Verbs Taking Clausal and Non-Finite Arguments as Signals of Modality –
Revisiting the Issue of Meaning Grounded in Syntax

Judith Eckle-Kohler
Research Training Group AIPHES and UKP Lab
Computer Science Department, Technische Universität Darmstadt
www.aiphes.tu-darmstadt.de, www.ukp.tu-darmstadt.de

Abstract

We revisit Levin’s theory about the correspondence of verb meaning and syntax and infer semantic classes from a large syntactic classification of more than 600 German verbs taking clausal and non-finite arguments. Grasping the meaning components of Levin-classes is known to be hard. We address this challenge by setting up a multi-perspective semantic characterization of the inferred classes. To this end, we link the inferred classes and their English translation to independently constructed semantic classes in three different lexicons – the German wordnet GermaNet, VerbNet and FrameNet – and perform a detailed analysis and evaluation of the resulting German–English classification (available at www.ukp.tu-darmstadt.de/modality-verbclasses/).

1 Introduction

Verbs taking clausal and non-finite arguments add a further meaning component to their embedded argument. For example, the embedded argument is realized as that-clause in (1) and (2), but understand in (1) marks it as factual and hope in (2) as uncertain. The verb pretend in (3) realizes its embedded argument as non-finite construction and marks it as non-factual.

(1) He understands that his computer has a hardware problem.

(2) She hopes that her experience will help others.

(3) He pretends to take notes on his laptop, but really is updating his Facebook profile.

The entities expressed by embedded clausal and non-finite arguments are also called “abstract object” (AO) in the rest of this paper (following Asher (1993)); we will use the linguistic term “modality” (Hacquard, 2011) to subsume the meanings (such as factuality, non-factuality and uncertainty) denoted by AO-selecting verbs.

As AO-selecting verbs can change the meaning of a text in important ways, text understanding systems should be sensitive to them. In particular, classifications of AO-selecting verbs according to semantic criteria are important knowledge sources for a wide range of NLP applications, such as event tagging (Saurí et al., 2005), committed belief tagging (Prabhakaran et al., 2010), reported speech tagging (Krestel et al., 2008), the detection of uncertainty (Szarvas et al., 2012) and future-oriented content (Eckle-Kohler et al., 2008), textual entailment (Saurí and Pustejovsky, 2007; Lotan et al., 2013), or determining the degree of factuality of a given text (Saurí and Pustejovsky, 2012; de Marneffe et al., 2012).

Accordingly, various semantic classifications of AO-selecting verbs have been developed, e.g., (Kiparsky and Kiparsky, 1970; Karttunen, 1971; Karttunen, 2012), some of them explicitly in the context of NLP (Nairn et al., 2006; Saurí, 2008).

However, these classifications are constructed manually and often quite limited in coverage. Consequently, extending or adapting them to specific domains or other languages is a major issue.

We propose to address this issue by exploiting the relationship between the syntactic behavior of verbs and their meaning following Levin’s theory (Levin, 1993). This has not been done yet for verbs signaling modality, as far as we are aware. For the particular category of AO-selecting verbs, Levin’s theory allows constructing verb classifications in a purely syntax-driven way, i.e. inducing semantic classes from syntactically defined
classes, and thus possibly also extending given classes using large corpora.\textsuperscript{1}

While the appeal of Levin’s hypotheses is clear, we are aware of a major difficulty, making our approach a challenging research problem: it is very hard to grasp the precise meaning components which are to be associated with a syntactic “Levin” class. At the same time, it is vital to have a good semantic characterization of the meaning components in order to apply such classes to NLP tasks in an informed way.

We address these issues and make the following contributions: (i) We consider a purely syntactic classification of more than 600 German AO-selecting verbs and induce semantic classes based on findings from formal semantics about correspondences between verb syntax and meaning. This yields an initial description of the meaning components associated with the classes, along with a tentative class name. (ii) In a second step, we refine and extend the semantic characterization of the verb classes by translating it to English and linking it to existing semantic classes in lexical resources at the word sense level: we consider the coarse semantic fields in the German wordnet GermaNet (Kunze and Lemnitzer, 2002), the verb classes in the English lexicon VerbNet (Kipper et al., 2006), and the semantic frames in the English lexicon FrameNet (Baker et al., 1998). As a result, we obtain a detailed semantic characterization of the verb classes, as well as insights into the validity of Levin’s theory across the related languages German and English. (iii) We also perform a task-oriented evaluation of the verb classes in textual entailment recognition, making use of insights from the previous two steps. The results suggest that the verb classes might be a promising resource for this task, for German and for English.

\section{Related Work}

This section summarizes related work about the correspondence between verb meaning and syntax and discusses related work on modality in NLP.

\textbf{Syntactic Reflections of Verb Meaning} Semantic verb classifications that are grounded in lexical-syntactic properties of verbs are particularly appealing, because they can automatically be recovered in corpora based on syntactic features. The most well known verb classification based on correspondences between verb syntax and verb meaning is Levin’s classification (Levin, 1993). According to Levin (2015a), verbs that share common syntactic argument alternation patterns also have particular meaning components in common, thus they can be grouped into a semantic verb class. For example, verbs participating in the dative alternation\textsuperscript{2} can be grouped into a semantic class of verbs sharing the particular meaning component “change of possession”, thus this shared meaning component characterizes the semantic class. Recent work on verb semantics provides additional evidence for this correspondence of verb syntax and meaning: Hartshorne et al. (2014) report that the syntactic behavior of some verbs can be predicted based on their meaning.

