The Grammar of English Deverbal Compounds and their Meaning

Gianina Iordăchioaia
University of Stuttgart
Stuttgart, Germany
gianina
@ifla.uni-stuttgart.de

Lonneke van der Plas
University of Malta
Valletta, Malta
lonneke.vanderplas
@um.edu.mt

Glorianna Jagfeld
University of Stuttgart
Stuttgart, Germany
glorianna.jagfeld
@gmail.com

Abstract

We present an interdisciplinary study on the interaction between the interpretation of noun-noun deverbal compounds (DCs; e.g., *task assignment*) and the morphosyntactic properties of their deverbal heads in English. Underlying hypotheses from theoretical linguistics are tested with tools and resources from computational linguistics. We start with Grimshaw's (1990) insight that deverbal nouns are ambiguous between argument-supporting nominal (ASN) readings, which inherit verbal arguments (e.g., *the assignment of the tasks)*, and the less verbal and more lexicalized Result Nominal and Simple Event readings (e.g., *a two-page assignment*). Following Grimshaw, our hypothesis is that the former will realize object arguments in DCs, while the latter will receive a wider range of interpretations like root compounds headed by non-derived nouns (e.g., *chocolate box*). Evidence from a large corpus assisted by machine learning techniques confirms this hypothesis, by showing that, besides other features, the realization of internal arguments by deverbal heads outside compounds (i.e., the most distinctive ASN-property in Grimshaw 1990) is a good predictor for an object interpretation of non-heads in DCs.

1 Introduction

Deverbal compounds (DCs) are noun-noun compounds whose head is derived from a verb by means of a productive nominalizing suffix such as *-al*, *-ance*, *-er*, *-ion*, *-ing*, or *-ment*, and whose non-head is usually interpreted as an object of the base verb, as illustrated in (1). *Root compounds* differ from DCs in that they need not be headed by deverbal nouns and their interpretation may vary with the context.[1] For instance, a root compound like *chocolate box* may refer to a box with chocolate or one that has chocolate color etc, depending on the context, while others like *gear box* have a more established meaning. DCs have been at the heart of theoretical linguistic research since the early days of generative grammar precisely due to their special status between lexicon and grammar (Roeper and Siegel, 1978; Selkirk, 1982; Grimshaw, 1990; Ackema and Neeleman, 2004; Lieber, 2004; Borer, 2013, among others). As compounds, they are new lexemes, i.e., they should be part of the lexicon, and yet, their structure and interpretation retain properties from argument-supporting nominals (ASNs) and correlated verb phrases, which suggests that they involve some grammar (cf. (2)).

(1) house rent**al**, title insur**ance**, oven-clean**er**, crop destruct**ion**, drug traffick**ing**, tax adjust**ment**

(2) a. crop destruction – destruction of crops – to destroy crops
b. tax adjustment – adjustment of taxes – to adjust taxes

Rooted in the long debate on synthetic compounds (see Olsen (2015) for an overview), two types of analyses have been proposed to DCs. What we call *the grammar-analysis* posits a grammar component

[1] Grimshaw (1990) argues that compounds headed by zero-derived nouns, e.g., *bee sting*, *dog bite*, are root compounds, since they do not preserve verbal event structure properties. In this paper, we will propose that even some of the DCs that are headed by suffix-based deverbal nouns form root compounds if their heads lack verbal properties and exhibit a more lexicalized meaning: cf. ASN vs. Result Nominal interpretation and the discussion in Section 2.1.

Proceedings of the Workshop on Grammar and Lexicon: Interactions and Interfaces,
pages 81–91, Osaka, Japan, December 11 2016.

in their structure and draws correlations between DCs and ASNs (or VPs), as in (2) (Roeper and Siegel, 1978; Grimshaw, 1990; Ackema and Neeleman, 2004, most notably). *The lexicon-analysis* argues that DCs are just like root compounds and derive their meaning from the lexical semantics of the nouns, whether derived or not (e.g., Selkirk 1982, Lieber 2004). The various implementations take different theory-driven shapes, but the baseline for our study is the question whether DCs retain properties from ASNs (and, implicitly, VPs) or behave like root compounds.

The main property that DCs share with ASNs and VPs is the realization of an argument as the non-head, more precisely, the internal argument of the original verb, which in the VP usually appears as a direct object: see *crop* and *tax* in (2). Importantly, unlike ASNs, which may realize both external and internal arguments as prepositional/genitive phrases, DCs have a reduced structure made up of two nouns, which can host only one argument besides the head (see (3)).

