Situation entity annotation

Annemarie Friedrich **Alexis Palmer**
Department of Computational Linguistics
Saarland University, Saarbrücken, Germany
{afried,apalmer}@coli.uni-saarland.de

Abstract

This paper presents an annotation scheme for a new semantic annotation task with relevance for analysis and computation at both the clause level and the discourse level. More specifically, we label the finite clauses of texts with the type of *situation entity* (e.g., eventualities, statements about kinds, or statements of belief) they introduce to the discourse, following and extending work by Smith (2003). We take a feature-driven approach to annotation, with the result that each clause is also annotated with fundamental aspectual class, whether the main NP referent is specific or generic, and whether the situation evoked is episodic or habitual. This annotation is performed (so far) on three sections of the MASC corpus, with each clause labeled by at least two annotators. In this paper we present the annotation scheme, statistics of the corpus in its current version, and analyses of both inter-annotator agreement and intra-annotator consistency.

1 Introduction

Linguistic expressions form patterns in discourse. Passages of text can be analyzed in terms of the individuals, concepts, times and *situations* that they introduce to the discourse. In this paper we introduce a new semantic annotation task which focuses on the latter and in particular their aspectual nature. Situations are expressed at the clause level; **situation entity (SE)** annotation is the task of associating individual clauses of text with the type of SE introduced to the discourse by the clause. Following Smith (2003), we distinguish the following *SE types* (see Sec. 3.1): EVENTS, STATES, GENERALIZING SENTENCES, GENERIC SENTENCES, FACTS, PROPOSITIONS, QUESTIONS and IMPERATIVES. Although these categories are clearly distinct from one another on theoretical grounds, in practice it can be difficult to cleanly draw boundaries between them. We improve annotation consistency by defining the SE types in terms of features whose values are easier for annotators to identify, and which provide guidance for distinguishing the more complex SE types.

As with most complex annotation tasks, multiple interpretations are often possible, and we cannot expect agreement on all instances. The feature-driven approach (see Sec. 3.2) is a valuable source of information for investigating annotator disagreements, as the features indicate precisely how annotators differ in their interpretation of the situation. Analysis of intra-annotator consistency shows that personal preferences of annotators play a role, and we conclude that disagreements often highlight cases where multiple interpretations are possible. We further argue that such cases should be handled carefully in supervised learning approaches targeting methods to automatically classify situation entity types.

As the first phase of the SE annotation project, we are in the process of annotating the written portion of MASC (Ide et al., 2010), the manually-annotated subcorpus of the Open American National Corpus. MASC provides texts from 20 different genres and has already been annotated with various linguistic and semantic phenomena.[1] MASC offers several benefits: it includes text from a wide variety of genres, it facilitates study of interactions between various levels of analysis, and the data is freely available with straightforward mechanisms for distribution. In this paper we report results for three of the MASC

This work is licenced under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: http://creativecommons.org/licenses/by/4.0/

[1] http://www.americannationalcorpus.org/MASC/Full_MASC.html

LAW VIII - The 8th Linguistic Annotation Workshop, pages 149–158,
Dublin, Ireland, August 23-24 2014.

genres: news, letters, and jokes. Once a larger portion of MASC has been labeled with SEs and their associated features, we will add our annotations to those currently available for MASC. We mark the SE types of clauses with the aim of providing a large corpus of annotated text for the following purposes:

(1) To assess the applicability of SE type classification as described by Smith (2003): to what extent can situations be classified easily, which borderline cases occur, and how do humans perform on this task? (see Sec. 4)

(2) Training, development and evaluation of automatic systems classifying situation entities, as well as sub-tasks which have (partially) been studied by the NLP community, but for which no large annotated corpora are available (for example, automatically predicting the fundamental aspectual class of verbs in context (Friedrich and Palmer, 2014) or the genericity of clauses and noun phrases).

(3) To provide a foundation for analysis of the theory of Discourse Modes (Smith, 2003), which we explain next (Sec. 2).

