The Role of the Head in the Interpretation of English Deverbal Compounds

Gianina Iordăchioaia\textsuperscript{i}, Lonneke van der Plas\textsuperscript{ii}, Glorianna Jagfeld\textsuperscript{i}
(\textit{Universität Stuttgart}\textsuperscript{i}, \textit{University of Malta}\textsuperscript{ii})

Wen wurmt der Ohrwurm? – An interdisciplinary, cross-lingual perspective on the role of constituents in multi-word expressions
39. DGfS, Universität des Saarlandes, Saarbücken, 8.-10. März 2017
Deverbal (DCs) vs. Root Compounds (RCs)

- N-N compounds that are interpreted on the basis of a relationship between the head and the non-head;
- RCs are headed by lexical nouns (usually non-derived); the relationship is determined by world knowledge or context:
  1. fireman, train station vs. book chair, chocolate box
- DCs are headed by deverbal Ns; the relationship is often identified to the one between the base verb and the non-head:
  2. snow removal < to remove (the) snow (OBJ)
     police questioning < the police questions somebody (SUBJ)
     safety instruction < to instruct somebody on safety (OTHER)
- Even DCs are often hard to interpret, in spite of the verbal base and especially due to the ambiguity of the deverbal noun head:
  3. marketing approval, committee assignment, security assistance
Argument Structure Nominals (ASNs) vs. Result Nominals (RNs)

- Grimshaw (1990): Deverbal Ns are ambiguous between compositional V-like ASN-readings and more lexicalized RN-readings:
  4. a. The examination/exam was on the table. (RN)
     b. The examination of the patients took a long time/*was on the table. (ASN).

- ASNs vs. RNs (presence/absence of event structure):

<table>
<thead>
<tr>
<th>Property</th>
<th>ASN-reading</th>
<th>RN-reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obligatory internal arguments</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Agent-oriented modifiers (careful, deliberate, intentional)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>By-phrases are arguments</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Aspectual in/for-X-time adverbials</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Frequent, constant appear with the singular form</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Must appear in the singular</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

(adapted from Alexiadou & Grimshaw 2008: 3, citing Grimshaw 1990; see Appendix-1 for details)
The Linguistic Debate on DCs

• **Grimshaw (1990):** DCs ~ ASNs: DCs obey AS-constraints; only lowest argument (Theme/OBJ) is possible (Agent<Goal<Theme):
  5. gift-giving to children - *child-giving of gifts (to give gifts to children)
     book-reading by students - *student-reading of books (Students read books)

• Cf. RCs (e.g., compounds headed by zero-derived nominals):
  6. bee sting; dog bite (vs. *bee-stinging, *dog-biting)

• **Borer (2013):** DCs = RCs; DCs have no AS or event structure:
  7. a. the house demolition (*by the army) (*in two hours) (DC)
     b. the demolition of the house by the army in two hours (ASN)

• As in RCs, non-heads are context-dependent: Agent/SUBJ is OK:
  8. teacher recommendation; court investigation; government decision
Contribution of this Talk

- **Hypothesis**: If a noun is used more like an ASN or a RN, this should be preserved in compounds => ASN-like nouns head DCs with OBJ/int. argument, RN-like nouns form RCs with context-dependent readings:

  9. snow_{OBJ}/waste_{OBJ} removal vs. health_{OBJ}/flood_{OTHER} insurance
     drug_{OBJ}/child_{OBJ} trafficking body_{OBJ}/protest_{OTHER}/student_{SUBJ} movement

- **Our study**: a balanced collection of DCs automatically extracted from the Annotated Gigaword Corpus (Napoles et al. 2012)

- Use machine learning techniques to check which morphosyntactic properties of DC heads are relevant for the (OBJ-NOBJ) interpretation of DCs and what correlations we find between the two

- Our results provide support for Grimshaw's analysis and our hypothesis that DCs headed by ASN-like nouns receive OBJ readings
Outline