(3) a. The **hurricane** destroyed the **crops**.
b. the destruction of the **crops** by the **hurricane**
c. **crop** destruction

For grammar-analyses it is crucial to argue that the argument realized in DCs must be the lowest one in the VP, namely, the internal/object argument (see the first sister principle of Roeper and Siegel (1978) and the thematic hierarchy in Grimshaw (1990)). Subject/external arguments are known to be introduced by higher/more complex event structure (Kratzer, 1996) and, given that DCs cannot accommodate the full event structure of the original verb, it must be the lowest component that includes the object. However, the main challenge to these analyses is that, indeed, we find DCs whose non-heads come very close to receiving a subject interpretation, as exemplified in (4). In view of this evidence, Borer (2013) offers a syntactic implementation of the lexicon-analysis, in which all DCs are root compounds and their (external or internal) argument interpretation should be resolved by the context.[2]

(4) **hurricane** destruction, **teacher** recommendation, **government** decision, **court** investigation

In this paper, we challenge Borer (2013) as the most recent lexicon-analysis by showing that high ASN-hood of a head noun predicts an object interpretation of the non-head. Our methodology draws on evidence from a large corpus of naturally occurring text. This realm of information is analysed using simple machine learning techniques. We designed extraction patterns tailored to the properties that are associated with ASNs to collect counts for the selected deverbal heads. These counts are used as features in a logistic regression classifier that tries to predict the covert relation between heads and non-heads. We find that a frequent realization of the internal argument (i.e., high ASN-hood) with particular deverbal nouns is indeed a good predictor for the object relation in compounds headed by these nouns. This confirms Grimshaw's claim and our hypothesis that DCs involve some minimal grammar of the base verb, which compositionally includes only the internal/object argument. Non-object readings are obtained with heads that do not preserve ASN-properties and implicitly build root compounds. The theoretical implication of this study is that compounds headed by deverbal nouns are ambiguous between real argumental DCs, as in (2), and root compounds, as in (4).

2 Previous work

In this section we introduce the linguistic background for our study and review previous work on interpreting deverbal compounds from the natural language processing literature.

2.1 Linguistic background

In this paper we build on two previous contributions to the theoretical debate on DCs: Grimshaw (1990) and Borer (2013). The former supports a grammar-analysis, the latter a lexicon-analysis.

[2] As is well-known, since Chomsky (1970), linguistic theories on word formation have been split between 'lexicalist' and 'syntactic'. The former assume that words have lexical entries and derivation is done by lexical rules, while the latter take word formation to follow general syntactic principles. This distinction is in fact orthogonal to the lexicon- vs. grammar-analyses of DCs that we refer to here, in that ironically Grimshaw (1990) is a lexicalist theory that posits grammar (i.e., event structure) in DCs, while Borer (2013) is a syntactic theory that denies the influence of any grammar principles in the make-up of DCs.

Grimshaw (1990) introduces a three-way distinction in the interpretation of deverbal nouns: Argument-supporting Nominals (ASNs; in her terminology, *complex event nominals*), Result Nominals, and Simple Event Nominals. The main difference is between ASNs and the other two categories in that only ASNs inherit event structure from the base verb and, implicitly, accommodate verbal arguments. Result Nominals are more lexicalized than ASNs; they refer to the result of the verbal action, often in the shape of a concrete object. Simple Event Nominals are also lexicalized but have an event/process interpretation like some non-derived nouns such as *event* or *ceremony*. The deverbal noun *examination* may receive a Result Nominal (RN) reading, synonymous to *exam*, when it doesn't realize arguments, as in (5a), or it may have an ASN reading, when it realizes the object, as in (5b). The predicate *was on the table* selects the RN reading, this is why, it is incompatible with the ASN. When the external/subject argument is realized in an ASN, the internal/object argument must be present as well (see (5c)); otherwise, the deverbal noun will receive a Result or Simple Event reading (see Grimshaw 1990: 53 for details).

(5) a. The examination/exam was on the table. (RN)
b. The examination **of the patients** took a long time/*was on the table. (ASN)
c. The (**doctor's**) examination of the patients (**by the doctor**) took a long time. (ASN)

Given that in Grimshaw's (1990) work Result Nominals and Simple Event Nominals pattern alike in not realizing argument structure and display similar morphosyntactic properties in contrast to ASNs, we refer to both as RNs, i.e., as more lexicalized and less compositional readings of the deverbal noun. Grimshaw (1990) enumerates several properties that distinguish ASNs from RNs, which we summarize in Table 1, a selection from Alexiadou and Grimshaw (2008). We come back to these properties in Section 3.4.

Property	**ASN-reading**	**RN-reading**
Obligatory internal arguments	Yes	No
Agent-oriented modifiers	Yes	No
By-phrases are arguments	Yes	No
Aspectual *in*/*for*-adverbials	Yes	No
Frequent, *constant* require plural	No	Yes
May appear in plural	No	Yes

Table 1: Morphosyntactic properties distinguishing between ASNs and RNs

Within this background, Grimshaw argues that DCs are headed by ASNs and fundamentally different from root compounds. In her approach, this means that the heads of DCs inherit event structure from the base verb, which accommodates argument structure like the ASNs in (5b-5c). Importantly, however, DCs have a reduced structure made up of two nouns, which entails that the head can realize only one argument inside the compound. In Grimshaw's approach, this predicts that the head inherits a reduced verbal event structure which should be able to accommodate only the lowest argument of the VP, namely, the object. In line with this reasoning and on the basis of her thematic hierarchy, Grimshaw argues that arguments other than themes (realized as direct objects) are excluded from DCs. These include both prepositional objects realizing goals or locations and subjects that realize external arguments, as illustrated in (6).