2 Background and related work

Within a text, one recognizes stretches that are intuitively of different types and can be clustered by their characteristic linguistic features and interpretations. Smith (2003) posits five *discourse modes*: Narrative, Report, Description, Informative and Argument/Commentary. Texts of almost all genre categories have passages of different modes. The discourse modes are characterized by (a) the type of situations (also called *situation entities*) introduced in a text passage, and (b) the principle of text progression in the mode (temporal or atemporal, and different manners of both temporal and atemporal progression). This annotation project directly addresses the first of these characteristics, the *situation entity types (SE types)*.

Some previous work has addressed the task of classifying SE types at the clause level. Palmer et al. (2004) enrich LFG parses with lexical information from both a database of lexical conceptual structures (Dorr, 2001) and hand-collected groups of predicates associated with particular SE types. The enriched parses are then fed to an ordered set of transfer rules which encode linguistic features indicative of SE types. The system is evaluated on roughly 200 manually-labeled clauses. Palmer et al. (2007) investigate various types of linguistic features in a maximum entropy model for SE type classification. The best results are still below 50% accuracy (with a most-frequent-class baseline of 38%), and incorporating features from neighboring clauses is shown to increase performance. Palmer et al. (2007) annotate data from one section of the Brown corpus and a small amount of newswire text, with two annotators and no clear set annotation guidelines. In addition, work by Cocco (2012) classifies clauses of French text according to a six-way scheme that falls somewhere between the SE level and the level of discourse modes. The types are: narrative, argumentative, descriptive, explicative, dialogal, and injunctive.

Other related works address tasks related to the features we annotate. One strand of work is in automatic classification of aspectual class (Siegel and McKeown, 2000; Siegel, 1999; Siegel, 1998; Klavans and Chodorow, 1992; Friedrich and Palmer, 2014) and its determination as part of temporal classification (UzZaman et al., 2013; Bethard, 2013; Costa and Branco, 2012). A second aims to distinguish generic vs. specific clauses (Louis and Nenkova, 2011) or to identify generic noun phrases (Reiter and Frank, 2010). The latter work leverages data with noun phrases annotated as either generic and specific from the ACE-2 corpus (Mitchell et al., 2003); their definitions of these two types match ours (see Sec. 3.2.1).

3 Annotation Scheme and Process

In this section, we first present the inventory of SE types (Sec. 3.1). We then describe our feature-driven approach to annotation (Sec. 3.2) and define the SE types with respect to three situation-related features: main referent type, fundamental aspectual class, and habituality. Some situation entity types are easier to recognize than others. While some can be identified on the basis of surface structure and clear linguistic indicators, others depend on internal temporal (and other) properties of the verb and its arguments. Annotators take the following approach: first, easily-identifiable *SE types* (Speech Acts and Abstract Entities) are marked. If the clause's SE type is not one of these, values for the three features are determined, and the final determination of SE type is based on the features.

3.1 Situation entity types

Following Smith (2003), we distinguish the following *SE types*:

Eventualities. These types describe particular situations such as STATES (1a) or EVENTS (2). The type REPORT, a subtype of EVENT, is used for situations introduced by verbs of speech (1b).

(1) *(a) "Carl is a tenacious fellow", (*STATE*)*
*(b) said a source close to USAir. (*EVENT – REPORT*)*

(2) *The lobster won the quadrille. (*EVENT*)*

General Statives. This class includes GENERALIZING SENTENCES (3), which report regularities related to specific main referents, and GENERIC SENTENCES (4), which make statements about kinds.

(3) *Mary often feeds my cats. (*GENERALIZING*)*

(4) *The lion has a bushy tail. (*GENERIC*)*

Abstract Entities are the third class of SE types, and comprise FACTS (5) and PROPOSITIONS (6). These situations differ from the other types in how they relate to the world: Eventualities and General Statives are located spatially and temporally in the world, but Abstract Entities are not. FACTS are objects of knowledge and PROPOSITIONS are objects of belief from the respective speaker's point of view.

