German Synthetic Compounds and the Architecture of the Grammar: A Behavioral Analysis

Livio Gaeta
Dipartimento di Filologia Moderna “Salvatore Battaglia”, Università degli Studi di Napoli Federico II

Mediterranean Morphology Meeting 8, Cagliari, 14-17 September 2011

Synthetic Compounds and their Behavior

- Synthetic compounds (SCs, German Rekombinationskomposita) are compounds in which the modifier saturates an argument of the head (Roepfer & Siegel 1978, Gaeta 2010), usually as a result of deverbal nominalization:

X fährt ein Auto  ‘X drives a car’
Auttfahrer  ‘car driver’

- Main questions:
  - Can SCs simply be derived from VPs (syntax below zero, see Spencer 2005) or are they an independent construction (Scalise & Guevara 2005)?
  - Can the selectional behavior of deverbal SCs in usage be predicted from that of corresponding VPs?
  - Focus on German agent nominalizations in -er (see Meibauer et al. 2004)


- Are the same objects attested? With similar frequency?
- Are there heads which prefer one pattern over the other?
- Does having many VP objects mean having many SCs?
- Is productivity as a VP head and as an SC head correlated?

Methodology

- Extract transitive VPs & SCs in -er from large corpus (deWaC, Baroni et al. 2009, ~1.7G tokens):
  - Use conservative patterns (verb final VPs with conjunction, subject, object compatible article not following a preposition)
  - Match verb as substring of compound
  - Correct for metathesis, Umlaut (Träger: tragen, Sammler: sammeln)
- Three groups of lexeme pairs are extracted:

- Group 1 – VP attestation only
  - Many very frequent VPs have no corresponding SC
  - Most cases can be divided into 3 groups:
    - Idiomatized phrase with preferred syntactic realization
    - Nominalization of head is avoided
    - Head nominalization has a different sense

<table>
<thead>
<tr>
<th>VP</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brauch machen</td>
<td>Verbrauch machen</td>
</tr>
<tr>
<td>Seilerei machen</td>
<td>Seilerei machen</td>
</tr>
<tr>
<td>Künstler machen</td>
<td>Künstler machen</td>
</tr>
<tr>
<td>Stadioner machen</td>
<td>Stadioner machen</td>
</tr>
<tr>
<td>Halbjunger machen</td>
<td>Halbjunger machen</td>
</tr>
</tbody>
</table>

- Group 2 – VP and SC attestation
  - Here we find a gradient from syntactic to morphological preference (sorted by ratio SC/VP):
    - Highly lexicalized but transparent compounds
    - Balanced attestation, including collocated AND lexicalized cases
    - More compositional but highly collocated idioms
  - However, there is no significant correlation between SC and VP attestation for each lexeme pair (r=0.0007, p=0.0)

- Group 3 – SC attestation only
  - Lexicalizations (Krankheitserreger ‘pathogen, lit. disease exciter’)
  - Suppletion (‘Unterrichter/Lehrer ‘teacher’, ‘Haber/Besitzer ‘owner’)
  - Metonymy / ellipsis (Erolitersteller ‘erotics-manufacturer’)
  - Archaisms (Staubsaugerverleih ‘vacuumcleaner sales rep’)

<table>
<thead>
<tr>
<th>SC</th>
<th>V</th>
<th>SCs only (Group 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbeitnehmer</td>
<td>Arbeitnehmer</td>
<td>10467</td>
</tr>
<tr>
<td>Wahrheitserzähler</td>
<td>Arbeitnehmer</td>
<td>115:135183</td>
</tr>
<tr>
<td>Vogelbeobachter</td>
<td>Vogelbeobachter</td>
<td>11:77</td>
</tr>
</tbody>
</table>

Productivity and Generation of Novel SCs

- The established lexeme types in Groups 1-3 may be lexicalized, and different lexicalizations for SC and VP may occur
- But if SCs are derived from VPs we expect productive behavior to correlate (non-lexicalized cases)
- Novel SCs should be based on VPs
- We use Baayen’s (2001) morphological productivity paradigm
  - Compare type frequency (V) and proportion of hapax legomena (HL, forms with frequency=1) for each head lexeme

<table>
<thead>
<tr>
<th>SC</th>
<th>V</th>
<th>SCs only (Group 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbeitnehmer</td>
<td>Arbeitnehmer</td>
<td>10467</td>
</tr>
</tbody>
</table>

- Correlation of type frequencies is fairly weak – many stems are much more prolific in object selection either as SCs or as VP
- Similarly, many heads have mainly VP-independent hapax SCs:

<table>
<thead>
<tr>
<th>SC head</th>
<th>V (VP)</th>
<th>SC head</th>
<th>f(SC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hersteller</td>
<td>1110</td>
<td>92</td>
<td>0.083416</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>1057</td>
<td>91</td>
<td>0.08425</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>802</td>
<td>178</td>
<td>0.22194</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>716</td>
<td>136</td>
<td>0.18994</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>664</td>
<td>71</td>
<td>0.10693</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>568</td>
<td>57</td>
<td>0.10035</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>392</td>
<td>30</td>
<td>0.07653</td>
</tr>
<tr>
<td>Arbeitnehmer</td>
<td>344</td>
<td>1</td>
<td>0.00290</td>
</tr>
</tbody>
</table>

Conclusion

- Lexical usage of SCs and VPs is different and unpredictable
- Constructional preferences, e.g. habitual/professional as SCs (Leiter ‘leader’, Sammler ‘collector’), others as VPs (sehen ‘see’, sagen ‘say’)
- Often little or no correlation of vocabulary size, productive behavior
- Frequent SC heads motivate novel SCs in same pattern, not extant VPs with same lexemes (cf. Construction Morphology, Booij 2010)
- Well-behaved exceptions confirm importance of lexical patterns: lexicalizations, head blocking, metonymy and partial suppletion
- More work needed on exhaustive classification of references

References