(6) a. **gift**-giving to children vs. ***child**-giving of gifts (to give gifts **to children**)
b. **flower**-arranging in vases vs. ***vase**-arranging of flowers (to arrange flowers **in vases**)
c. **book**-reading by students vs. ***student**-reading of books (**Students** read books.)

In her discussion of DCs, Grimshaw (1990) does not address the properties from Table 1 on DC heads to show that they behave like ASNs. Borer (2013) uses some of these properties to argue precisely against the ASN status of DC heads. We retain two of her arguments: the lack of aspectual modifiers and argumental *by*-phrases. Borer uses data as in (7) to argue that, unlike the corresponding ASNs, DCs disallow aspectual *in-*/*for*-adverbials and fail to realize *by*-phrases (contra Grimshaw's (6c)).

(7) a. the demolition of the house **by the army** *in two hours* (ASN)
b. the maintenance of the facility **by the management** *for two years* (ASN)
c. the house demolition (*__by the army__) (**in two hours*) (DC)
d. the facility maintenance (*__by the management__) (**for two years*) (DC)

For Borer, the unavailability of aspectual modifiers indicates that event structure is entirely missing from DCs, so they cannot be headed by ASNs. Her conclusion is that DCs are headed by RNs and behave like root compounds. Thus the object interpretation of their non-head is just as valid as a subject or prepositional interpretation, depending on the context of use. In support of this, she quotes DCs as in (4) above, whose non-heads are most likely interpreted as subjects.

In our study, we will show that the presence of *of*-phrases that realize the internal argument with head nouns outside compounds is a good predictor for an object interpretation of the non-head in DCs, supporting Grimshaw's approach. Moreover, the appearance of a *by*-phrase in DCs – which pace Borer (2013) is well attested in the corpus – seems to be harmful to our model, which shows us that the *by*-phrase test is not very telling for the structure and interpretation of DCs.[3]

2.2 Interpretation of DCs in natural language processing literature

Research on deverbal compounds (referred to with the term *nominalizations*) in the NLP literature has focused on the task of predicting the underlying relation between deverbal heads and non-heads. Relation inventories range from 2-class (Lapata, 2002) to 3-class (Nicholson and Baldwin, 2006), and 13-class (Grover et al., 2005), where the 2-class inventory is restricted to the subject and direct object relations, the 3-class adds prepositional complements, and the 13-class further specifies the prepositional complement.

Although we are performing the same task, our underlying aim is different. Instead of trying to reach state-of-the-art performance in the prediction task, we are interested in the contribution of a range of features based on linguistic literature, in particular, morphosyntactic features of the deverbal head. Features used in the NLP literature mainly rely on occurrences of the verb associated with the deverbal head and the non-head in large corpora. The idea behind this is simple. For example, a verb associated with a deverbal noun – such as *slaughter* from *slaughtering* – is often seen in a direct object relation with a specific noun, such as *animal*. The covert relation between head and non-head in *animal slaughtering* is therefore predicted to be direct object. To remedy problems related to data sparseness, several smoothing techniques are introduced (Lapata, 2002) as well as the use of Z-scores (Nicholson and Baldwin, 2006). In addition to these statistics on verb-argument relations, Lapata (2002) uses features such as the suffix and the direct context of the compound.

Apart from the suffix, the features used in these works are meant to capture encyclopaedic knowledge, usually building on lexicalist theoretical approaches that list several covert semantic relations typically available in compounds (cf. most notably, Levi 1978; see Fokkens 2007, for a critical overview). The morphosyntactic features we use are fundamentally different from the syntactic relations used in this NLP literature and described above (cf. also Rösiger et al. (2015), for German). Our features are head-specific and rely on insights from linguistic theories that posit an abstract structural correlation between DCs and the compositional event structure of the original verb, as mirrored in the behavior of the derived nominals (as ASNs or RNs).

In addition, our selection of DCs is different. We carefully selected a balanced number of DCs based on the suffixes *-al*, *-ance*, *-ing*, *-ion*, and *-ment* within three different frequency bands. These suffixes derive eventive nouns which should allow both object and subject readings of their non-heads in compounds, unlike *-ee* and *-er*, which are biased for one or the other. Moreover, previous studies also included zero-derived nouns, which we excluded because they mostly behave like RNs (Grimshaw, 1990).

3 Materials and methods

This section presents the corpora and tools we used, the methods we adopted for the selection of DCs, the annotation effort, the feature extraction, as well as the machine learning techniques we employed.

[3] We haven't included aspectual adverbials in our study for now, because acceptability judgements on ASNs usually decrease in their presence and we expect them to be very rare in the corpus.