(5) *I know that Mary refused the offer. (*FACT*)*

(6) *I believe that Mary refused the offer . (*PROPOSITION*)*

We limit the annotation of Abstract Entities to the clausal complements of certain licensing predicates, as well as clauses modified by a certain class of adverbs, as it is not always possible to identify sentences directly expressing Facts or Propositions on linguistic grounds (Smith, 2003). In (6), *believe* is the licensing predicate, and *Mary refused the offer* is a situation that is introduced as not being *in* the world, but *about* the world (Smith, 2003). Annotators are asked to additionally label the embedded SE type when possible. For example, *that Mary refused the offer* in (5) and (6) would be labeled as EVENT.

Speech Acts. This class comprises QUESTIONS and IMPERATIVE clauses (Searle, 1969).

Derived SE types. In some cases, the SE type of a clause changes based on the addition of some linguistic indication of uncertainty about the status of the situation described. We refer to these as derived SE types. More specifically, clauses that would otherwise be marked as EVENT may be coerced to the type STATE due to negation, modality, future tense, conditionality, and sometimes subjectivity: e.g. *John did not win the lottery*, a negated event, introduces a STATE to the discourse.

3.2 Features for distinguishing situation entity types

In this section, we describe three features that allow for the clear expression of differences between SE types. Fleshing out the descriptions of SE types with these underlying features is useful to convey the annotation scheme to new annotators, to get partial information when an annotator has trouble making a decision on SE type, and to analyze disagreements between annotators.

3.2.1 Main referent type: specific or generic

This feature indicates the type of the most central entity mentioned in the clause as a noun phrase. We refer to this entity as the clause's *main referent*. This referent can be found by asking the question: *What is this clause about?* Usually, but not always, the main referent of a clause is realized as its grammatical subject. We appeal to the annotator's intuitions in order to determine the main referent of a clause. In case the main referent does not coincide with the grammatical subject as in example (7), this is to be indicated during annotation.

(7) *There are two books on the table. (**specific** main referent,* STATE*)*

Some SE types (STATES, GENERALIZING SENTENCES and GENERIC SENTENCES, for details see Table 1) are distinguished by whether they make a statement about some *specific* main referent or about a *generic* main referent. Specific main referents are particular entities (8), particular groups of entities (9), organizations (10), particular situations (11) or particular instantiations of a concept (12).

(8) *Mary likes popcorn. (particular entity → **specific**,* STATE*)*

(9) *The students met at the cafeteria. (a particular group → **specific**,* STATE*)*

(10) *IBM was a very popular company in the 80s. (organization → **specific**,* STATE*)*

(11) *That she didn't answer her phone really upset me. (particular situation → **specific**,* EVENT*)*

(12) *Today's weather was really nice. (particular instantiation of a concept → **specific**,* STATE*)*

The majority of generic main referents are noun phrases referring to a *kind* rather than to a particular entity, and generic mentions of concepts or notions (14). Definite NPs and bare plural NPs (13) are the main kind-referring NP types (Smith, 2003).

(13) *The lion has a bushy tail. / Dinosaurs are extinct. (**generic**,* GENERIC SENTENCE*)*

(14) *Security is an important issue in US electoral campaigns. (**generic**,* GENERIC SENTENCE*)*

While some NPs clearly make reference to a well-established kind, other cases are not so clear cut, as humans tend to make up a context in which an NP describes some kind (Krifka et al., 1995). Sentence (15) gives an example for such a case: while *lions in captivity* are not a generally well-established kind, this term describes a class of entities rather than a specific group of lions in this context.

(15) *Lions in captivity have trouble producing offspring. (**generic**,* GENERIC SENTENCE*)*

Gerunds may occur as the subject in English sentences. When they describe a *specific* process as in (16a), we mark them as specific. If they instead describe a *kind* of process as in (16b), we mark them as generic.