VerbNet is a broad-coverage verb lexicon organized in verb classes based on Levin-style syntactic alternations: verbs with common subcategorization frames and syntactic alternation behavior that also share common semantic roles are grouped into VerbNet classes. VerbNet not only includes the verbs from the original verb classification by Levin, but also more than 50 additional verb classes (Kipper et al., 2006) automatically acquired from corpora (Korhonen and Briscoe, 2004). These classes contain many AO-selecting verbs that were not covered by Levin’s classification. However, VerbNet does not provide information about the modal meaning of AO-selecting verbs and does not reflect fine-grained distinctions between various kinds of modality.

There is also some criticism in previous work regarding the validity of Levin’s approach. Baker and Ruppenhofer (2002) and Schnorbusch (2004) both discuss various issues with Levin’s original classification, in particular the difficulty to grasp the meaning components, which are to be associated with a Levin class.

While approaches to exploit the syntactic behavior of verbs for the automatic acquisition of semantic verb classes from corpora have been developed in the past, they were used to recover only small verb classifications: Schulte im Walde (2006)’s work considered a semantically balanced set of 168 German verbs, Merlo and Stevenson (2001) used 60 English verbs from three particular semantic classes.

In contrast to previous work, we consider a large

\textsuperscript{1}Abstract objects already characterize the possible semantic roles to a certain extent.

\textsuperscript{2}These verbs can realize an argument syntactically either as noun phrase or as prepositional phrase with \textit{to}.
set of more than 600 German AO-selecting verbs and focus on their modal meaning (i.e., expressing factuality or uncertainty).

Related Work on Modality in NLP Previous work in NLP on the automatic (and manual) annotation of modality has often tailored the concept of modality to particular applications. Szarvas et al. (2012) introduce a taxonomy of different kinds of modality expressing uncertainty, such as deontic, bouletic, abilitative modality, and use it for detecting uncertainty in an Information Extraction setting. Their uncertainty cues also include verbs.

Saurí and Pustejovsky (2012) use discrete values in a modality continuum ranging from uncertain to absolutely certain in order to automatically determine the factuality of events mentioned in text. Their automatic approach is based on the FactBank corpus (Saurí and Pustejovsky, 2009), a corpus of newswire data with manually annotated event mentions. For the factuality annotation of the event mentions, the human annotators were instructed to primarily base their decision on lexical cues. For example, they used verbs of belief and opinion, perception verbs, or verbs expressing proof.

Nissim et al. (2013) introduce an annotation scheme for the cross-linguistic annotation of modality in corpora. Their annotation scheme defines two dimensions which are to be annotated (called layers): factuality (characterizing the embedded proposition or concept) and speaker’s attitude (characterizing the embedding predicate). Their annotation scheme starts from a fixed set of modal meanings and aims at finding previously unknown triggers of modality. However, some modal meanings are not distinguished, in particular those involving future-orientation. A classification approach grounded in syntax – as in our work – can be considered as complementary: it starts from the syntactic analysis of a large set of trigger words, and induces a broad range of modal meanings based on correspondences between verbs syntax and meaning.

Our semantic classification for AO-selecting verbs covers a wide range of different kinds of modality in text, thus considerably extending previous work.

3 Inferring Semantic Verb Classes

In this section, we infer semantic verb classes from the syntactic alteration behavior of a large dataset of German AO-selecting verbs. The research hypotheses underlying our method can be summarized as follows: There are correspondences between verb syntax and meaning: certain syntactic alternations correspond to particular meaning components (Levin, 2015a).

3.1 German Subcategorization Lexicon

We consider a set of 637 AO-selecting verbs given in (Eckle-Kohler, 1999). These verbs are a subset of a subcategorization lexicon (i.e., pairs of lemma and subcategorization frame) that has automatically been extracted from large newspaper corpora using a shallow regular expression grammar covering more than 240 subcategorization frames (short: subcat frames). All the subcat frames extracted for a given verb were manually checked and only the correct ones were included in the final lexicon, because high quality lexical information was crucial in the target application Lexical Functional Grammar parsing.\(^3\)

Eckle-Kohler (1999) specified the alteration behavior of each AO-selecting verb regarding different types of clausal and non-finite arguments, yielding a syntactic signature for each verb (e.g., 111101 for the verb *einsehen* (realize) using the encoding in Table 1, top to bottom corresponding to left to right).\(^4\) For this, each verb was inspected regarding its ability to take any of the considered clausal and non-finite constructions as argument – either on the basis of the automatically acquired subcat frames or by making use of linguistic introspection. Linguistic introspection is necessary to reliably identify nonpossible argument types, since missing subcat frames that were not extracted automatically are not sufficient as evidence.

Although there are 64 possible syntactic signatures according to basic combinatorics, in the data only 46 signatures were found, which group the verbs into 46 classes. While Eckle-Kohler (1999) points out a few semantic characteristics of these classes, most of them lack a semantic characterization. Our goal is to address this gap and to infer shared meaning components for all the classes.

\(^3\)Today, this lexicon is part of the larger resource “IMSLex German Lexicon” (Fitschen, 2004).

\(^4\)The automatically extracted subcategorization lexicon also contains adjectives and nouns taking clausal or infinitival arguments. However, many of the 1191 nouns and 666 adjectives are derived from verbs, which makes them the central word class.
occurs with non-objective verbs (e.g.,
whom
answer.

infinitive in the past tense (e.g.,
responding to a particular fact (Ginzburg, 1996).

ative and declarative clauses into factual AOs, cor-
that
dependent witness, whereas the
to objective verbs entailing the existence of an in-

Recently, linguistic research has increasingly
English. In English, the ing-form in the past would be more

Factuals: the that-wh and the that-wh/if alternation. Verbs that are able to alternatively take
that and wh-clauses coerce the embedded interrog-
ative and declarative clauses into factual AOs, cor-
responding to a particular fact (Ginzburg, 1996).