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

Table 2: Samples of the suffix-based selection of deverbal head nouns

3.1 Corpus and tools

For the selection of DCs and to gather corpus statistics on them, we used the Annotated Gigaword corpus (Napoles et al., 2012) as one of the largest general-domain English corpora that contains several layers of linguistic annotation. We used the following available automatic preprocessing tools and annotations: sentence segmentation (Gillick, 2009), tokenization, lemmatization and POS tags (Stanford's CoreNLP toolkit[4]), as well as constituency parses (Huang et al., 2010) converted to syntactic dependency trees with Stanford's CoreNLP toolkit. The corpus encompasses 10M documents from 7 news sources and more than 4G words. As news outlets often repeat news items in subsequent news streams, the corpus contains a substantial amount of duplication. To improve the reliability of our corpus counts, we removed exact duplicate sentences within each of the 1010 corpus files, resulting in a 16% decrease of the corpus size.

3.2 Selection of deverbal compounds

We selected a varied yet controlled set of DCs from the Gigaword corpus. We first collected 25 nouns (over three frequency bands: low, medium, and high) for each of the suffixes *-ing*, *-ion*, *-ment*, *-al* and *-ance*. These suffixes usually derive both ASNs and RNs, unlike zero-derived nouns like *attack*, *abuse*, which, according to Grimshaw (1990), mostly behave like RNs. We excluded nouns based on the suffixes *-er* and *-ee* because they denote one participant (subject, respectively, object) of the verb, implicitly blocking this interpretation on the non-head (cf. $police_{subj}$ *train***ee** – dog_{obj} *train***er**). The base verbs were selected to allow transitive uses, i.e., both subjects and objects are in principle possible.[5] For illustration, Table 2 offers samples of five deverbal nouns per each frequency range and suffix. For each such selected noun we then extracted the 25 most frequent compounds that they appeared as heads of, where available.[6] After removing some repetitions we ended up with a total of 3111 DCs.

3.3 Annotation effort

We had all DCs annotated by two trained American English speakers, who were asked to label them as OBJ, SUBJ, OTHER, or ERROR, depending on the relation that the DC establishes between the verb from which its head noun is derived and the non-head. For instance, DCs such as in (2) would be labeled as OBJ, while those in (4) would be SUBJ. OTHER was the label for prepositional objects (e.g., *adoption*

[4]http://nlp.stanford.edu/software/corenlp.shtml

[5]*Arrive* is the only intransitive verb we included, but it is unaccusative, so it realizes an internal argument.

[6]Some deverbal nouns were not so productive in compounds and appeared with fewer than 25 different non-heads. To ensure cleaner results, in future work we aim to balance the dataset for the number of compounds per each head noun.

counseling 'somebody counsels somebody on adoption') or attributive uses in compounds (e.g., *surprise arrival* 'an arrival that was a surprise'). ERROR was meant to identify the errors of the POS tagger (e.g., *face abandonment* originates in 'they face$_V$ abandonment'). We allowed the annotators to use multiple labels and let them indicate the preference ordering (using '>') or ambiguity (using '–').

We used the original annotations from both annotators to create a final list of compounds and the labels that they both agreed on. We kept the multiple labels for ambiguous cases and selected only the preferred reading, when there was one. We labeled them as OBJ, NOBJ, DIS(agreement between annotators), AMBIG, or ERROR. If one annotator indicated ambiguity and the other selected only one of the readings, we selected the reading available to both. We found only two cases of ambiguity where both annotators agreed. We removed the fully ambiguous DCs together with the 163 errors and the 547 cases of disagreement.[7] This left us with 2399 DCs that we could process for our purposes. To allow for multi-class and binary classification of the data, we kept two versions: one in which OTHER and SUBJ were conflated to NOBJ, and one in which we kept them separate. In this paper, we focus on the binary classification. The resulting data set was skewed with OBJ prevailing: 1502 OBJ vs. 897 NOBJ.

3.4 Extracting features for ASN-hood

To determine how the ASN-hood of DC heads fares in predicting an OBJ or NOBJ interpretation of the non-head we constructed 7 indicative patterns mostly inspired by Grimshaw's ASN properties in Table 1, for which we collected evidence from the Gigaword corpus. They are all summarized in Table 3 with illustrations.

The feature *suffix* is meant to show us whether a particular suffix is prone to realizing an OBJ or NOBJ relation in compounds (Grimshaw, for instance, was arguing that *ing*-nominals are mostly ASNs, predicting a preference for OBJ in DCs). The features from 2 to 4 are indicative of an ASN status of the head noun when it occurs outside compounds. As shown in Table 1, Grimshaw argued that ASNs appear only in the singular (see feature 2. *percentage_sg_outside*). In the same spirit, we counted the frequency of *of*-phrases (feature 3) when the head is in the singular. The occurrence with adjectives is very infrequent, so we counted them all together under feature 4. *sum_adjectives*. For *frequent* and *constant* we again counted only the singular forms, given that Grimshaw argued that these may also combine with RNs in the plural. As agent-oriented modifiers, *intentional*, *deliberate*, and *careful* are only expected with ASNs. The features 5. *percentage_sg_inside* and 6. *percentage_by_sg_inside* test ASN-hood of the head noun when appearing inside compounds. Remember that Grimshaw documented *by*-phrases in DCs (see (6c), contra Borer's (7c-7d)). We didn't consider the parallel realization of *of*-phrases *inside* compounds, since there is no theoretical claim on them and they would have produced too much noise, given that the object argument that *of*-phrases realize would typically appear as a non-head in DCs.[8]