(16) *(a) Knitting this scarf took me 3 days. (**specific**,* EVENT*)*
*(b) Knitting a scarf is generally fun. (**generic**,* GENERIC SENTENCE*)*

We also give annotators the option to explicitly mark the main referent as *expletive*, as in (17).

(17) *It seemed like (**expletive** = no main referent,* STATE*)*
*he would win. (**specific**,* STATE*)*

3.2.2 Fundamental aspectual class: stative or dynamic

Following Siegel and McKeown (2000), we determine the *fundamental aspectual class* of a clause. This notion is the extension of *lexical aspect* or *aktionsart*, which describe the "real life shape" of situations denoted by verbs, to the level of clauses. More specifically, aspectual class is a feature of the main verb and a select group of modifiers, which may differ per verb. The stative/dynamic distinction is the most fundamental distinction in taxonomies of aspectual class (Vendler, 1967; Bach, 1986; Mourelatos, 1978).

We allow three labels for this feature: **dynamic** for cases where the verb and its arguments describe some event (something happens), **stative** for cases where they introduce some properties of the main referent to the discourse, or **both** for cases where annotators see both interpretations.

It is important to note that the fundamental aspectual class of a verb can be different from the type of situation entity introduced by the clause as a whole. The basic situation type of *building a house* is **dynamic**, and in the examples below we see this fundamental aspectual class appearing in clauses with different situation entity types. Example (18) describes an EVENT. Clause (19), on the other hand, is a GENERALIZING SENTENCE, as it describes a pattern of events; this is a situation with a *derived* type. The same is true for example (20), which is a STATE because of its future tense.

(18) *John built a house. (*EVENT*, **dynamic** fundamental aspectual class)*

(19) *John builds houses. (*GENERALIZING SENTENCE*, **dynamic** fundamental aspectual class)*

(20) *John is going to build a house. (*STATE*, **dynamic** fundamental aspectual class)*

3.2.3 Habituality

Another dimension along which situations can be distinguished is whether they describe a **static** state, a one-time (**episodic**) event (21) or some regularity of an event (22) or a state (23), which is labeled **habitual**. The term *habitual* as used in this annotation project covers more than what is usually considered a matter of habit, extending to any clauses describing regularities (24). The discussion related to this linguistic feature in this section follows Carlson (2005). If one can add a frequency adverbial such as *typically/usually* to the clause and the meaning of the resulting sentence differs at most slightly from the meaning of the original sentence, or the sentence contains a frequency adverbial such as *never*, the sentence expresses a regularity, i.e., is habitual. Another property of habituals is that they are generalizations and hence have the property of tolerating exceptions. If we learn that Mary eats oatmeal for breakfast, it does not necessarily need to be true that she eats oatmeal at every breakfast. It is important to note that unlike fundamental aspectual class, *habituality* is an attribute of the entire situation.

(21) *Mary ate oatmeal for breakfast this morning. (**episodic**,* EVENT*)*

(22) *Mary eats oatmeal for breakfast. (**habitual**,* GENERALIZING SENTENCE*)*

(23) *I often feel as if I only get half the story. (**habitual**, **stative** fundamental aspectual class,* GENERALIZING SENTENCE*)*

(24) *Glass breaks easily. (**habitual**,* GENERIC SENTENCE*)*

3.3 SE types and their features

The feature-driven approach to annotation taken here is defined such that, ideally, each unique combination of values for the three features leads to one SE type. Table 1 shows the assignment of SE types to various combinations of feature values. This table covers all SE types except ABSTRACT ENTITIES and SPEECH ACTS, which are more easily identifiable based on lexical and/or syntactic grounds. Annotators are also provided with information about linguistic tests for some SE types and feature values, both for making feature value determinations and to support selection of clause-level SE type labels.

SE type	main referent	aspectual class	habituality
EVENT	specific	eventive	episodic
	generic		
STATE	specific	stative	static
GENERIC SENTENCE	generic	eventive	habitual
		stative	static, habitual
GENERALIZING SENTENCE	specific	eventive	habitual
		stative	
General Stative	specific	eventive	habitual
	generic		

Table 1: Situation entity types and their features.

