonymous base form.

**Consequence:** Diminutive usage is reduced when the base noun is common-gender and the syntactic context includes a definite determiner.

**Question:** Is this result observable in a corpus?
General hypothesis: Word-external phenomena (here, syntactic context) can constrain morphological derivation and usage

Specific hypothesis: The presence of an agreeing determiner will constrain the production of Dutch diminutives

- Speakers should be less likely to use *incongruent* diminutives than congruent diminutives after a definite determiner
The source

- INL Corpus of spoken Dutch (Schuurman et al., 2003)
- 8.9 million words
  - 227,572 standard case neuter nouns
  - 471,827 standard case common nouns
  - 40,495 standard case diminutives
Our database

- Extracted all standard case nouns, coding for
  - Diminutive (Y/N)
  - Log token frequency
  - Lemma number (to match diminutives with their corresponding bases)
  - Speech context (face-to-face conversation, telephone, sermon, interview, etc.)
  - Region (Dutch vs. Flemish)
  - Presence of immediately preceding definite determiner (Y/N)
  - Gender of base (neuter=congruent, common=incongruent)
Pruning

- Theoretical discards:
  - Lemmas whose diminutives were more frequent than the base (Hay, 2001)
  - Observations that did not occur in an informal, spontaneous speech context

- Statistical discards: items for which...
  - No non-diminutive instance was attested
  - No diminutive instance was attested
  - No instance occurred after a definite determiner
Final database

133,597 observations

<table>
<thead>
<tr>
<th></th>
<th>Non-diminutive</th>
<th>Diminutive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuter/Congruent</td>
<td>38,100</td>
<td>3,255</td>
</tr>
<tr>
<td>Common/Incongruent</td>
<td>85,679</td>
<td>6,563</td>
</tr>
</tbody>
</table>

1468 distinct lemmas

- 345 neuter/congruent
- 1042 common/incongruent
Our model

- **Outcome variable:** Whether an observation will be a diminutive

- **Fixed effects**
  - Frequency (log-transformed)
  - Region (Flemish)
  - Congruency (base word neuter)
  - Context (preceding *het/de*)
  - Frequency x Congruency
  - Frequency x Context

- **Random effect**
  - Lemma
Our model

- **Outcome variable**: Whether an observation will be a diminutive

- **Fixed effects**
  - Frequency (log-transformed)
  - Region (Flemish)
  - Congruency (base word neuter)
  - Context (preceding *het/de*)
  - Frequency $\times$ Congruency
  - Frequency $\times$ Context
  - **Congruency $\times$ Context**

- **Random effect**
  - Lemma
Our model

- **Outcome variable**: Whether an observation will be a diminutive
- **Fixed effects**
  - Frequency (log-transformed)
  - Region (Flemish)
  - Congruency (base word neuter)
  - Context (preceding *het/de*)
  - Frequency x Congruency
  - Frequency x Context
  - **Congruency x Context** X
- **Random effect**
  - Lemma
## Analysis

| Fixed effects                        | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------------------|----------|------------|---------|----------|
| (Intercept)                          | 4.10     | 0.14       | 29.47   | 0.000    |
| Log frequency                        | -2.29    | 0.03       | -90.18  | 0.000    |
| Region                               | -0.90    | 0.047      | -19.50  | 0.000    |
| Congruency                           | 2.20     | 0.30       | 7.48    | 0.000    |
| Context                              | -2.83    | 0.16       | -17.72  | 0.000    |
| Log Fq x Congruency                  | -0.57    | 0.06       | -10.33  | 0.000    |
| Log Fq x Context                     | 0.38     | 0.03       | 12.18   | 0.000    |
| Congruency x Context                 | 0.18     | 0.14       | 1.33    | 0.18     |
Gender and syntactic context

- **Congruency (base word neuter) : 2.20**
  Neuter base gender (i.e., forming a congruent diminutive) makes a diminutive form more likely

- **Context (preceding het/de) : -2.83**
  The context of a definite determiner makes a diminutive less likely

- **However**, these are main effects, not interactions
  - No interaction between context and congruency
  - Gender congruency effect is unlikely to be driving preference for congruent diminutives or dispreference for definite diminutives
Frequency

- **Log frequency**: 4.10
  Higher frequency lowers likelihood of diminutive
  - Probably due to exclusion of diminutive-dominant lemmas, whose diminutives have higher frequency than the base forms

- **Log Fq x Congruency**: -0.57
  Effect is increased for neuter-gender nouns

- **Log Fq x Context**: 0.38
  Effect is reversed in the context of a definite determiner
Future work

- Refine model to control for other variables (e.g., animacy)
- Include other agreeing determiners
Can syntax influence morphological complexity?

Not obviously
References I


References II


Implications for morphological productivity

*Morphological productivity* (informally): the ability of a morphological category to increase its membership

- Measured here by *potential productivity* or *category-conditioned degree of probability* (Baayen, 2009)

\[ P = \frac{V(1,C,N)}{N(C)} \]

- Represents the ratio of the number of hapax legomena of a category to the total number of tokens of a category within a corpus
- (Tricky to make statistical comparisons)
If syntactic context puts a category member at a disadvantage in production

- Speakers might use the morphological category less in that context
- This can result in less usage of that category
- Final effect: lower productivity for that category
Incongruent Dutch diminutives are at a disadvantage when produced after definite determiners

- Speakers might use incongruent diminutives less in that context than congruent diminutives

- This might cause incongruent diminutives to be used less often than congruent diminutives

- Final effect: lower productivity for incongruent diminutives compared to congruent ones
Incongruent Dutch diminutives are at a disadvantage when produced after definite determiners
  ▶ Speakers might use incongruent diminutives less in that context than congruent diminutives
    ▶ Except they don’t
  ▶ This might cause incongruent diminutives to be used less often than congruent diminutives
  ▶ Final effect: lower productivity for incongruent diminutives compared to congruent ones
  ▶ Potential productivity for incongruent diminutives: $7.1 \times 10^{-2}$
  ▶ Potential productivity for congruent diminutives: $3.9 \times 10^{-2}$
Perhaps the effect is not present cumulatively, but does show up in the relevant syntactic context:
Perhaps the effect is not present cumulatively, but does show up in the relevant syntactic context:

- Potential productivity for incongruent diminutives after *het*: $5.4 \times 10^{-2}$
- Potential productivity for congruent diminutives after *het*: $4.2 \times 10^{-2}$