Among the verbs showing the that-wh alternation
are the well-known factive verbs (Kiparsky and
Kiparsky, 1970) (e.g., She proves that she exists.
vs. She proves who she is. vs. He proves whether
he can mine gold.).

There is a further distinction among these verbs regarding the ability to take an embedded
if/whether-question: Schwabe and Fittler (2009)
show that the that-wh/if alternation is connected
to objective verbs entailing the existence of an
independent witness, whereas the that-wh alternation (i.e., an if/whether-question is not possible)
occurs with non-objective verbs (e.g., He regrets
whom he ended up with. vs. *He regrets whether
he ended up playing this game.).

“Aspectuals”: the inability to take that-
clauses and to-infinitives in the past tense. Recently,
linguistic research has increasingly addressed particular semantic aspects of to-
infinitives. Kush (2011) has investigated AOs that
can neither be realized as that-clause nor as to-
infinite in the past tense (e.g., She hesitates to
answer. vs. *She hesitates to have answered.)

Table 1: Clausal and infinitival arguments distin-
guished in the syntactic classification; possibility
of each type is encoded as 1 (possible) or 0 (not possible).

For this, we use linguistic research findings as de-
scribed in the next section.

3.2 Findings from Formal Semantics

We employ the following findings on correpon-
dences between verb meaning and syntax in order
to infer semantic classes from the syntactic sig-
natures. This gives also rise to tentative names (la-

dels) for the corresponding meaning components.

Factuals: the that-wh and the that-wh/if alternation. Verbs that are able to alternatively take
that and wh-clauses coerce the embedded interrog-
ative and declarative clauses into factual AOs, cor-
responding to a particular fact (Ginzburg, 1996).

Among the verbs showing the that-wh alternation
are the well-known factive verbs (Kiparsky and
Kiparsky, 1970) (e.g., She proves that she exists.
vs. She proves who she is. vs. He proves whether
he can mine gold.).

There is a further distinction among these verbs regarding the ability to take an embedded
if/whether-question: Schwabe and Fittler (2009)
show that the that-wh/if alternation is connected
to objective verbs entailing the existence of an
independent witness, whereas the that-wh alternation (i.e., an if/whether-question is not possible)
occurs with non-objective verbs (e.g., He regrets
whom he ended up with. vs. *He regrets whether
he ended up playing this game.).

“Aspectuals”: the inability to take that-
clauses and to-infinitives in the past tense. Recently,
linguistic research has increasingly addressed particular semantic aspects of to-
infinitives. Kush (2011) has investigated AOs that
can neither be realized as that-clause nor as to-
infinite in the past tense (e.g., She hesitates to
answer. vs. *She hesitates to have answered.)

※She hesitates that ...）These AOs are selected by
control verbs⑧ and can be characterized as men-
tal actions. Kush (2011) points out that the verbs
selecting those AOs have a aspectual meaning in
common.

Future orientation: to-infinitives in the
present tense and the inability to take to-
infinitives in the past tense. Luca (2013) has in-
vestedigted verbs across English and Spanish that
embed future-oriented AOs. Only future-oriented
AOS can be used with future-oriented adverbials,
such as tomorrow, and these AOs are often real-
ized as non-finite constructions, e.g., to-infinitives.
She points out that not only control verbs take
future-oriented AOs, but also verbs expressing at-
itudes of preference. This finding implies that
such future-oriented AOS are typically incompati-
ble with past-oriented adverbials (e.g., yesterday)
and verb forms in the past tense (e.g., *She plans
having finished the assignment yesterday.).

3.3 Mapping to Meaning Components

We automatically infer semantic classes based on
a manually constructed mapping between the syn-
tactic signatures from Eckle-Kohler (1999) and the
meaning components grounded in syntax summa-
rized in Section 3.2.⑨

We constructed this mapping in two steps: In a
first step, the signatures are aligned to the meaning
components from Section 3.2 based on substrings
of the signatures: future-orientation matches the
110 prefix, aspectual the 010 prefix, and factuality
matches 1’s in fourth or fifth position. It is
important to point out that future-orientation can
be combined with factuality: this corresponds to
an independent matching of the 110 prefix and the
factuality substring. While this combination may
seem contradictory, it reflects the lexical data and
shows that also weak forms of factuality (“it will
most likely be factual at some point in the future”) are
expressed in language.

In a second step, the pre-aligned signatures are
merged, if the remaining slots of the signature are
either 1 or 0 (i.e. the respective argument types
can or can not occur); in the resulting merged sig-

Table 1: Clausal and infinitival arguments distin-
guished in the syntactic classification; possibility
of each type is encoded as 1 (possible) or 0 (not possible).

| Argument Type                  | Y/N | Example |
|--------------------------------|-----|---------|
| daft(that)-clause             | 1/0 | sehen (see) |
| zu(to)-infinitive, present    | 1/0 | versuchen (try) |
| zu(to)-infinitive, past       | 1/0 | bereuen (regret) |
| wh-clause                     | 1/0 | einsehen (realize) |
| ob(whether/if)-clause         | 1/0 | fragen (ask) |
| declarative clause           | 1/0 | schreien (shout) |

⑧This is the literal translation of the German equivalent to
English. In English, the ing-form in the past would be more

⑨"Control" refers to the co-reference between the implicit
subject of the infinitival argument and syntactic arguments
in the main clause, either the subject (subject control) or direct
object (object control).