1) Our Methodology: Data Extraction and Annotation
2) Verification by Machine Learning Techniques
3) Discussion of Results
4) Conclusion and Future Plans
Outline

1) Our Methodology: Data Extraction and Annotation
2) Verification by Machine Learning Techniques
3) Discussion of Results
4) Conclusion and Future Plans
Our Plan

- Test if heads of DCs are more like ASNs or RNs in the corpus
- **Hypothesis**: DCs ≠ RCs
  Two types of *compounds headed by ASN/RN-like deverbal Ns*:
  - **True DCs**: non-head = only internal argument (OBJ)
  - **RCs**: non-head = ext. arg. (SUBJ); OTHER; int. arg. (OBJ)
- **Expectation** to test:
  - Correlation between **ASN-properties in heads** of DCs and an **OBJ interpretation** of the DC
- **Corpus and Tools**: see details in Appendix-2
Procedure

1) We created a frequency-balanced list of 25 heads for each of the suffixes -ing, -ion, -al, -ance, -ment (see Appendix-3)
Procedure

1) We created a frequency-balanced list of 25 heads for each of the suffixes -ing, -ion, -al, -ance, -ment (see Appendix-3)

2) We then extracted the 25 most frequent compounds that they appeared as heads of => a total of 3111 compounds
Procedure

1) We created a frequency-balanced list of 25 heads for each of the suffixes -\textit{ing}, -\textit{ion}, -\textit{al}, -\textit{ance}, -\textit{ment} (see Appendix-3)

2) We then extracted the 25 most frequent compounds that they appeared as heads of \(\Rightarrow\) a total of 3111 compounds

3) Annotate each compound's interpretation: OBJ, SUBJ, OTHER
3) Annotation of Compounds

- Two trained annotators (native speakers of American English)
- **Annotate** the relation between head and non-head:
  - **SUBJ**: ext. Arg. (*police* questioning, *designer* creation)
  - **OTHER** (*contract* killing, *safety* instruction)
  - **ERROR** (PoS tag errors or uninterpretable compounds: e.g. *face*<sub>V</sub> abandonment, *fond*<sub>A</sub> remembrance, percent assurance)
- Allow for ambiguity & preference order: SUBJ – OBJ, SUBJ > OBJ
- Post-processing (Appendix-4) => binary classification OBJ-NOBJ
- Simple interannotator agreement after post-processing: 81.5%
- **Result**: 2399 DCs: 1502 OBJ - 897 NOBJ
Procedure

1) We created a frequency-balanced list of 25 heads for each of the suffixes -ing, -ion, -al, -ance, -ment (see Appendix-3)

2) We then extracted the 25 most frequent compounds that they appeared as heads of => a total of 3111 compounds

3) Annotate each compound's interpretation: OBJ, SUBJ, OTHER

4) Determine ASN vs. RN properties of heads based on some of Grimshaw's (1990) tests by extracting contexts from the Gigaword
4) Morphosyntactic Features to Test

- 2. - 4. are Grimshaw's ASN-properties; 3. is the crucial one!
- 5. & 6. - comparable properties when the head is part of DCs

<table>
<thead>
<tr>
<th>Feature label</th>
<th>Description and illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. suffix</td>
<td>The suffix of the head noun: AL (rental), ANCE (insurance), ING (killing), ION (destruction), MENT (treatment)</td>
</tr>
<tr>
<td>2. sg_head_outside_DC</td>
<td>Percentage of the head’s occurrences as singular outside compounds.</td>
</tr>
<tr>
<td>3. sg_head+of_outside_DC</td>
<td>Percentage of the head’s occurrences as singular outside compounds which realize a syntactic relation with an <em>of</em>-phrase. (e.g., assignment of problems).</td>
</tr>
<tr>
<td>4. sum_adjectives</td>
<td>Percentage of the head’s occurrences in a modifier relation with one of the adjectives frequent, constant, intentional, deliberate, or careful.</td>
</tr>
<tr>
<td>5. sg_head_inside_DC</td>
<td>Percentage of the head’s occurrences as singular inside compounds.</td>
</tr>
<tr>
<td>6. sg_head+by_inside_DC</td>
<td>Percentage of the head’s occurrences as singular inside compounds which realize a syntactic relation with a <em>by</em>-phrase. (e.g., task assignment by teachers)</td>
</tr>
<tr>
<td>7. head_in_DC</td>
<td>Percentage of the head’s occurrences within a compound out of its total occurrences in the corpus.</td>
</tr>
</tbody>
</table>
Outline