4 Annotator agreement and consistency

This section presents analyses of inter-annotator agreement and intra-annotator consistency, looking at agreement for individual feature values as well as clause-level SE type.

4.1 Data and annotators

The current version of our corpus consists of three sections (news, letters and jokes) of MASC corpus (Ide et al., 2010). We hired three annotators, all either native or highly-skilled speakers of English, and had a training phase of 3 weeks using several Wikipedia documents. Afterwards, annotation of the texts began and annotators had no further communication with each other. Two annotators (A and B) each marked the complete data set, and one additional annotator (C) marked the news section only.

ANNOTATORS	NUMBER OF SEGMENTS	MAIN REFERENT	ASPECTUAL CLASS	HABITUALITY	SE TYPE	SE TYPE (REP=EVT)
A:B	2563	0.35	0.81	0.77	0.56	0.66
A:C	2524	0.29	0.77	0.76	0.55	0.65
B:C	2556	0.45	0.73	0.76	0.76	0.74
average	2545	0.36	0.77	0.76	0.62	0.68

Table 2: **Cohen's** κ, for pairs of annotators on the MASC `news` section.

GENRE	NUMBER OF SEGMENTS	MAIN REFERENT	ASPECTUAL CLASS	HABITUALITY	SE TYPE	SE TYPE (REP=EVT)
jokes	3455	0.57	0.85	0.81	0.74	0.73
news	2563	0.35	0.81	0.77	0.56	0.66
letters	1851	0.41	0.71	0.65	0.56	0.56
all	7869	0.47	0.80	0.77	0.64	0.68

Table 3: **Cohen's** κ, for two annotators on three different sections of MASC.

4.2 Segmentation into clauses

We segment the texts into finite clauses using the SPADE discourse parser (Soricut and Marcu, 2003), applying some heuristic post-processing and allowing annotators to mark segments that do not contain a situation (for instance, headlines or by-lines) or that should be merged with another segment in order to describe a complete situation. We filter out all segments marked by any annotator as having a *segmentation problem*. Of the 2823 segments automatically created for the news section, 4% were marked as containing no situation by at least one of the three annotators, and 7% were merged to a different segment by at least one annotator. All three annotators agree on the remaining 2515 segments (89%). Of the 9428 automatically-created segments in the full data set, 11.5% were marked as no-situation by at least one of two annotators, and a further 5% were merged to other segments by at least one annotator. 7869 segments remain for studying agreement between two annotators on the full data set.

The three genres vary as to the average segment length. Segments in the letters texts have the longest average length (11.1 tokens), segments in jokes are the shortest (6.9 tokens on average), and segments in news fall in the middle with an average length of 9.9 tokens.

4.3 Inter-annotator agreement

As we allow annotators to mark a segment as Speech Acts or Abstract Entities and in addition mark the SE type of the embedded situation with a non-surface type, we compute agreement for Eventualities and General Statives in the following, and present the results for Speech Acts and Abstract Entities separately.

news section, 3 annotators. We compute Cohen's unweighted κ between all three pairs of annotators for the news section, as shown in Table 2. We compute agreement for the segments where both respective annotators agree on the segmention, i.e., that the segment describes a situation. For aspectual class, we compute agreement over the three labels *stative*, *dynamic* and *both*; for main referents, we compute agreement over the three labels *specific*, *dynamic* and *expletive*; for habituality, we compute agreement over the three labels *episodic*, *habitual* and *static*. In each case, we omit segments for which one of the annotators did not give a label, which in each case are fewer than 26 segments.

