Methodological Aspects of Semantic Annotation

Harry Bunt and Amanda Schiffrin

Department of Language and Information, Faculty of Arts, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, Netherlands
harry.bunt@uvt.nl, mandy@uvt.nl

Abstract

This paper constitutes a preliminary report on the work carried out on semantic content annotation in the LIRICS project, in close collaboration with the activities of ISO TC 37/SC 4/TDG 3[1]. This consists primarily of: (1) identifying commonalities in alternative approaches to the annotation and representation of various types of semantic information; and (2) developing methodological principles and concepts for identifying and characterising representational concepts for semantic content. The LIRICS project does not aim to develop a standard *format* for the annotation and representation of semantic content, but at providing well-defined *descriptive concepts*. In particular, the aim is to build an on-line registry of definitions of such concepts, called '*data categories*', in accordance with ISO standard 12620. These semantic data categories are abstract concepts, whose use is not restricted to any particular format or representation language. We advocate the use of the metamodel as a tool to extract the most important of these abstract overarching concepts, with examples from dialogue act, temporal, reference and semantic role annotation.

1. Models and Metamodels

Alternative approaches to the marking up of linguistic resources differ most importantly in the categories of information that they aim to capture. The choices made in this respect can be represented by specifying the classes of objects and relations that are covered by their markup tags. Such a characterisation is called a *model*. Looking for commonalities in alternative approaches implies comparing their underlying models. This can be done by moving to a more abstract level than that of the models themselves, building a so-called *metamodel*. Metamodels are well known from software engineering, where they are loosely defined as a model that describes a set of models. Bunt and Romary (2004) have proposed a more formal interpretation of the term metamodel by relating it to the notion of model as used in model-theoretic semantics. This notion of metamodel can be used as a methodological tool for the definition of semantic concepts, and for the isolation of corresponding semantic data categories of importance.

We argue that by using metamodels we can find an overarching conceptualisation for diverging linguistic theories. A metamodel is constructed by identifying the data categories of differing models that represent identical, similar or related items conceptually, and then by introducing a broader concept that includes the variations. In this way, one can retain the individual distinctions of the specific theories, while at the same time capturing the generalities. This is not always an easy or straightforward process, but when a metamodel is abstracted from individual models within the same theoretical area, and can also be shown to 'fit' the phenomena and structure to be found in the varying component theories, then this provides a good basis for consensus within the research community, as well as the first step in the standardisation of core concepts within a field.

2. Types of Semantic Annotation

The LIRICS project will tackle at least the following specific areas of semantic interest: dialogue acts, temporal entities and relations, reference and semantic roles. There are very clear motivations for considering the areas discussed here in particular. Firstly, they largely coincide with similar areas of interest in ISO. Secondly, each of these areas has achieved a certain level of maturity in the semantics (and pragmatics) research communities.

2.1. Dialogue Acts

One area of language and speech technology that could benefit greatly from internationally agreed annotation standards is that of dialogue modelling. Annotated dialogue corpora are important sources of information for the design of spoken dialogue systems, for the development of interactive text-based multimodal human-computer interfaces and for the creation of systems that mediate in a useful way in human-human dialogue. Dialogue acts have become popular for annotating dialogues in order to indicate what the participants are doing (see e.g. Jurafsky & Martin 2000, Chapter 19).

The term 'dialogue act' is sometimes used rather informally, in the sense of 'speech act used in dialogue'. Accordingly, a dialogue act has a certain function or purpose, roughly corresponding to the 'illocutionary force' of speech act theory, and a propositional or referential content ('propositional content' in speech act theory). A dialogue act may also be said to have 'locutionary' and 'perlocutionary' aspects, although these are often not considered, because the locutionary act is constitutive of the dialogue act itself, while the perlocutionary act does not lend itself to systematic formalisation and identification. In order to be optimally useful for dialogue analysis purposes, however, a more precise and self-contained notion of dialogue act would be necessary, focusing on its role in assigning meanings to utterances in dialogue. Following the 'information-state update' or 'context-change' approach to dialogue, we may define a dialogue act as a semantic unit in the description of dialogue utterance meaning, having two main components: a *communicative function* and a *semantic content*. The semantic content is information that the source of the dialogue act is bringing to the addressee's attention, and the communicative function specifies what the addressee should do with the semantic content, i.e., in what way the addressee should use that information to update his information state (or context model), upon understanding the utterance.