The rationale behind checking the ASN-properties of DC heads when they appear outside and inside DCs is that deverbal nouns exhibit different degrees of ambiguity between ASNs and RNs with tendencies towards one or the other, which would be preserved in compounds and reflected in the OBJ/ NOBJ interpretation. Our last individual feature 7. *percentage_head_in_DC* indicates the frequency of the head in compounds and is meant to measure whether the noun heads that appear very frequently in compounds exhibit any preference for an OBJ or NOBJ relation. If such a relation associates with a high preference of the head to appear in compounds, this tells us that this relation is typical of DCs.

3.4.1 Technical details of the feature extraction method

For the *inside_compound* features we collected the compounds by matching the DCs in the corpus with the word form of the non-head and the lemma of the head and required that there be no other directly preceding or succeeding nouns or proper nouns. Conversely, the *outside_compound* features apply to head nouns appearing without any noun or proper noun next to them. We determined the grammatical number of a noun (compound) by its POS tag (the POS tag of its head).

[7] We will base our studies on the agreed-upon relations only. However, for the sake of completeness and for showing that the task is well-defined we computed the simple inter-annotator agreement excluding the error class and the negligible class for ambiguous cases. It amounts to 81.5%.

[8] We also discarded *by*-phrases outside compounds, since they would have to be considered in combination with an *of*-phrase (Grimshaw, 1990, cf. (5c)), and we wanted to focus on individual features in relation to the head noun.

Feature label	Description and illustration
1. *suffix*	The suffix of the head noun: AL (rent**al**), ANCE (insur**ance**), ING (kill**ing**), ION (destruct**ion**), MENT (treat**ment**)
2. *percentage_sg_outside*	Percentage of the head's occurrences as singular outside compounds.
3. *percentage_of_sg_outside*	Percentage of the head's occurrences as singular outside compounds which realize a syntactic relation with an *of*-phrase. (e.g., *assignment* **of** *problems*).
4. *sum_adjectives*	Percentage of the head's occurrences in a modifier relation with one of the adjectives *frequent*, *constant*, *intentional*, *deliberate*, or *careful*.
5. *percentage_sg_inside*	Percentage of the head's occurrences as singular inside compounds.
6. *percentage_by_sg_inside*	Percentage of the head's occurrences as singular inside compounds which realize a syntactic relation with a *by*-phrase. (e.g., task assignment **by** teachers)
7. *percentage_head_in_DC*	Percentage of the head's occurrences within a compound out of its total occurrences in the corpus.

Table 3: Indicative features

We counted a noun (or DC) as being in a syntactic relation with an *of-phrase* or *by-phrase*, if it (respectively, its head) governs a collapsed dependency labeled 'prep_of'/'prep_by'. As we were interested in prepositional phrases that realize internal, respectively, external arguments, but not in the ones appearing in temporal phrases (e.g., 'by Monday') or fixed expressions (e.g., 'by chance'), we excluded phrases headed by nouns that typically appear in these undesired phrases. We semi-automatically compiled these lists based on a multiword expression lexicon[9] and manually added entries. The lists comprise 161 entries for *by*-phrases and 53 for *of*-phrases. To compute the feature *sum_adjectives* we counted how often each noun appearing outside compounds governs a dependency relation labeled 'amod' where the dependent is an adjective (POS tag 'JJ') out of the lemmas *intentional, deliberate, careful, constant,* and *frequent*.

3.5 Using Machine Learning techniques for data exploration

The features listed in Table 3 are a mix of numerical (2 to 7) and categorical features (1). The dependent variable is a categorical feature that varies between one of the two annotation labels, OBJ and NOBJ. Thus, in order to test our hypotheses that the features in Table 3 are useful to predict the relation between the deverbal head and the non-head, we trained a Logistic Regression classifier[10] using these features.

The resulting model was tested on a test set for which we ensured that neither compounds, nor heads[11] were seen in the training data. To this end, we randomly selected two mid-frequency heads for each suffix and removed these from the training data to be put in the test data. We selected these for this initial experiment, because we expect mid-frequency heads to lead to most reliable results. High-frequency heads may show higher levels of idiosyncrasy and low-frequency heads may suffer from data sparseness. Since our goal is not to determine the realistic performance of our predictor, but to measure the contribution of features, this bias is acceptable. In future experiments, we plan to investigate the impact of frequency, which is not in the scope of the present study. This resulted in a division of roughly 90% training and 10% testing data.[12] Because the data set resulting from the annotation effort is skewed, and our selection of test instances introduces a different proportion of OBJ and NOBJ in the test and training sets, we balanced both sets by randomly removing instances with the OBJ relation from the training and test sets until both classes had equal numbers. The balanced training set consisted of 1614 examples,

[9] http://www.cs.cmu.edu/ ark/LexSem/

[10] We used version 3.8 for Linux of the Weka toolkit (Hall et al., 2009) and experimented with several other classifiers, focusing on those that have interpretable models (Decision Tree classifier, but also SVMs and Naive Bayes). All underperformed on our test set. However, the Decision Tree classifier also selects *percentage_head_in_DC* and *percentage_of_sg_outside* as the strongest predictors, just like the Logistic Regression classifier we are reporting on in Table 4.