⑩We did not consider verbs can be used with all kinds of
clausal and infinitival arguments, such as the majority of
communication verbs (e.g., comment, whisper).
Table 2: The 632 verbs in 8 semantic classes (5 verbs show idiosyncratic behavior). Signature substrings in bold correspond to meaning components, which (along with tentative class names) are based on Sec. 3.2. The cross-lingual semantic characterization shows aligned VerbNet (VN) classes covering 265 (42%) verbs and aligned FrameNet (FN) frames covering 126 (20%) verbs, see Sec. 4.1.6

| signature | #verbs – examples | meaning components | semantic characterization (#linked verbs) |
|-----------|-------------------|--------------------|-----------------------------------------|
| 010 ---   | 36 (6%) – wagen (dare), zögern (hesitate), weigern (refuse) | aspectual: verbs expressing the ability of doing an action | VN (2): consider-29.9, wish-62; FN (2): purpose, cogitation |
| 110 0--   | 195 (31%) – anbieten (offer), empfehlen (recommend), fordern (demand) | future-oriented: verbs marking AOs as anticipated, planned | VN (89): force-59, forbid-67, wish-62, promote-102, urge-58.1, order-60, admire-31.2, order-60, promise-37.13; FN (43): request, preventing |
| 000 11-   | 15 (2%) – nachfragen (inquire), anfragen (ask) | interrogative: verbs marking AOs as under investigation | VN (3): estimate-34.2, inquire-37.1.2, order-60; FN (1): questioning, request |
| 111 1--   | 122 (19%) – bedauern (regret), überwinden (overcome), danken (thank) | wh-factual: opinion verbs marking AOs as factual | VN (45): transfer-mesg-37.1.1, wish-62, admire-31.2, complain-37.8, conjecture-29.5, say-37.7; FN (18): statement, reveal-secret |
| 110 10-   | 30 (5%) – befürworten (approve), verteidigen (defend), loben (praise) | future-oriented wh-factual: opinion verbs marking AOs as future-oriented and factual | VN (15): admire-31.2, allow-64, transfer-mesg-37.1.1, suspect-81, characterize-29.2, neglect-75, want-32.1, defend-85, comprehend-87.2; FN (10): judgment, grant-permission, defend, experiencer-focus, judgment-communication, justifying, hit-or-miss, statement, reasoning, tolerating, grasp |
| 1-- 11-   | 120 (19%) – beschreiben (describe), hören (hear), erinnern (remember) | wh/if -factual: objective verbs marking AOs as factual | VN (55): discover-84, say-37.7, see-30.1, comprehend-87.2, rely-70, seem-109, consider-29.9, transfer-mesg-37.1.1, estimate-34.2, inquire-37.1.2; FN (23): perception-experience, statement, cogitation, grasp |
| 110 11-   | 48 (8%) – festlegen (determine), abschätzen (assess), lehren (teach) | future-oriented wh/if-factual: objective verbs marking AOs as future-oriented and factual | VN (28): estimate-34.2, rely-70, indicate-78, transfer-mesg-37.1.1, correspond-36.1, conjecture-29.5, discover-84, say-37.7; FN (16): predicting, education-teaching, assessing, relying, reasoning |
| 111 0--   | 66 (10%) – vorwerfen (accuse), bestreiten (deny), färben (fear) | non-factual: verbs marking AOs as not resolvable re. their factuality | VN (28): conjecture-29.5, wish-62, complain-37.8, admire-31.2; FN (13): statement, reveal-secret, experiencer-focus, certainty |

Table 2: The 632 verbs in 8 semantic classes (5 verbs show idiosyncratic behavior). Signature substrings in bold correspond to meaning components, which (along with tentative class names) are based on Sec. 3.2. The cross-lingual semantic characterization shows aligned VerbNet (VN) classes covering 265 (42%) verbs and aligned FrameNet (FN) frames covering 126 (20%) verbs, see Sec. 4.1.6

While the descriptions of the meaning components and the class names are inspired from research in linguistics (typically a very deep analysis of only few verbs), transferring them to our verb resource – which is of much larger scale – inevitably leads to outlier verbs in the classes, e.g., verbs that do not strictly match the class label. Examples include verbs such as überlegen (consider) in the wh/if-factual class (not covering the future-oriented meaning component) or schaden (harm) as non-factual rather than as wh-factual. For this reason, and also because of the assignment of highly polysemous verbs to only one class, the definitions of meaning components and the class names should rather be considered as loose, providing a first tentative semantic characterization of the modality classes.

In sum, this section presented an inventory...
of modal meaning components that we primarily synthesized from research in linguistics. The classification work is strictly grounded in syntactic properties of the verbs and was not targeted a priori at modal meanings.

4 Evaluation

4.1 Linking to Semantic Classes

Our first set of experiments aims at refining the initial semantic characterization of the classes by linking them to independently constructed semantic classifications at the word sense level. Specifically, we consider three different semantic classifications from computational lexicons, which have been created by linguistic experts: (i) the so-called semantic fields in GermaNet, grouping verb senses into 15 coarse classes, such as perception, emotion, (ii) the verb classes given in VerbNet, and (iii) the Frame-semantic frames in FrameNet. As the GermaNet and FrameNet classes are based on different lexicographic and linguistic theories, we expect an additional semantic characterization from the linking. The VerbNet classes, which also follow Levin’s hypotheses, however, are used to investigate if the syntax-semantics correspondence is maintained across the related languages German and English.

For this linking experiment, we used the UBY framework (Gurevych et al., 2012)\(^{10}\), containing standardized versions of the above lexicons, as well as a linking between VerbNet and FrameNet on the word sense level.

**Approach** In order to link our classes to verb senses in GermaNet and VerbNet, we developed an automatic linking method based on subcat frame similarity. Recognizing subcat frame similarity requires a common standardized format for the otherwise incomparable frames. UBY provides such a standardized format which has been presented in detail by Eckle-Kohler and Gurevych (2012). It represents subcat frames uniformly across German and English, and at a fine-grained level of individual syntactic arguments. Our linking approach is based on the following hypothesis: Two verb senses with equivalent lemmas are equivalent, if they have similar subcat frames.\(^{11}\) Our method interprets the pairs of verb and subcat frame listed in our classification\(^{12}\) as senses. While we do not claim that this hypothesis is sufficient in general, i.e., for all verb senses, we found that it is valid for the subset of senses belonging to the class of AO-selecting verbs.