Dialogue acts are thus tied to utterances, which are parts of turns, which are parts of a dialogue. Mostly, a

---

[1] See: `http://let.uvt.nl/research/ti/iso-tdg3`.

dialogue act is realized by a single utterance, but it may be that the dialogue act is expressed by two utterances together, for instance when a speaker is interrupted by some dialogue-external circumstance. It may also happen that a speaker is unable to complete an utterance that he intends to make in order to express a certain dialogue act, and that the addressee helps the speaker; in such a case, we believe it is still best to consider only the first participant as playing the sender role. Similarly if an addressee performs backchannel acts to give positive feedback. In sum, a dialogue act is related via an utterance to a turn and thus to a single participant in the sender role, and to another participant in the addressee role.

In the case of multiparty dialogue, there may be multiple addressees. Also, there may be additional participants who witness the dialogue without belonging to the intended audience (but both sender and addressee(s) may be aware of their presence, and take that into account); their role might be called that of 'overhearer'. The category of overhearer includes the ratified participants in a conversation, such as the side participant and the bystander, of which sender and addressee are aware, and the non-ratified participants, eavesdroppers, of which they are not.
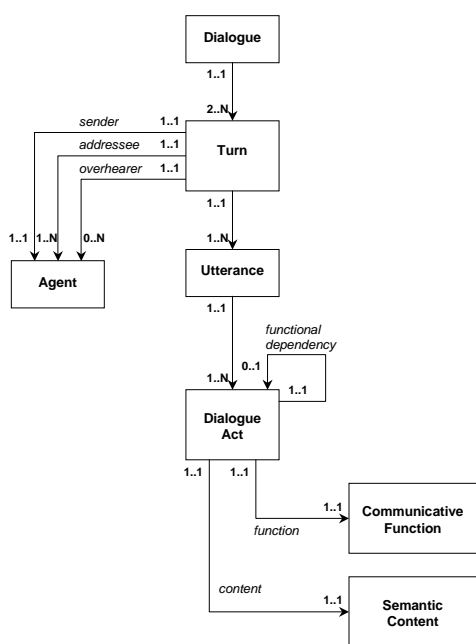
Figure 1 Preliminary metamodel for dialogue acts.

The metamodel shown in Figure **1** captures these considerations, with the additional inclusion of possible functional dependencies between dialogue acts, which encompass such things as indicating to which question a response is intended to be an answer, etc. Note also, and most importantly, that an utterance may correspond to multiple dialogue acts, due to the multidimensional nature of communication and the multifunctionality of natural language utterances (see Allwood, 2000; Bunt, 2000; Bunt & Girard, 2005).

The most important work that has been done to model concepts for dialogue annotation is that of the Discourse Research Initiative who produced the DAMSL annotation scheme (Allen & Core, 1997). Other efforts are those undertaken in the TRINDI project (see Larsson, 1998), in Dynamic Interpretation Theory (Bunt & Girard, 2005) and quite recently in the ISO TC37/SC 4 Task Domain Group 3

on Semantic Content Representation, which takes these various efforts into account (see reference in Footnote 1).

2.2. Temporal Information

There are currently two major developing standards for the annotation of temporal entities and relations: OWL-Time (which used to be called 'DAML-Time') and TimeML.

2.2.1 OWL-Time

OWL-Time (Hobbs & Pan, 2004) is an ontology that provides and defines a structure for the logical relationships between different temporal expressions, with the aim of marking up textual elements (primarily on the web) for the rapid extraction of 'surface-level' temporal information. It was developed as "an ontology of temporal concepts, for describing the temporal content of Web pages and the temporal properties of Web services" (Hobbs & Pan, 2004:1); in other words, mainly with the idea of reasoning about temporal events in mind, not for the annotation of natural language texts. As a consequence, only the static, topological qualities of temporal entities are captured. The idea was to create temporal constructs to enable web service providers to describe the semantic time properties of their services. The developers of OWL-Time were principally concerned with the representation of such information as, for example, the opening times of doctor's surgeries, the times of meetings across different time zones, the times/durations of theatre performances, etc.
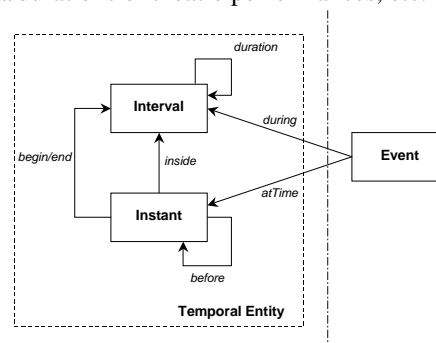
Figure 2 Model underlying OWL-Time.