[11] As can be seen in Table 3 the features are all head-specific.

[12] Multiple divisions of training and test data would lead to more reliable results, but we have to leave this for future work.

Features	Accuracy
All features	66.7%
All features, except *sg_percentage_outside*	66.7%
All features, except *sum_adjectives*	66.7%
All features, except *sg_percentage_inside*	66.7%
All features, except *percentage_head_in_DC*	46.7%†
All features, except *percentage_of_sg_outside*	56.1%†
All features, except *suffix*	61.7%†
All features, except *percentage_by_sg_inside*	71.1%†
Percentage_head_in_DC, *percentage_of_sg_outside*, and *suffix* combined	**76.1%**†

Table 4: Percent accuracy in ablation experiments. † indicates a statistically significant difference from the performance when including all features

and the test set of 180 examples. We ran ablation experiments to determine the individual contribution of each feature and combined the top-n features to see the predictive potential of the model.

4 Discussion of results

Our main concern was to determine whether our morphosyntactic head-related features have any predictive power. For this we compared the results using these features with a random baseline that lacked any information.[13] When using all features from Table 3, the classifier significantly outperforms[14] the random baseline (50%) with a reasonable margin (66.7%), showing that our features driven by linguistic theory have predictive power.

The ablation experiments in Table 4 shed light on the contribution of each feature.[15] The experiments show that *percentage_head_in_DC* has a high contribution to the overall model when it comes to predicting the relation between the deverbal head and the non-head, as its removal leads to a large drop in performance. The second strongest feature is *percentage_of_sg_outside*, and third comes *suffix*. One feature is actually harmful to the model: *percentage_by_sg_inside*, as its removal improves the accuracy of the classifier. The remaining features seem unimportant as their individual removal does not lead to performance differences. The best result we get on this task is 76,1%, when combining just the top-3 features (*percentage_head_in_DC*, *percentage_of_sg_outside*, and *suffix*). Although our test set is small, the performances indicated with the dagger symbol (†) lead to a statistically significant difference from the performance when including all features.

After inspecting the coefficients of the model, we are able to determine whether higher values of a given feature are indicating higher chances of an OBJ or NOBJ relation. Higher values of both *percentage_head_in_DC* and *percentage_of_sg_outside* lead to higher chances of predicting the OBJ class. For the categorical feature in our model, *suffix*, some point in the direction of a NOBJ interpretation (*-ance* and, less strongly, *-ment*), while others point in the direction of OBJ (*-ion* and, less strongly, *-al*), or do not have much predicting power as in the case of *-ing*.

From a theoretical point of view, these results have several implications. First, a high percentage of occurrences of the head inside compounds (i.e., *percentage_head_in_DC*) predicts an OBJ reading, which means that OBJ is the default interpretation of non-heads in DCs. Although this feature is not related to ASN-hood and previous linguistic literature does not mention it, we find it highly relevant as it defines the profile of DCs, in that deverbal heads that typically occur in compounds show a tendency to trigger an

[13] While stronger baselines would be needed to test the practical use of these features as compared to features used in the NLP literature, the random baseline is perfectly suitable to determine the predictive power of head features in theory.

[14] Significance numbers for these experiments in which training and test data are fixed were computed using a McNemar test with $p < .05$, because it makes relatively few type I errors (Dietterich, 1998).

[15] We realize that ablation does not lead to a complete picture of the strength of the features and their interplay. In addition, we tested several feature selection procedures by running the AttributeSelectedClassifier in Weka, because this allowed us to provide a separate test set. The best of these procedures (CfsSubsetEval) prefers subsets of features that are highly correlated with the class while having low inter-correlation and resulted in a non-optimal score of 70.3%. In future work, we would like to experiment with regularization to see which of the features' weights will be set to 0.

Head noun	Percentage_head_in_DC	OBJ
laundering	94.80%	95.45%
mongering	91.77%	100%
growing	68.68%	95.23%
trafficking	61.99%	100%
enforcement	53.68%	66.66%

Table 5: Head nouns with high compoundhood

Head noun	*Of*-phrases	OBJ
creation	80.51%	72.72%
avoidance	70.40%	100%
obstruction	65.25%	90.47%
removal	63.53%	92%
abandonment	55.90%	90%

Table 6: Head nouns with frequent *of*-phrases

OBJ interpretation of the non-head. This supports Grimshaw's claim that DCs structurally embed event structures with internal arguments.