1) Our Methodology: Data Extraction and Annotation
2) Verification by Machine Learning Techniques
3) Discussion of Results
4) Conclusion and Future Plans
Logistic Regression for Data Analysis

- **Questions** for the experiments:
  1) Can the head's **ASN-properties** help in predicting the meaning of DCs (**OBJ** or **NOBJ**)?
  2) Which properties are the **strongest predictors**?
- 7 independent variables (one categorical: suffix)
- Categorical dependent variable (**OBJ-NOBJ**)
- Split up data so that no head in test data is seen in training
- Balanced data set for two classes (by removing **OBJ** instances)
- **Data used:** 1614 training, 180 test compounds
Results in Ablation Experiments

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except \textit{sg_head_outside_DC}</td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except \textit{sum_adjectives}</td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except \textit{sg_head_inside_DC}</td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except \textit{head_in_DC}</td>
<td>46.7%†</td>
</tr>
<tr>
<td>All features, except \textit{sg_head_plus_of_outside_DC}</td>
<td>56.1%†</td>
</tr>
<tr>
<td>All features, except \textit{suffix}</td>
<td>61.7%†</td>
</tr>
<tr>
<td>All features, except \textit{sg_head_plus_by_inside}</td>
<td>71.1%†</td>
</tr>
<tr>
<td>\textit{Head_in_DC, sg_head_plus_of_outside_DC, and suffix combined}</td>
<td>76.1%†</td>
</tr>
</tbody>
</table>

† indicates a statistically significant difference from the performance when all features are included.
Answers to our Questions

1) Are the features predictive? **YES** – cf. random baseline: 66.7% vs. 50%; best performance: 76.1% vs. 50% (see Appendix-5 & 6)

2) Which features are strongest?
   - **Head_in_DC**: how often a head noun appears within a compound out of its total occurrences in the corpus
   - **Sg_head+of_outside_DC**: how often a head noun (in the singular) realizes an *of*-phrase outside compounds

<table>
<thead>
<tr>
<th>Features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except <em>sg_head_outside_DC</em></td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except <em>sum_adjectives</em></td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except <em>sg_head_inside_DC</em></td>
<td>66.7%</td>
</tr>
<tr>
<td>All features, except <em>head_in_DC</em></td>
<td>46.7%†</td>
</tr>
<tr>
<td>All features, except <em>sg_head+of_outside_DC</em></td>
<td>56.1%†</td>
</tr>
<tr>
<td>All features, except <em>suffix</em></td>
<td>61.7%†</td>
</tr>
<tr>
<td>All features, except <em>sg_head+by_inside</em></td>
<td>71.1%†</td>
</tr>
<tr>
<td><em>Head_in_DC, sg_head+of_outside_DC</em>, and suffix combined*</td>
<td>76.1%†</td>
</tr>
</tbody>
</table>
Outline

1) Our Methodology: Data Extraction and Annotation
2) Verification by Machine Learning Techniques
3) Discussion of Results
4) Conclusion and Future Plans
Head_in_DC (46.7% vs. 66.7%)

- High percentage of occurrences of a head inside compounds
- It indicates an **OBJ interpretation** (see Appendix-6)
  - Not related to ASN-hood and not mentioned in previous literature
  - High compoundhood of a head noun indicates its specialization for compounds
  - The fact that it correlates with an OBJ reading shows us that **if a deverbal noun typically forms a compound with one of its arguments, then this argument will be the object**
- This supports Grimshaw’s claim that DCs embed event structure with internal arguments
## Head\_in\_DC: Examples