We can visualise the basic model for OWL-Time as shown in Figure 2. The 'Event' element shown to the right of the dashed and dotted line is not in fact explicitly specified by Hobbs & Pan (2004), but is presumably the kind of 'element' that can take place in either an instant, an interval, or (if interrupted) a sequence of intervals. So, concerning the information to be extracted in the examples given earlier, the 'Event' might be equivalent to being seen by the doctor, or the meeting itself, or the play, and so on. Hobbs and Pan (2004) admit that in the future, simply marking up items in this way will be insufficient for retrieving all relevant content expressed in natural language form from the web.

OWL-Time does not provide an event ontology; Hobbs & Pan (2004) explicitly state that they believe that any event ontology should be kept separate from a time ontology. They make this their reason for not providing descriptions of event elements themselves. However, in Pan & Hobbs (2004:2), there is an attempt to link a temporal Instant with an Instant Event (a Punctual Event in the phraseology of TimeML), and a temporal Interval with an Interval Event (an Extended Event in TimeML). This

would seem to indicate that the developers of OWL-Time do indeed link events inextricably to temporal entities, even including both as children of a generic 'Temporal Thing'. How these two types of ontologies might be combined (i.e. by merging OWL-Time with TimeML) is explored in Hobbs & Pustejovsky (2003).

One conceptual problem with OWL-Time is its treatment of durations, which are viewed as intervals. For example, a duration of one week is modelled in OWL-Time as the concatenation of seven intervals of one day. This is conceptually confusing, however, since concatenation is an operation that applies to intervals, not to their lengths. A duration is the *length* of an interval, measured with the help of a certain unit; the relation between a duration of one week and one of seven days is not one of concatenation but one of equivalence, due to the possible conversion from weeks to days. More generally, a duration is an amount of time, and just like amounts in other dimensions (amount of weight, volume, velocity, etc.), it is formally an equivalence class defined through the conversions between units of the same dimension (see Bunt, 1985, Chapter 6).

To sum up, there are a number of reasons why OWL-Time on its own would not be a suitable base for the development of a standard for the linguistic annotation of temporal elements:

- OWL-Time was not designed with the full annotation of natural language in mind.
- It does not deal with deictic time, temporal aggregates (although see Pan & Hobbs (2005) and Pan (2005) for how this is being incorporated), or vagueness.
- There is no satisfactory treatment of durations.
- It has not been fully tested yet using a real-world application domain with a temporal reasoner.
- OWL-Time does not indicate the role of tense in the extraction of information from the web in any way, which would seem of crucial importance when trying to reason about whether an event has already occurred or whether it will occur at some point in the future.

Some of these criticisms and shortcomings of OWL-Time are addressed by the temporal annotation scheme proposed by the developers of TimeML.

2.2.2 TIMEML

TimeML provides a rather extensive scheme for temporal annotation of natural language texts. The annotation scheme not only marks up temporal expressions, but also provides links describing relations between these and the events being described (see Figure 3).
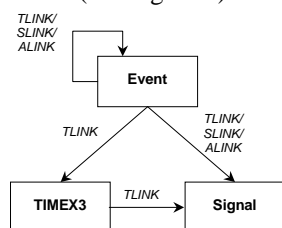

Figure 3 Model underlying TimeML (simplified).

Although TimeML does represent the temporal structures and relationships in natural language in a more complete way than the OWL-Time ontology, there is little indication of how the information marked up might be interpreted and reasoned with. The TimeML tags essentially do not have a semantics, which would be needed for temporal reasoning.

Hobbs & Pustejovsky (2003) suggest that the OWL-Time ontology can be used to give a semantics to TimeML annotations, and illustrate this with a number of examples. However, for certain cases that they do not discuss, unifying the schemes would prove problematic. This is for instance the case for tense information, which Hobbs & Pustejovsky mention as temporarily left out of consideration. While the two schemes are roughly complementary, there are some further difficulties to be faced when attempting to combine the two, as we shall see below.