We observe good agreement for the features aspectual class and habituality, and for SE type between annotators B and C. Pairs involving annotator A reach lower agreement; we identify two causes. Annotator A marks many segments marked as REPORT by the others as the corresponding supertype EVENT. This shows up in Table 2 as higher values of κ when considering REPORT to match its supertype EVENT. The second cause is A's different preference for marking main referents, causing lower κ scores for agreement on the main referent type and also influencing agreement for situation entity types. In more than 92% of the 183 clauses on which annotators B and C agree with each other, but disagree with A, B and C assigned the value specific while A marked the main referent as generic. Early in the annotation project, a revision was made to the scheme for labeling main referents – one hypothesis is that A might not have updated her way of labeling these. We estimate that roughly 40% of these cases were due to

A's misunderstanding of feature value definitions, but around 30% of these cases do allow for both interpretations. In the following sentence, the main referent of the second segment could either refer to the specific set of all kids in New York, or to the class of children in New York: *As governor, I'll make sure // that every kid in New York has the same opportunity.* Another frequent case is the main referent *you*, which can be interpreted in a generic way or as specifically addressing the reader (e.g. of a letter). Such disagreements at the level of feature annotations allow us to detect cases where several interpretations are possible. Having annotators with different preferences on difficult cases can actually be a valuable source of information for identifying such cases.

The distribution of labels for main referents is highly skewed towards specific main referents for the `news` section; when comparing B and C, they agree on 2358 segments to have a specific main referent. However, only 122 segments are labeled as having a generic main referent by at least one annotator, and they agree only on 43 of them. A further 49 are labeled generic by B but specific by C and a further 30 vice versa. In order to collect more reliable data and agreement numbers for the task of labeling main referent types, we plan to conduct a focused study with a carefully-balanced data set.

news, jokes, letters: 2 annotators. We report agreement for three sections, corresponding to three genres, for two annotators (A and B) in Table 3. We observe higher agreement for jokes than for news, and higher agreement for news than for letters. Figure 1 shows the distribution of situation entity types per genre. The numbers express averages of percentages of label types assigned to the clauses of one genre by the two annotators. The letters genre is different in that it has more STATES, far fewer EVENTS, which are usually easy to detect, and more General Statives. Most cases of confusion between annotators occur between General Statives and STATES, so the more EVENTS texts have, the higher the agreement.

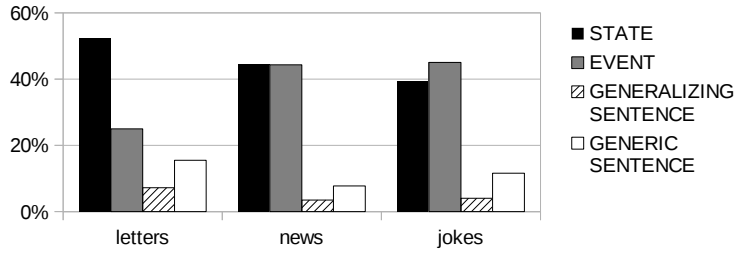

Figure 1: Distribution of situation entity types in three different genres.

Speech Acts and Abstract Entities. Figure 2 shows the percentage of segments of each genre that were marked as a Speech Act or an Abstract Entity by at least one annotator. QUESTIONS are most frequent in the jokes genre, but about half of them are just marked by one annotator, which has to do with how consistently indirect questions are marked. The two annotators agree on almost all segments labeled as imperatives; while there are only very few IMPERATIVES in the news section, there are more in the jokes and letters sections. The letters are mainly fund-raising letters, which explains the high percentage of IMPERATIVES (*Please help Goodwill. // Use the enclosed card // and give a generous gift today.*). FACTS and PROPOSITIONS, on the other hand, are rather infrequent in any genre, and annotators tend to mark them inconsistently. We take from this analysis that we need to offer some help to the annotators in detecting Abstract Entities. We plan to compile a list of verbs that may introduce Abstract Entities and specifically highlight potential licensing constructions in order to increase recall for these types.