The cross-lingual linking of our classes to VerbNet senses requires an additional translation step, which we describe first.

**Manual Translation** While UBY also provides translations between German and English verb senses, e.g., as part of the Interlingual Index from EuroWordnet (ILI), we found that many of the translations were not present in our target lexicon VerbNet. Therefore, the main author of this paper, a native speaker of German with a good proficiency in English, translated the AO-compatible verbs (i.e., word senses) manually using Linguee\(^{13}\) and dictionaries. This took about 7 hours.

For 23 German verbs, we could not find any equivalent lexicalized translation, because these verbs express very fine-grained semantic nuances. For example, we did not find an equivalent English verb for a few verbs in the aspectual class but only a translation consisting of an adjective in combination with to be. Examples include be easy (leichtfallen), be willing (sich bereitfinden), be capable (vermögen), which have German equivalents that are lexicalized as verbs. As a result, we arrived at translations for 614 out of 637 German verbs. These 614 German verbs are translated to 413 English verbs, indicating that the English translation has a more general meaning in many cases.

**Automatic Verb Sense Linking** Our algorithm links a German verb sense (or its English translation) with a GermaNet (or VerbNet) sense, if the subcat frames of both verb senses have the same number of arguments and if the arguments have certain features in common.\(^{14}\) For example, to create a link to GermaNet, features such as the complementizer of clausal arguments and the case of noun phrase arguments have to agree. In a similar way, the linking to VerbNet is based on a comparison of German subcat frames and English subcat

\(^{10}\)http://www.ukp.tu-darmstadt.de/uby/

\(^{11}\)This approach is applicable for GermaNet, because GermaNet contains fine-grained syntactic subcat frames.

\(^{12}\)We consider only verb senses that are compatible with AOs, as indicated by subcat frames with clausal or non-finite arguments.

\(^{13}\)Linguee (http://www.linguee.de/) is a translation tool combining an editorial dictionary and a search engine processing bilingual texts. In particular, it provides a large variety of contextual translation examples.

\(^{14}\)We do not link the subcat frames, but we do compare them across the related languages German and English to determine their similarity in the context of linking.
frames – which are represented uniformly across German and English. In Section A.2, we provide more details about the algorithm.

**Results** According to a manual evaluation of a random sample of 200 sense pairs, the automatic verb sense linking yielded an accuracy of 89.95% for the linking to GermaNet, and 87.54% for the linking to VerbNet (κ agreement on the sample annotated by two annotators was 0.7 and 0.8, respectively). The main types of errors in the linking to GermaNet and VerbNet are due to specific syntactic features of the subcat frames which diverge and are not considered in the automatic linking. The differences regarding these specific features are due to cross-lingual differences (VerbNet, e.g., verb phrase arguments with ing-form) and diverging linguistic analyses of particular constructions (GermaNet, e.g., constructions with es (it)), see also Eckle-Kohler and Gurevych (2012).

By linking the verbs in our classification to semantic classes in GermaNet, VerbNet and FrameNet, we obtain a three-way semantic characterization of our classes. The linking to the GermaNet semantic fields covers 270 (43%) of the source verbs. Of these, 219 (81%) are linked to the three semantic fields cognition, communication and social. Fewer verbs (32 (12%)) are linked to the semantic fields emotion, perception, change. Semantic fields not among the target classes are consumption, competition, contact, body and weather.

Table 2 summarizes the linking to VerbNet and FrameNet and shows how many verbs from each source class could be linked to any of the classes in VerbNet or FrameNet. As the class distribution of the verb subsets covered by our linking-based evaluation is similar as for the original classes, we consider our evaluation as valid, although less than 50% of all verbs could be evaluated this way.

The target classes in VerbNet and FrameNet reveal meaning components that are on the one hand unique for individual classes, and on the other hand shared across several German classes.

The **future-oriented** class contains object control verbs (e.g., force-59, forbid-67 in VerbNet, and request, preventing in FrameNet). The wh/if-factual class is unique regarding the cognition and perception verbs (e.g., discover-84, see-30.1-1, and perception-experience). The **future-directed**

| Verb class          | Wiki | Web | News | News Eng. |
|---------------------|------|-----|------|-----------|
| all                 | 25.85| 50.58| 33.91| 25.31     |
| aspectual           | 0.90 | 0.80| 1.44 | 1.96      |
| future-oriented     | 9.45 | 23.04| 13.65| 12.58     |
| interrogative       | 0.01 | 0.05| 0.05 | 0.65      |
| wh-factual          | 4.26 | 17.89| 4.99 | 3.48      |
| fo. wh-factual      | 0.29 | 0.28| 0.85 | 1.14      |
| wh/if-factual       | 3.02 | 2.54| 3.53 | 5.20      |
| fo. wh/if-factual   | 2.36 | 1.77| 3.14 | 5.75      |
| non-factual         | 4.29 | 3.36| 4.84 | 3.57      |

Table 3: Percentage of classes in corpora: German Wikipedia (Wiki), SDeWaC (Web), Tiger (News); English Reuters corpus (News Eng.).

The **wh/if-factual** class also contains **objective assessment** verbs, as shown by the estimate-34.2 class. The verbs in the two **wh-factual** classes share meaning components as well, as shown by the **opinion verb** classes admire-31.2 and defend-85 in VerbNet or judgment, tolerating in FrameNet.