Second, the next most predictive feature we found is the presence of an *of*-phrase realizing the internal argument of the head/verb (i.e., *percentage_of_sg_outside*), which again indicates an OBJ reading. In Grimshaw's approach, the realization of the internal argument is most indicative of the ASN status of a deverbal noun. This finding provides the strongest support for Grimshaw's claims and proves our hypothesis that high ASN-hood of the head triggers an OBJ interpretation of the non-head in DCs.

Tables 5 and 6 illustrate some examples related to these two most predictive features. Table 4 shows the five noun heads that present the highest percentage of appearance in a compound context and the percentage of OBJ readings among its compounds. Table 5 illustrates the five nouns heads that present highest percentage of *of*-phrases outside compounds and the corresponding percentage of OBJ readings.[16]

Third, the *suffix* feature is also predictive, with *-ion* and *-ance* indicating strong affinity for OBJ, respectively, NOBJ, and *-ing* being undetermined. These findings need further investigation, since the theoretical literature usually takes the suffixes *-ion*, *-ment*, *-ance*, and *-al* to be very similar. Should one wonder whether our overall results were not foreseeable given the selection of the suffixes in our dataset, we can name at least two reasons why this cannot be the case. First, the suffixes *-ing*, *-ion*, *-ment*, *-ance*, and *-al* all allow both OBJ and NOBJ readings and in terms of ASN-properties they all exhibit the ambiguity that Grimshaw describes. If any, then *-ing* should theoretically show a preference for OBJ, according to both Grimshaw's (1990) and Borer's (2013) approaches. Grimshaw takes *-ing* to mostly introduce ASNs, while Borer extensively argues that *-ing* lexically contributes an external argument, leaving only the internal argument available to be filled by the non-head in a compound. Thus, both approaches predict a preference for OBJ readings in DCs with *-ing*, while this suffix came out as undetermined between the two readings in our experiment. This means that the suffix that theoretically should have triggered the most foreseeable results, did not do so. Second, if the selection of the suffixes had had a decisive influence on the results, we would have expected the *suffix* feature to have more predictive power than it does and to trigger more unitary readings. But, as shown above, our results are mixed: while *-ion* prefers OBJ, *-ance* favors NOBJ. Within this background, more careful inspection of the data and further study on the individual suffixes would be necessary before we can conclude anything on the influence of each suffix.

Finally, we would like to comment on the noise that the feature *percentage_by_sg_inside* introduces into our model. Remember that the theoretical debate is unsettled as to whether *by*-phrases are possible in DCs. With (6c), Grimshaw indirectly states that they are possible, while Borer explicitly argues the opposite (see (7c-7d)). While theoretical analyses except for Borer (2013) show no clear stand, the fact that our model found this feature to be noisiest might be an indicator that *by*-phrases do not play a crucial role in the structure and interpretation of DCs.

Comparisons to performances in the NLP literature for the task of relation prediction make little sense at this time. The evaluation data in previous work are less carefully selected and cover a wider range of DC types (including zero-derived nominals and nouns ending in *-er* or *-ee*, among others). They used different statistical methods for prediction and different features. Moreover, it was not our aim to

[16]Note that despite having balanced the dataset for suffixes and number of compounds at the beginning, after the manual annotation some noun heads presented more errors than others as well as more disagreement among the annotators. This means that our final dataset is not so balanced as we initially designed it, but we will seek to do this in future work. In these tables we only included the nouns heads that were represented by at least 20 compounds in our final dataset.

reach state-of-the-art performance for a particular application. Our aim has been theoretical in nature and focused on understanding whether features of the head reported in the linguistic literature have predictive power, as well as determining which of these features are strongest.

For the sake of completeness, we would like to give the performance numbers reached by previous work. In the two-class prediction task, Lapata (2002) reaches an accuracy of 86.1% compared to a baseline of 61.5%, i.e, 24,6% above the baseline. The accuracy we achieve is 26,1% above the lower baseline of 50%.[17] Moreover, given the approach we used to discover useful indicators – that is, by means of a prediction task, in line with previous NLP work – it should be relatively easy to compare these results with previous studies in future work. We could test how our features based on the head compare with their encyclopaedic features for the prediction task, by evaluating our methods on their test sets and/or by incorporating their encyclopaedic features into our models for a direct comparison.

5 Conclusions

In this study, we identified two properties of deverbal noun heads with high predictive power in the interpretation of DCs: a head's predilection for DC-contexts and its frequent realization of internal arguments outside DCs. The latter is the most distinctive ASN-property that Grimshaw (1990) uses in her disambiguation of deverbal nouns, confirming her claim and our assumption that DCs have some event structure hosting internal arguments.