4.4 Intra-annotator consistency

After the first round of annotation, we identified 11 documents with low inter-annotator agreement on SE type (5 news, 5 letters, 1 jokes) and presented them to two annotators for re-annotation. For each annotator, the elapsed time between the first and second rounds was at least 3 weeks. We observe that in general, the agreement of each annotator with herself is greater than agreement with the other annotator. This shows that the disagreements are not pure random noise, but that annotators have different preferences for certain difficult decisions. It is interesting to note that annotator B apparently changed how

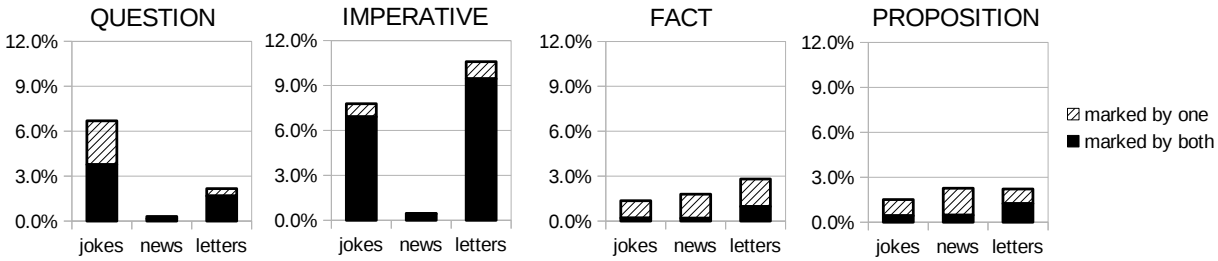

Figure 2: Percentage of segments marked as Speech Act or Abstract Entity by at least one annotator.

GENRE	NUMBER OF SEGMENTS	MAIN REFERENT	ASPECTUAL CLASS	HABITUALITY	SE TYPE	SE TYPE (REP=EVT)
A1:B1	636	0.15	0.79	0.64	0.40	0.45
A2:B2	599	0.12	0.78	0.70	0.42	0.48
A1:A2	596	0.79	0.88	0.78	0.75	0.75
B1:B2	620	0.55	0.84	0.78	0.75	0.75

Table 4: Consistency study: **Cohen's** κ, for two annotators, comparing against each other and against themselves (re-annotated data). A1 = annotator A in first pass, B2 = annotator B in second pass etc.

she annotates main referents; possibly this is also due to the above mentioned revision to the annotation scheme. On the other hand, B annotated very few segments as generic (only 61 segments were marked as having a generic main referent in either the first or second pass, 27 of them in both passes), which may also have led to the low κ value. The fact that annotators *do* disagree with themselves indicates that there are noisy cases in our data set, where multiple interpretations are possible. However, we want to point out that the level of noise estimated by this intra-annotator consistency study is an upper bound as we chose the most difficult documents for re-annotation; the overall level of noise in the data set can be assumed to be much lower.

5 Conclusion

We have presented an annotation scheme for labeling clauses with their situation entity type along with features indicating the type of main referent, fundamental aspectual class and habituality. The feature-driven approach allows for a detailed analysis of annotator disagreements, showing in which way the annotators' understandings of a clause differ. The analysis in the previous chapter showed that while good inter-annotator agreement can be reached for most decisions required by our annotation schema, there remain hard cases, on which annotators disagree with each other or with their own first round of annotations. We do not yet observe satisfying agreement for main referent types or for identifying abstract entities. In both cases, data sparseness is a problem; there are only very few generic main referents and abstract entities in our current corpus. We plan to conduct case studies on data that is specifically selected for these phenomena.

However, in many of the hard cases, several readings are possible. Rather than using an adjudicated data set for training and evaluation of supervised classifiers for labeling clauses with situation entities, we plan to leverage such disagreements for training, following proposals by Beigman Klebanov and Beigman (2009) and Plank et al. (2014).

The annotation reported here is ongoing; our next goal is to extend annotation to additional genres within MASC, starting with essays, journal, fiction, and travel guides. Following SE annotation, we will extend the project to annotation of discourse modes. Finally, we are very interested in exploring and annotating SEs in other languages, as we expect a similar inventory but different linguistic realizations.