Generally speaking, TimeML's strengths are OWL-Time's weaknesses, and vice versa. Some of the limitations of TimeML are as follows:

- TimeML tags in general do not have a semantics. This means that TimeML annotations are not a reliable basis for temporal reasoning.
- TimeML has some redundancy in allowing the temporal relations between events to be indicated both directly (using TLINK relations between events) and indirectly (via TLINK relations to their temporal anchoring).
- TimeML has no satisfactory treatment of durations. For instance, a duration of 30 minutes (when mentioned as such in a text) is annotated as <TIMEX3 (...) type="DURATION" value="PT30M" (...)>30 minutes</TIMEX3>. This makes it next to impossible to reason that this duration was half an hour.
- TimeML has no provision for time zones and daylight saving conventions, which OWL-Time has. This makes it difficult to produce annotations that denote exact times.
- The 'SIGNAL' tag in TimeML seems a kind of wastebasket, as it is used to annotate a wide variety of semantically very different entities, such as temporal prepositions, temporal modifiers ("twice", "every"), negatives ("not", "never"), modals ("might", "may"), and still others.

2.2.3 Merging OWL-Time and TIMEML

We are not here aiming to present a finalised metamodel for temporal entities and relations; but we do intend to indicate some of the issues that need to be addressed before such a unified metamodel can be developed.

One important respect in which the underlying models of TimeML and OWL-Time are not compatible is in how interrupted events are viewed as being anchored in time. OWL-Time allows an extended event to be interrupted, and thus to occur during a sequence of temporal intervals; TimeML, by contrast, considers the various parts of an interrupted event to be separate events, which are called 'event instances', and which are related as instances of the same 'event' through the function 'makeinstance'.

TimeML distinguishes three kinds of relations: SLINKs (subordination relations), ALINKs (aspectual relations) and TLINKs (temporal relations). Two events are for instance temporally related in cases like "He smiled while he looked at the picture", relating a 'smile' event and a 'look-at' event. Allowing such direct temporal relations between events creates redundancy in the model, since alternatively it would be possible to link the 'smile' event to a temporal interval that is included in the interval during which the 'look-at' event occurred. Similarly this is so for aspectual links. Once an integrated model of events and temporal entities is in place, as the result of merging the

underlying models of OWL-Time and TimeML, it seems best to represent all temporal relations between events through their anchoring in time. Only subordination relations between events remain, as they are fundamentally not just temporal relations, but rather semantic relations from which temporal relations may be inferred.

Besides dealing with these aspects of relating TimeML and OWL-Time, we would suggest amending OWL-Time's treatment of durations, by treating them as lengths of intervals, defined by the combination of a numeral and a unit (which is related to other units through a table of conversion).

Finally, in order to encompass the problem of anchoring an instant to a concrete point in time OWL-Time provides for the definition of a timestamp, with the time zone information included as part of the timestamp. However, although the timestamp is always interpretable with respect to the location of the event, this itself is not a part of the *temporal* model of information rather than of the *spatial*. The same observation applies to the adjustment of the timestamp for the purposes of daylight saving. The timestamp should inherit this information from the geographic location of the event, and so comparisons between different timestamps will require the ability to convert a timestamp of one location into the timestamp of another.

2.3. Reference Annotation

Reference annotation covers the annotation of coreferential, anaphoric and referential relations. Following van Deemter & Kibble (2000), we consider:

- *Reference* as a relation between a referring expression and a unique representation of its referent (but not another referring expression);
- *Coreference* as an equivalence relation between two (or more) referring expressions in respect to the identity of their referent (a special case of anaphoric relation);
- *Anaphoric relations* as various types of (not necessarily equivalent) semantic dependency relations between expressions, whose range depends strongly of the theoretical and practical goals of specific approaches and currently include: coreference, lexical relations, function-value relations, and even discourse relations or frame-element relations.