<table>
<thead>
<tr>
<th>Head noun</th>
<th>Head_in_DC</th>
<th>OBJ-reading</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>laundering</em></td>
<td>94.80%</td>
<td>95.45%</td>
</tr>
<tr>
<td><em>mongering</em></td>
<td>91.77%</td>
<td>100%</td>
</tr>
<tr>
<td><em>growing</em></td>
<td>68.68%</td>
<td>95.23%</td>
</tr>
<tr>
<td><em>trafficking</em></td>
<td>61.99%</td>
<td>100%</td>
</tr>
<tr>
<td><em>enforcement</em></td>
<td>53.68%</td>
<td>66.66%</td>
</tr>
<tr>
<td><em>insurance</em></td>
<td>43.73%</td>
<td><strong>46.15%</strong></td>
</tr>
<tr>
<td><em>chasing</em></td>
<td>44.74%</td>
<td>90%</td>
</tr>
<tr>
<td><em>rental</em></td>
<td>42.95%</td>
<td>87.5%</td>
</tr>
<tr>
<td><em>acquittal</em></td>
<td>1.80%</td>
<td>12.5%</td>
</tr>
<tr>
<td><em>ignorance</em></td>
<td>0.85%</td>
<td>0%</td>
</tr>
<tr>
<td><em>refusal</em></td>
<td>0.77%</td>
<td>43.75%</td>
</tr>
<tr>
<td><em>anticipation</em></td>
<td>0.70%</td>
<td>37.5%</td>
</tr>
<tr>
<td><em>defiance</em></td>
<td>0.64%</td>
<td>35.29%</td>
</tr>
</tbody>
</table>

Heads with most/least frequent occurrence in compounds; outliers in bold
\textit{Sg\_head+of\_outside\_DC} (56.1\% vs. 66.7\%)

- The presence of an \textit{of}-phrase realizing the internal argument of the head/verb (cf. \textit{the examination of the patient})

- It predicts an OBJ reading (see Appendix-6)
  - In Grimshaw (1990), the realization of the internal argument is most indicative of the ASN status of a deverbal noun.

- This proves our hypothesis to be right: high ASN-hood of the head $\Rightarrow$ OBJ reading in compound
  - Precision & recall in the extraction of \textit{of}-phrases is pretty good:
    - \textbf{Precision}: 90.96
    - \textbf{Recall}: 90.08
### $Sg_{\text{head}}+\text{of}_\text{outside}_DC$: Examples

<table>
<thead>
<tr>
<th>Head noun</th>
<th>Of-phrases</th>
<th>OBJ-reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>creation</td>
<td>80.51%</td>
<td>72.72%</td>
</tr>
<tr>
<td>avoidance</td>
<td>70.40%</td>
<td>100%</td>
</tr>
<tr>
<td>obstruction</td>
<td>65.25%</td>
<td>90.47%</td>
</tr>
<tr>
<td>removal</td>
<td>63.53%</td>
<td>92%</td>
</tr>
<tr>
<td>breaking</td>
<td>58.83%</td>
<td>94.11%</td>
</tr>
<tr>
<td>abandonment</td>
<td>55.90%</td>
<td>90%</td>
</tr>
<tr>
<td>assassination</td>
<td>52.27%</td>
<td><strong>11.76%</strong></td>
</tr>
<tr>
<td>preservation</td>
<td>52.14%</td>
<td>100%</td>
</tr>
<tr>
<td>education</td>
<td>1.81%</td>
<td>30%</td>
</tr>
<tr>
<td>proposal</td>
<td>1.08%</td>
<td><strong>76.19%</strong></td>
</tr>
<tr>
<td>counseling</td>
<td>0.53%</td>
<td>10%</td>
</tr>
<tr>
<td>insurance</td>
<td>0.42%</td>
<td>46.15%</td>
</tr>
<tr>
<td>mongering</td>
<td>0%</td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Heads with (in)frequent of-phrases outside compounds; outliers in bold
Sg_head+by_inside_DC (71.1% vs. 66.7%)