While there are also other VerbNet and FrameNet classes shared across several classes, they turned out to be very general and underspecified regarding their meaning, thus not contributing to a more fine-grained semantic characterization. For example, the conjecture-29.5 class assembles quite diverse conjecture verbs, e.g. verbs expressing opinion (feel, trust) and factuality (observe, discover). A similar observation holds for the statement frame in FrameNet.

### 4.2 Analysis of Frequency and Polysemy

In order to assess the usefulness of the verb resource for NLP tasks, we determined the lemma frequency of all verbs in the 8 classes in SDeWaC (Faiß and Eckart, 2013), a cleaned version of the German DeWaC corpus (Baroni and Kilgarriff, 2006). A ranking of the verbs according to their lemma frequency showed that 89% of the verbs occur more than 50 times in SDeWaC.16

We also analyzed the frequency distribution of the 8 verb classes in two other German corpora belonging to different genres, and also for English, see Table 3:17 encyclopedic text (the German Wikipedia18), German newspaper text (the Tiger corpus (Brants et al., 2004)), and the English

---

15Based on the percentage of source class members linked to any of the target classes, we only display target classes with an overlap of at least 1.8% due to space constraints.

16In the verb resource we provide for download, we included this frequency information in order to enable frequency-based filtering.

17Details of the computation of the verb lemma frequency lists are given in the appendix A.1.

18www.wikipedia.de, dump of 2009-06-18
usefulness of the German
For an extrinsic evaluation, we investigated the
4.3 Textual Entailment Experiment
whether the polysemy is an issue for any applica-
the category of AO-selecting verbs we consider.
Whether the polysemy is an issue for any applica-
where the verb classes are used as features is
not a priori clear and depends on the task at hand.

4.3 Textual Entailment Experiment
For an extrinsic evaluation, we investigated the usefulness of the German and the English verb classes as features in recognizing textual entailment (RTE). In RTE, the task is to determine whether for a pair of text fragments – the text T and the hypothesis H – the meaning of H is entailed by T (Dagan et al., 2006); for non-entailing pairs, sometimes a further category “unknown” is used as a label.

We employed a simple classification-based approach to RTE and trained and evaluated a Naive Bayes classifier on the test sets of three RTE benchmarks, using 10-fold cross validation: the English RTE-3 data (Giampiccolo et al., 2009) and their German translation (the development sets and the test sets each consist of 800 pairs), and an expanded version of the English RTE-3 data from the Sagan Textual Entailment Test Suite (Castillo, 2010) consisting of 2974 pairs. While the German dataset provides a two-way classification of the T-H pairs, the two English datasets provide a three-way classification, also using the “unknown” label. We used the DKPro TC framework (Daxenberger et al., 2014) for classification and applied POS tagging and lemmatization as preprocessing.

As a baseline feature, we use the word overlap measure between two T-H pairs (no stopword filtering, no lemmatization, no normalization of overlap score), which is quite competitive on the RTE-3 data, because this dataset shows a high difference in word overlap between positive (entailment) and negative (no entailment) pairs (Bentivogli et al., 2009).

An analysis of the development set of the German RTE-3 data showed that 62% of the pairs contain at least one occurrence of any of the verbs from the classification in either T or H. However, T and H fragments display no statistically significant differences regarding the occurrences of any of the verb classes.

A detailed analysis revealed that pairs without entailment are often characterized by a mismatch between T and H regarding the presence of factuality markers. For example, the presence of verbs indicating uncertainty (all classes apart from wh-factual and wh/if-factual) in T and an absence of such verbs in H might indicate non-entailment as in the following not entailing pair from the English RTE3 development set where “long” signals non-factuality, but “researching” signals factuality:

T: The BBC’s Americas editor Will Grant says many Mexicans are tired of conflict and long for a return to normality.
H: Will Grant is researching a conflict with Mexicans.

Thus, an insufficient overlap of modality markers in T and H might actually indicate non-entailment, but lead to an incorrect classification as entailment when considering only word overlap.

Accordingly, we implemented a factuality-mismatch feature both for German and for English, based on our new German and English classes. This feature is similar to the word overlap feature but with lemmatization and normalization of overlap score. Verb class counts are based on verb lemma counts of the member verbs; for English verbs that are members of more than one class, we included all verb classes in our factuality-mismatch feature.

Table 4 shows the results. While the differences

---

19 Reuters-21578, Distribution 1.0, see http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html.
20 http://www.dfki.de/~neumann/resources/RTE3_DE_V1.2_2013-12-02.zip
21 All significance scores in this paper are based on Fisher’s exact test at significance level $p<0.05$.
22 In the German part, every verb is assigned to one class, while the translation to English resulted in 22% of the English verbs being members in more than one class. However, only 11% of the multiple class assignments involve a combination of factual and uncertainty classes.
for RTE-3 DE and RTE-3 EN are not statistically significant, the factuality-mismatch feature yielded a small but significant improvement on the expanded RTE-3 EN dataset. This is due to the different nature of the expanded RTE dataset, which was created using a paraphrasing technique. As a result, the number of occurrences of verbs from our classes increased, and the factuality-mismatch became a discriminative feature for distinguishing between CONTRADICTION and UNKNOWN/ENTAILMENT.

Considering the fact that we employed only simple overlap features that do not rely on dependency parsing and did not perform any word sense disambiguation, these results suggest that the verb classes might be promising features for RTE, both for German and English. As factuality can be expressed by a variety of further linguistic means, including modal verbs, negation, tense and certain adverbs, investigating the combination of our verb classes with other modality signals might be especially promising as part of future work.

| RTE-3 DE | RTE-3 EN | RTE-3 EN exp. |
|----------|----------|--------------|
| WO       | 59.87    | 54.75        |
| WO+FM    | 59.25    | 54.62        |
|          |          | **58.81**    |

Table 4: Accuracy of a Naive Bayes classifier (10-fold cross validation on the test sets) with word overlap (WO) and additional factuality-mismatch (WO+FM) features.