The annotation of these relations has already been the subject of considerable practical and theoretical research in order to establish basic principles and to unify different approaches for coherent and consistent annotation (see for example, amongst others, Hirschman & Chinchor, 1997; Davies and Poesio, 2000; Salmon-Alt & Romary, 2004a and Poesio, 2004).
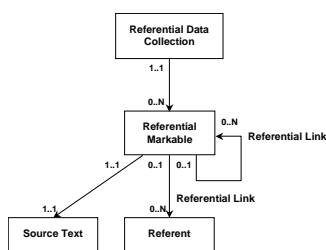

Figure 4 Reference metamodel.

Figure 4 shows an adapted version of the metamodel suggested by Salmon-Alt & Romary (2004b). The

Referential Data Collection component should be understood as the collection of all annotation data solely related to reference annotation. It should not to be confused with the annotated source text. Generally, linguistic data relevant to reference annotation is only a subset of the source text.

The *Reference Markable* component stands for any linguistic expression that instantiates a referential link or links within a text. The specific identification procedure of those chunks is left open. Therefore, the reference annotation metamodel provides a representation for linguistic objects to be linked independently from their actual selection procedure (manually or automatically annotated), their linguistic nature (noun phrases, clauses, referring expressions, bound anaphora, zero morphemes, "universe entities" (Davies & Poesio, 2000), etc.) and their physical location (contiguous or non contiguous text spans within the same text document or in an external text, video or audio source file). Because of the latter point, Referential Markables are associated with a mandatory data category *Source Text* that either contains the annotated text string (in-line annotation) or points to externalised source data (stand-off annotation). This feature allows for building markables recursively, i.e. combining elementary markables into more complex ones, as is sometimes required for the reference annotation of coordinated noun phrases.

The *Reference Link* relation represents any relation between Reference Markables that are relevant to reference annotation. Depending on the underlying theory, this could be coreference, reference or anaphoric relations as defined by van Deemter and Kibble (2000). A Reference Link relates a source markable (typically, the "anaphor") to a target markable (typically, the "antecedent") or referent.

In the design of data categories for coreference, it will be the types of referential expression and the types of relations that hold between them, and how they may be instantiated that will be of concern to the LIRICS project to define and exemplify in the definition of data categories. These types of referential expression and relation may also have considerable overlap with other levels of linguistic description, both within the level of semantics itself as well as with other levels such as morpho-syntax and syntax.

2.4. Semantic Roles

There have been intensive efforts to produce syntactically annotated text; however syntactic annotation by itself is not enough to represent meaning. For example:

(1) John broke the window.
(2) The window broke.

Syntax alone would tell us that 'the window' is the direct object of the verb in (1) and the subject in (2), but fails to represent that in both these cases, it is the thing that is broken, i.e. that even though it plays a different syntactic role, it plays the same semantic one.

Promising approaches to this representational problem are those based on frame semantics (Fillmore 1976) and the classes of verb defined by Levin (1993). In terms of annotation, there have been two major projects to annotate semantic roles utilising frame structures (primarily of verbs): FrameNet (Baker et al., 1998) and PropBank (Palmer et al., 2005). Both are built around definitions in VerbNet and WordNet, but are designed with different priorities in mind. FrameNet starts from the text itself with

no predetermined structure, whereas PropBank uses the pre-parsed data from the Penn Treebank. This has the side effect that in PropBank, every node in the parse tree is labelled with a semantic role of some sort, whereas in FrameNet, this is not the case. FrameNet is more concerned with the semantic roles of the constituents of a frame within a situation, and with how those frames are related to each other hierarchically. For the developers of PropBank, the main concern is the predicate-argument structures of syntactic constituents. PropBank's framesets are hinged on the verb; whereas the roles are fixed in FrameNet, in PropBank these are determined with respect to the verb.

Future work in the area of semantic roles within the LIRICS project will focus upon deciding what would be an appropriate set of role data categories by looking carefully at the roles defined in the FrameNet and PropBank schemes to see where there is a significant overlap and where a difference. Decisions will have to be made concerning the granularity of the data categories; for example, whether to concentrate on the types of roles (FrameNet) or the part that the roles play with respect to the verb (PropBank). It should be possible to develop some form of hierarchical structure to encompass both of these considerations. An attempt to demonstrate this is shown in Figure 5.
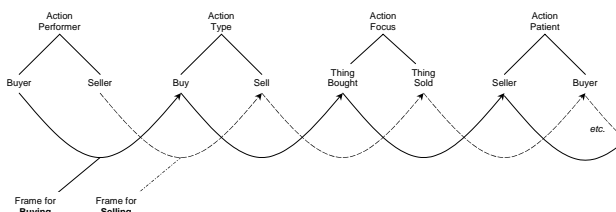
Figure 5 An abstracted model for the verb frames of buying and selling.