- Frequency of a by-phrase (i.e., ext. argument) with a compound
- It is noisy – results improve when feature is dismissed
  - Grimshaw (1990): book-reading by students
  - Borer (2013): the house demolition (*by the army)
- Possible interferences:
  - by is ambiguous between ext. arg. and 'author'-by: e.g., a book by Chomsky => in principle, both ASNs and RNs should be OK
  - Precision 85.02 & recall 72.78 in our by-phrase extractions
  - Further investigation is needed
Outline

1) Our Methodology: Data Extraction and Annotation
2) Verification by Machine Learning Techniques
3) Discussion of Results
4) Conclusions and Future Plans
Conclusions

- Heads of DCs are ambiguous between ASNs and RNs and this influences the interpretation of DCs.
- We find two correlations:
  - realization of internal arguments as of-phrases and OBJ readings
  - high compoundhood and OBJ readings
- These support Grimshaw's claim that DCs include event structure with internal arguments.
- The by-phrase in compounds is a noisy feature – this may be due to its ambiguity.
- Suffixes: see Appendix-7.
Future Plans

- Add third annotator (majority vote)
- Add annotation feature *result* (RN) vs. *process* (ASN) (1 to 5)
- We extracted the base verbs and their objects/subj ects – check whether:
  - the high frequency of a direct object with a verb correlates with an OBJ reading of the DCs
  - the non-heads that appear in DCs correlate with the objects/subjects of the verb – close to Borer's (2013) suggestions
- Would descriptive statistics be able to explain the correlations in our data better than ML techniques?
Acknowledgments

• Annotators: Katherine Fraser & Whitney Frazier Peterson
• Technical support from the SFB 732 INF-project – thanks to Kerstin Eckart
• Alla Abrosimova helped with other technical details
• Research funded by the DFG for the projects B1 – The form and interpretation of derived nominals – and D11 – A Crosslingual Approach to the Analysis of Compound Nouns – as part of the SFB 732 at the University of Stuttgart
Appendix
Appendix-1: ASNs vs. RNs (Grimshaw 1990)

- Arguments are introduced by verbs via their event structure (aspectual properties, argument licensing, verbal properties)
- ASNs preserve event structure & AS from verbs; RNs do not
- ASN: obligatory internal arguments (vs. RNs) (Grimshaw 1990: 50-52)

(7) a. The assignment is to be avoided. (RN)
   b. *The constant assignment is to be avoided. (ASN-RN)
   c. The constant assignment of unsolvable problems is to be avoided. (ASN)

- *Constant and frequent are aspectual modifiers when they appear with a singular noun => they require event structure (7b, c); if the noun is plural, it can be a RN:

(9) The constant assignments were avoided by the students. (RN)
Appendix-1: ASNs vs. RNs (Grimshaw 1990)

- *Intentional, deliberate, careful* are agent-oriented modifiers and only appear with event structure => ASNs but not RNs

(11) a. *The instructor's intentional examination took a long time.*
    b. The instructor's intentional examination of the papers took a long time.

- ASNs reject plural (not nominal enough) vs. RNs (Grimshaw 1990: 54)

(18) a. The assignments were long. (RN)
    b. *The assignments of the problems took a long time.* (ASN)
Appendix-2: Corpus and Tools