Acknowledgments We thank the anonymous reviewers, Bonnie Webber and Andreas Peldszus for helpful comments, and our annotators Ambika Kirkland, Ruth Kühn and Fernando Ardente. This research was supported in part by the MMCI Cluster of Excellence, and the first author is supported by an IBM PhD Fellowship.

References

Emmon Bach. 1986. The algebra of events. *Linguistics and philosophy*, 9(1):5–16.

Beata Beigman Klebanov and Eyal Beigman. 2009. From annotator agreement to noise models. *Computational Linguistics*, 35(4):495–503.

Steven Bethard. 2013. ClearTK-TimeML: A minimalist approach to TempEval 2013. In *Second Joint Conference on Lexical and Computational Semantics (* SEM)*, volume 2, pages 10–14.

Greg Carlson. 2005. Generics, habituals and iteratives. *The Encyclopedia of Language and Linguistics*.

Christelle Cocco. 2012. Discourse type clustering using pos n-gram profiles and high-dimensional embeddings. In *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL 2012.

Francisco Costa and António Branco. 2012. Aspectual type and temporal relation classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 266–275.

Bonnie J. Dorr. 2001. LCS verb database. Online software database of Lexical Conceptual Structures, University of Maryland, College Park, MD.

Annemarie Friedrich and Alexis Palmer. 2014. Automatic prediction of aspectual class of verbs in context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). Baltimore, USA*.

Nancy Ide, Christiane Fellbaum, Collin Baker, and Rebecca Passonneau. 2010. The manually annotated sub-corpus: A community resource for and by the people. In *Proceedings of the ACL 2010 conference short papers*, pages 68–73.

Judith L. Klavans and Martin S. Chodorow. 1992. Degrees of stativity: The lexical representation of verb aspect. In *Proceedings of the 14th COLING*, Nantes, France.

Manfred Krifka, Francis Jeffry Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. 1995. Genericity: an introduction. *The Generic Book*, pages 1–124.

Annie Louis and Ani Nenkova. 2011. Automatic identification of general and specific sentences by leveraging discourse annotations. In *Proceedings of IJCNLP 2011*.

Alexis Mitchell, Stephanie Strassel, Mark Przybocki, JK Davis, George Doddington, Ralph Grishman, Adam Meyers, Ada Brunstein, Lisa Ferro, and Beth Sundheim. 2003. ACE-2 Version 1.0. *Linguistic Data Consortium, Philadelphia*.

Alexander PD Mourelatos. 1978. Events, processes, and states. *Linguistics and philosophy*, 2(3):415–434.

Alexis Palmer, Jonas Kuhn, and Carlota Smith. 2004. Utilization of multiple language resources for robust grammar-based tense and aspect classification. In *Proceedings of LREC 2004*.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. A sequencing model for situation entity classification. *Proceedings of ACL 2007*.

Barbara Plank, Dirk Hovy, and Anders Søgaard. 2014. Learning part-of-speech taggers with inter-annotator agreement loss. In *Proceedings of EACL 2014*.

Nils Reiter and Anette Frank. 2010. Identifying generic noun phrases. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*.

John Searle. 1969. *Speech Acts*. Cambridge University Press.

Eric V Siegel and Kathleen R McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.

Eric V. Siegel. 1998. Disambiguating verbs with the WordNet category of the direct object. In *Proceedings of Workshop on Usage of WordNet in Natural Language Processing Systems*, Universite de Montreal.

Eric V. Siegel. 1999. Corpus-based linguistic indicators for aspectual classification. In *Proceedings of ACL37*, University of Maryland, College Park.

Carlota S Smith. 2003. *Modes of discourse: The local structure of texts*. Cambridge University Press.

Radu Soricut and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics.

Naushad UzZaman, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second joint conference on lexical and computational semantics (* SEM)*, volume 2, pages 1–9.

Zeno Vendler, 1967. *Linguistics in Philosophy*, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, New York.