Whether it is possible to abstract so far away from the event type has yet to be determined. Perhaps the structure shown in Figure 5 will only fit the events of buying and selling because they are in effect the same transaction taken from differing points of view (the buyer's and the seller's). This is essentially what is done in the FrameNet scheme by treating 'buying' and 'selling' as instances of the frame of commerce, which has set role types. FrameNet however has nothing at all to say about the relationship between the semantic roles defined and the type of commerce (buying or selling), nor about how this will affect the order of these roles within the text.

The above example highlights a problem with the frame representation of semantics, namely that of granularity and of inheritance of characteristics. There is little point in defining frames that uniquely correspond with every verb and noun, because this would be expensive to search. We would ideally like to be able to infer certain patterns from the frames that are defined. So, in short, while a great deal of the work in the definition of semantic roles and descriptions has been carried out using frames as a basic structure, the status of frames themselves from a semantic point of view is not yet clear.

Again, as in the case of reference annotation, in the design of the data categories for semantic roles, it will be the *types* of roles and the *types* of relations that hold between them that will be of most importance to choose and define. How these are then instantiated and utilised in a specific annotation scheme is not of primary concern to the

LIRICS project. These types of semantic roles and relations may also have considerable overlap with other levels of linguistic description, both within the level of semantics itself (temporal information for instance) as well as with other levels, especially that of syntax.

At this stage, when we are principally considering the methodological aspects of approaching the problem of semantic role annotation, we need not go into further detail about the choice of semantic role labels that would be required for the provision of a comprehensive set of data categories for semantic roles. We note that, although FrameNet and PropBank approach semantic roles from different perspectives, these differences are largely cosmetic and there is a great overlap between these schemes.

When linguistic or multimodal input occurs, there is, first of all, a communicative event, which occurs at some point in time (or during a certain time interval). Second, it has at least two participants, who have different roles: sender and addressee. So we minimally want to distinguish (1) events; (2) temporal objects and relations for anchoring events in time; (3) entities participating in events; and (4) semantic roles relating events to participants. This corresponds to the metamodel depicted in Figure 6.
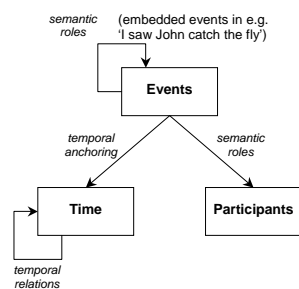
Figure 6 Metamodel of events, temporal objects, participants and semantic roles.

This metamodel contains semantic roles not only for relating events to participants, but also for relating events to other events. This is because events are needed not only for describing communicative events, but also for the semantic analysis of natural language utterances and sentences, where 'roles' correspond to semantic relations between a verb and its arguments. Such events may have other events as 'participants', as in the sentence 'I saw John catch the fly', where a see event has an embedded catch event. To avoid overlapping between the types of entities in different boxes, we intend the box 'participants' in Figure 6 to be 'participants which are themselves not events', and thus we get semantic roles as relations in two places.

3. Conclusion

We have put forward a number of fundamental, methodological considerations for the development of semantic data categories within the scope of the LIRICS project. We have stressed the importance of establishing metamodels as a first step to the design of annotation schemes and the specification of their elements in the form of data categories, and we have discussed potential metamodels for dialogue act annotation, temporal annotation, reference annotation, and semantic role annotation, and how they might fit together in a generic metamodel for semantic annotation. We also consider some other potentially relevant areas of semantic information.

All this is intended primarily to spark further discussion with and among experts in the various areas, rather than to suggest definitive metamodels or annotation schemes at this point.