5 Results and Discussion

Our construction of semantic classes from the syntactic behavior of AO-selecting verbs results in an inventory of modal meanings that emerged from a large lexical resource. The main result of the linking based evaluation is a detailed semantic characterization of the inferred classes – a prerequisite for using them in NLP tasks in an informed way. The semantic classes seem to be particular suited for tasks related to opinion analysis, textual inference, or argumentation mining. In this context, the relationship between our large resource of lexical verbs and the closed class of modal verbs might be an interesting question for future research.

Most of all, the linking to GermaNet and FrameNet shows that it is indeed possible to narrow down meaning components for Levin classes. Moreover, the results of the linking to VerbNet also provide support for Levin’s hypothesis that the correspondences between verb syntax and meaning described for English largely apply to the related language German as well (Levin, 2015b).

The English version of the semantic classes which we created by means of translation has the same semantic properties as the German classes. However, the syntactic properties of the English classes are not fully specified, because English has additional kinds of non-finite arguments, such as ing-forms or bare infinitives. Therefore, it might be interesting to address this question in the future and to build a similar semantic classification for English from scratch, in particular in the context of extracting modality classes from corpora. This would require an adaptation of the syntactic signatures, considering the various kinds of non-finite arguments particular to English. Based on large subcategorization lexicons available for English (e.g. COMLEX (Grishman et al., 1994) or VerbNet), it should be feasible to derive such signatures and to construct a mapping of signatures to modality aspects in a similar way as for German.

The question whether the syntactic signatures can be recovered in large corpora is particularly interesting, because this would allow extending the existing classes and to also acquire AO-selecting adjectives and nouns. We plan to investigate this question as part of future work.

6 Conclusion

We inferred semantic classes from a large syntactic classification of German AO-selecting verbs based on findings from formal semantics about correspondences between verb syntax and meaning. Our thorough evaluation and analysis yields detailed insights into the semantic characteristics of the inferred classes, and we hope that this allows an informed use of the resulting resource in various semantic NLP tasks.

Acknowledgments

This work has been supported by the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant No. I/82806 and by the German Research Foundation under grant No. GU 798/17-1 and No. GRK 1994/1. We thank the anonymous reviewers for their valuable comments. Additional thanks go to Anette Frank, Iryna Gurevych and Ani Nenkova for their helpful feedback on earlier versions of this work.
References

Nicholas Asher. 1993. Reference to Abstract Objects in Discourse. Studies in Linguistics and Philosophy (Book 50). Springer.

Collin F. Baker and Josef Ruppenhofer. 2002. FrameNet’s Frames vs. Levin’s Verb Classes. In Proceedings of 28th Annual Meeting of the Berkeley Linguistics Society, pages 27–38, Berkeley, CA, USA.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL), pages 86–90, Montreal, Canada.

Marco Baroni and Adam Kilgarriff. 2006. Large Linguistically-Processed Web Corpora for Multiple Languages. In Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics (EACL), pages 87–90, Trento, Italy.

Luisa Bentivogli, Bernardo Magnini, Ido Dagan, Hoa Trang Dang, and Danilo Giampiccolo. 2009. The Fifth Pascal Recognizing Textual Entailment Challenge. In Proceedings of the Text Analysis Conference (TAC), pages 14–24, Gaithersburg, Maryland, USA.

Sabine Brants, Stefanie Dipper, Peter Eisenberg, Silvia Hansen, Esther König, Wolfgang Lezius, Christian Rohrer, George Smith, and Hans Uszkoreit. 2004. TIGER: linguistic interpretation of a German corpus. Research on Language and Computation, 2(4):597–620.

Julio J. Castillo. 2010. Using Machine Translation Systems to Expand a Corpus in Textual Entailment. In Hrafn Loftsson, Eirkur Rgnvaldsdsson, and Sigur Helgadþtir, editors, Advances in Natural Language Processing, volume 6233 of Lecture Notes in Computer Science, pages 97–102. Springer, Berlin, Heidelberg.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The PASCAL Recognising Textual Entailment Challenge. In Joaquin Quiñonero-Candela, Ido Dagan, Bernardo Magnini, and Florence dAlch Buc, editors, Machine Learning Challenges, Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment, volume 3944 of Lecture Notes in Computer Science, pages 177–190. Springer Berlin Heidelberg.

Johannes Daxenberger, Oliver Ferschke, Iryna Gurevych, and Torsten Zesch. 2014. Dkpro tc: A java-based framework for supervised learning experiments on textual data. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 61–66, Baltimore, MD, USA.

Marie-Catherine de Marneffe, Christopher D. Manning, and Christopher Potts. 2012. Did It Happen? The Pragmatic Complexity of Veridicality Assessment. Computational Linguistics, 38(2):301–333, June.

Richard Eckart de Castilho and Iryna Gurevych. 2014. A Broad-Coverage Collection of Portable NLP Components for Building Shareable Analysis Pipelines. In Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (OIAF4HLT) at COLING 2014, pages 1–11, Dublin, Ireland.

Judith Eckle-Kohler and Iryna Gurevych. 2012. Subcat-LMF: Fleshing Out a Standardized Format for Subcategorization Frame Interoperability. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pages 550–560, Avignon, France.

Judith Eckle-Kohler, Michael Kohler, and Jens Meinert. 2008. Automatic recognition of german news focusing on future-directed beliefs and intentions. Computer Speech and Language, 22(4):394–414, October.

Judith Eckle-Kohler. 1999. Linguistisches Wissen zur automatischen Lexikon-Akquisition aus deutschen Textcorpora. Logos-Verlag, Berlin, Germany. PhD Thesis, Universität Stuttgart, Germany.

Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC – A Corpus of Parsable Sentences from the Web. In Iryna Gurevych, Chris Biemann, and Torsten Zesch, editors, Language Processing and Knowledge in the Web: Proceedings of the 25th Conference of the German Society for Computational Linguistics (GSCL 2013), Darmstadt, Germany, September 25–27, 2013., pages 61–68. Springer, Berlin, Heidelberg.

Arne Fitschen. 2004. Ein Computerlinguistisches Lexikon als komplexes System. PhD Thesis, Universität Stuttgart, Germany.

Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. 2009. The Third PASCAL Recognizing Textual Entailment Challenge. In Proceedings of the Workshop on Textual Entailment and Paraphrasing at ACL 2009, pages 1–9, Prague, Czech Republic.

Jonathan Ginzburg. 1996. Interrogatives: Questions, Facts, and Dialogue. In Shalom Lappin, editor, The Handbook of Contemporary Semantic Theory, pages 385–422. Blackwell, Oxford, UK.

Ralph Grishman, Catherine Macleod, and Adam Meyers. 1994. Comlex Syntax: Building a Computational Lexicon. In Proceedings of the 15th International Conference on Computational Linguistics (COLING), pages 268–272, Kyoto, Japan.
Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pages 580–590, Avignon, France.

Valentine Hacquard. 2011. Modality. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, Semantics: An International Handbook of Natural Language Meaning. HSK 33.2, pages 1484–1515. Berlin: Mouton de Gruyter.

Joshua K. Hartshorne, Claire Bonial, and Martha Palmer. 2014. The VerbCorner Project: Findings from Phase 1 of Crowd-Sourcing a Semantic Decomposition of Verbs. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), pages 397–402, Baltimore, MD, USA.

Lauri Karttunen. 1971. Implicative Verbs. Language, pages 340–358.

Lauri Karttunen. 2012. Simple and Phrasal Implicatives. In *SEM 2012: The First Joint Conference on Lexical and Computational Semantics, pages 124–131, Montréal, Canada.

Paul Kiparsky and Carol Kiparsky, 1970. Fact. Mouton, The Hague.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with Novel Verb Classes. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), pages 1027–1032, Genoa, Italy.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A Large-scale Classification of English Verbs. Language Resources and Evaluation, 42:21–40.

Anna Korhonen and Ted Briscoe. 2004. Extended Lexical-Semantic Classification of English Verbs. In Proceedings of the Workshop on Computational Lexical Semantics at HLT-NAACL 2004, pages 38–45, Boston, Massachusetts, USA.

Ralf Krestel, Sabine Bergler, and Ren Witte. 2008. Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles. In Nicolletta Calzolari et al., editor, Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC), pages 2823–2828, Marrakech, Morocco.

Claudia Kunze and Lothar Lemnitzer. 2002. GermaNet – Representation, Visualization, Application. In Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC), pages 1485–1491, Las Palmas, Canary Islands, Spain.

Dave Kush. 2011. Mental Action and Event Structure in the Semantics of ‘try’. In Proceedings of the 21st Semantics and Linguistic Theory Conference, pages 413–425, New Brunswick, New Jersey, USA.

Brenda Laca. 2013. Temporal Orientation and the Semantics of Attitude Verbs. In Karina Veronica Molsing and Ana Maria Tramunt Ibáñez, editors, Time and TAME in Language, pages 158–180. Cambridge Scholars Publishing, Newcastle upon Tyne, UK.

Beth Levin. 1993. English Verb Classes and Alternations. The University of Chicago Press, Chicago, USA.

Beth Levin. 2015a. Semantics and Pragmatics of Argument Alternations. Annual Review of Linguistics, 1(1):63–83.

Beth Levin. 2015b. Verb Classes Within and Across Languages. In Andrej Malchukov and Bernard Comrie, editors, Valency Classes in the Worlds Languages (Volume 2): Case Studies from Austronesia, the Pacific, the Americas, and Theoretical Outlook, pages 1627–1670. Berlin, Boston: De Gruyter Mouton.

Amnon Lotan, Asher Stern, and Ido Dagan. 2013. TruthTeller: Annotating Predicate Truth. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 752–757, Atlanta, Georgia.

Paola Merlo and Suzanne Stevenson. 2001. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. Computational Linguistics, 27(3):373–408, September.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing Relative Polarity for Textual Inference. Inference in Computational Semantics (ICoS-5), pages 20–21.

Malvina Nissim, Paola Pietrandrea, Andrea Sanso, and Caterina Mauri. 2013. Cross-Linguistic Annotation of Modality: a Data-Driven Hierarchical Model. In Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation, pages 7–14, Potsdam, Germany.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. 2010. Automatic Committed Belief Tagging. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING), pages 1014–1022, Beijing, China.

Rosar Saurí and James Pustejovsky. 2007. Determining Modality and Factuality for Text Entailment. In Proceedings of the International Conference on Semantic Computing, ICSC ’07, pages 509–516, Washington, DC, USA. IEEE Computer Society.

Rosar Saurí and James Pustejovsky. 2009. FactBank: a Corpus Annotated with Event Factuality. Language Resources and Evaluation, 43(3):227–268.
A.1 Verb Lemma Frequency List

In order to count the occurrences of verb lemmas in the German corpus SDeWaC, we used a reader and pre-processing components (i.e., the LanguageTool segmenter and the TreeTagger for POS tagging and lemmatization) from the DKPro Core collection (Eckart de Castilho and Gurevych, 2014). From DKPro Core, we also used a component that detects separated particles of German particle verbs and replaces the lemma of the verb base form annotated by the TreeTagger by the true lemma of the particle verb. Our verb lemma counting pipeline is available at github.com/UKPLab/acl2016-modality-verbclasses.