For the theoretical debate between the presence or absence of grammar in DCs, this means that we have two categories of DCs: some are part of the grammar and some should be part of the lexicon. On the one hand, we have DCs whose heads have ASN-properties and realize OBJ non-heads as predicted by the grammar-analysis (e.g., *drug trafficking*, *money laundering*, *law enforcement*). These DCs are part of the grammar and their OBJ interpretation can be compositionally derived from their event structure, so they do not need to be listed in the lexicon. On the other hand, DCs whose heads behave like RNs and realize NOBJ non-heads do not involve any event structure, so they cannot be compositionally interpreted and should be listed in the lexicon. Without previous (lexical) knowledge of the respective compound, one would interpret *adult supervision* or *government announcement* as involving an object by default, which would be infelicitous, since these compounds have a lexicalized NOBJ (i.e., subject-like) interpretation.

From a computational perspective, these experiments are a first attempt at trying to discover the complex set of dependencies that underlie the interpretation of deverbal compounds. Further work is necessary to determine the interdependence between the individual features, as well as to find out why adjectives and suffixes do not yield better results. Subsequently, taking into account the picture that arises from these additional experiments, we would like to compare our model based on head-dependent features with models that stem from NLP research and focus on encyclopaedic knowledge gathered from large corpora.

Acknowledgements

We thank our three anonymous reviewers for insightful and stimulating comments, which resulted in an improved and hopefully clearer version of our paper. Our work has been supported by the German Research Foundation (DFG) via a research grant to the Projects B1 *The Form and Interpretation of Derived Nominals* and D11 *A Crosslingual Approach to the Analysis of Compound Nouns*, within the Collaborative Research Center (SFB) 732, at the University of Stuttgart. We are also grateful to Katherine Fraser and Whitney Peterson for annotating our database and to Kerstin Eckart for technical support with respect to the Gigaword Corpus.

[17] Remember that we eventually balanced our dataset for 50% OBJ and 50% NOBJ compounds (cf. Section 3.5). Note that, given the biased interpretation of suffixes such as *-er* and *-ee*, including them into our dataset would have resulted in better accuracy and a higher predictive power for the *suffix* feature. But unlike, for instance, in Lapata (2002), we excluded this potential aid from our study, since we aimed to determine the morphosyntactic features of the head that are most relevant for the prediction task.

References

Peter Ackema and Ad Neeleman. 2004. *Beyond Morphology*. Oxford University Press, Oxford.

Artemis Alexiadou and Jane Grimshaw. 2008. Verbs, nouns, and affixation. In Florian Schäfer, editor, *Working Papers of the SFB 732 Incremental Specification in Context*, volume 1, pages 1–16. Universität Stuttgart.

Hagit Borer. 2013. *Taking Form*. Oxford University Press, Oxford.

Noam Chomsky. 1970. Remarks on nominalization. In Roderick A. Jacobs and Peter S. Rosenbaum, editors, *Readings in English transformational grammar*, pages 184–221. Waltham, MA: Ginn.

Thomas G Dietterich. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923.

Antske Sibelle Fokkens. 2007. A hybrid approach to compound noun disambiguation. MA-thesis. Universität des Saarladens, Saarbrücken.

Dan Gillick. 2009. Sentence boundary detection and the problem with the u.s. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, NAACL-Short '09, pages 241–244, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jane Grimshaw. 1990. *Argument Structure*. MIT Press, Cambridge, MA.

Claire Grover, Mirella Lapata, and Alex Lascarides. 2005. A comparison of parsing technologies for the biomedical domain. *Journal of Natural Language Engineering*, 11:01:27–65.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explorations*, 11(1):10–18.

Zhongqiang Huang, Mary Harper, and Slav Petrov. 2010. Self-training with products of latent variable grammars. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 12–22, Stroudsburg, PA, USA. Association for Computational Linguistics.

Angelika Kratzer. 1996. Severing the external argument from its verb. In Johan Rooryck and Laurie Zaring, editors, *Phrase Structure and the Lexicon*, pages 109–137. Kluwer Academic Publishers.

Mirella Lapata. 2002. The disambiguation of nominalizations. *Journal of Computational Linguistics*, 28:3:357–388.

Judith N. Levi. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.

Rochelle Lieber. 2004. *Morphology and Lexical Semantics*. Cambridge University Press, Cambridge.

Courtney Napoles, Matthew Gormley, and Benjamin Van Durme. 2012. Annotated gigaword. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, AKBC-WEKEX '12, pages 95–100, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jeremy Nicholson and Timothy Baldwin. 2006. Interpretation of compound nominalisations using corpus and web statistics. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 54–61, Sydney, Australia.

Susan Olsen. 2015. Composition. In Peter O. Müller et al., editors, *Word-Formation: An International Handbook of the Languages of Europe*, volume I, pages 364–386. De Gruyter.

Thomas Roeper and Muffy Siegel. 1978. A lexical transformation for verbal compounds. *Linguistic Inquiry*, 9:199–260.

Ina Rösiger, Johannes Schäfer, Tanja George, Simon Tannert, Ulrich Heid, and Michael Dorna. 2015. Extracting terms and their relations from german texts: Nlp tools for the preparation of raw material for e-dictionaries. In *Proceedings of eLex 2015*, pages 486–503. Herstmonceux Castle, UK.

Elisabeth O. Selkirk. 1982. *The Syntax of Words*. MIT Press, Cambridge, MA.