- The Annotated Gigaword Corpus (Napoles et al. 2012) – LDC Catalog No. LDC2012T21
- 10-million documents from seven news outlets
- Total of more than 4-billion words
- Automatic processing and annotation we use:
  1. Segmentation (using Splitta - Gillick, 2009) and tokenization (using Stanford's CoreNLP pipeline)
  2. Lemmatization and POS tags (Stanford's CoreNLP pipeline)
  3. Treebank-style constituent parse trees (Huang et al. 2010, Avg. F score = 91.4 on WSJ sec 22)
  4. Syntactic dependency trees (Using Stanford's CoreNLP pipeline for the conversion from constituency to dependency trees)
- We removed within-file (1010 files) duplicate sentences (170 >143 GB)
Appendix-3: Selection of Target Head Nouns

- For each suffix, we selected 25 nouns derived from transitive verbs, which head NN compounds (no N before or after) in Gigaword;
- *Arrival* – the only unaccusative verb

<table>
<thead>
<tr>
<th>Frequency</th>
<th>ING</th>
<th>ION</th>
<th>MENT</th>
<th>AL</th>
<th>ANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>spending building training bombing trafficking</td>
<td>production protection reduction construction consumption</td>
<td>enforcement development movement treatment punishment</td>
<td>proposal approval withdrawal arrival rental</td>
<td>insurance performance assistance clearance surveillance</td>
</tr>
<tr>
<td>Medium</td>
<td>killing writing counseling firing teaching</td>
<td>supervision destruction cultivation deprivation instruction</td>
<td>deployment replacement placement assignment adjustment</td>
<td>renewal burial survival denial upheaval</td>
<td>assurance disturbance dominance acceptance tolerance</td>
</tr>
<tr>
<td>Low</td>
<td>weighting baking chasing measuring mongering</td>
<td>demolition anticipation expulsion obstruction deportation</td>
<td>reinforcement realignment empowerment mistreatment abandonment</td>
<td>retrieval acquittal disapproval rebuttal dispersal</td>
<td>defiance reassurance endurance remembrance ignorance</td>
</tr>
</tbody>
</table>
Appendix-4: Post-processing of Annotations

- Initial database of 3111 compounds
- Conflate OTHER and SUBJ to NOBJ (=> binary classification)
- Remove errors (163)
- Remove disagreements (547)
- Remove true ambiguous cases (for both annotators) (2)
- DCs headed by *arrival*: SUBJ > OBJ (but we didn’t check alternating verbs – on our to do list)
- For ambiguous vs. unambiguous annotations, take overall preference (e.g., A1: NOBJ-OBJ; A2: NOBJ => NOBJ)
Appendix-5: Comparison to NLP Studies

- Our best performance: 76.1% vs. 50% => 26.1% improvement
- Previous work in the NLP literature targets state-of-the-art performance in prediction with methods different from ours
- Our purpose was to start from linguistic theory and test linguistic hypotheses
- These studies include more suffixes (-er, -ee) and zero-derived nouns; -er and -ee are biased, so they are more predictive;
- We had only 'event'-denoting suffixes, where SUBJ/OBJ are similarly conceivable
- Lapata (2002): 86.1% vs. 61.5% => 24.6% above the baseline
## Appendix-6: Predicted Interpretation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Class OBJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>[list of features and their values]</td>
<td>[table of values]</td>
</tr>
</tbody>
</table>

- The two most predictive features correlate with an OBJ-reading (see `head_in_DC, sg_head+of_outside_DC`).
- For the suffix feature we get some variation:
  - **Suffix**: `-ion, -al : OBJ`
  - `-ance, -ment, -ing : NOBJ`
Appendix-7: Suffixes (61.7% vs. 66.7%)

- It is the weakest predictive feature
- Grimshaw (1990): *ing*-nominals are always ASNs => OBJ
- Borer (2013): *ing* introduces the Originator (ext. arg.) itself and biases the DC towards an OBJ reading
  - Both theories predict a correlation between *ing* and OBJ, which we did not find
- Latinate suffixes (*-ion, -ment, -al, -ance*) are taken to behave similarly in theory, but we find a bias for OBJ in *-ion* and *-al*, and for NOBJ in *-ance* and *-ment*
- Further research is needed: both cleaner data on our side and linguistic research on the selectional preferences of suffixes