When trying to design an annotation scheme there are a number of issues and problems to consider, not just in the annotation of semantic content, but also in any area in linguistics. There is a choice to be made between two opposing and mutually contradictory standpoints: whether to go for a simple, shallow approach, which is less expressive and less representative of the information to be modelled, but easier to annotate consistently and to process; or for a complex, in-depth approach which allows for more detailed expression and representation but is more difficult to annotate consistently and more 'expensive' to process. The complexity of annotation has an enormous effect on the ability to produce consistent and reliable markup of texts, both for human- and machine-annotation. Especially where it concerns natural language, with its inherent vagueness and ambiguity, a small weakness in specification will lead to multiple potential interpretations and therefore to multiple potential diverging applications of the annotation. Corollary to this, as most attempts to tag text automatically use human annotated texts to train the programs, complexity will also have consequences upon the efficacy of such taggers.

Developers of different annotation schemes within the same field, or even from within different but related fields, will often come to the task with different aims and objectives in mind, as well as from different theoretical standpoints. Different backgrounds will focus on different aspects, often conflating two or more distinct areas for annotation into one. In other words, there is a problem not only in deciding the granularity of an annotation scheme, but also what exactly it should encompass. It is especially for dealing with issues of this kind that the establishment of a metamodel can be of great help.

4. References

Allen, J. & M. Core (1997) *Draft of DAMSL: Dialog Act Markup in Several Layers*. Department. of Computer Science, University of Rochester, Rochester, NY.

Allwood, J. (2000). 'An activity-based approach to pragmatics'. In H. Bunt & W. Black (eds.) *Abduction, Belief and Context in Dialogue*. Benjamins, Amsterdam: 47-80.

Baker, C. F., C. J. Fillmore & J. B. Lowe (1998) 'The Berkley FrameNet Project' in *Proceedings of the COLING-ACL*, Montreal, Canada.

Bunt, H. (1985). *Mass terms and model-theoretic semantics*. Cambridge University Press, Cambridge, UK.

Bunt, H. (2000). 'Dialogue Pragmatics and Context Specification'. In H. Bunt & W. Black (eds.) *Abduction, Belief and Context in Dialogue*. Benjamins, Amsterdam: 81-150.

Bunt, H. & Y. Girard (2005). 'Designing an open, multimodal dialogue act taxonomy'. In C. Gardent & B Gaiffe (eds.), *Proceedings of the Ninth Workshop on the Semantics and Pragmatics of Dialogue (DIALOR'05)*, Nancy, June 2005.

Bunt, H. & L. Romary (2004). 'Standardization in multimodal content representation: Some methodological issues'. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, 2219-2222.

Davies, S. & M. Poesio (2000). 'Coreference'. In *MATE Dialogue Annotation Guidelines-Deliverable D2.1*, January 2000, 126-182.

Fillmore, C. J. (1976). 'Frame semantics and the nature of language'. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280: 20-32.

Gildea, D. & D. Jurafsky (2000). 'Automatic labeling of semantic roles'. In *Proceedings of ACL 2000*, Hong Kong.

Grosz, B.J., & Sidner, C.L. (1986). 'Attention, intentions, and the structure of discourse'. *Computational Linguistics*, 12 (3): 175-204.

Hirschman, L. & N. Chinchor (1997). 'MUC-7 coreference task definition'. In *MUC-7 Proceedings*.

Hobbs, J. R. & F. Pan (2004). 'An ontology of time for the semantic web'. *ACM Transactions on Asian Language Information Processing (TALIP)*: 66-85

Hobbs, J.R.,& J. Pustejovsky (2003). 'Annotating and Reasoning about Time and Events'. In *Proceedings of AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning,* Stanford, California, March 2003.

Jurafsky, D. & J. H. Martin (2000). *Speech and Language Processing.* Prentice-Hall, New York.

Larsson, S. (1998) 'Coding schemes for dialog moves'. Unpublished paper, available at http://www.ling.gu.se/sl

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press, Chicago.

Palmer, M., D. Gildea & P. Kingsbury (2005). 'The Proposition Bank: An Annotated Corpus of Semantic Roles'. Computational Linguistics, 31(1): 71-106.

Pan, F (2005). 'A Temporal Aggregates Ontology in OWL for the Semantic Web'. In *Proceedings of the AAAI Fall Symposium on Agents and the Semantic Web*, Arlington, Virginia, 2005.

Pan, F. & J. R. Hobbs (2004). 'Time in OWL-S'. In *Proceedings of the AAAI Spring Symposium on Semantic Web Services*, Stanford University, CA: 29-36.

Pan, F. & J. R. Hobbs (2005). 'Temporal aggregates in OWL-Time'. In *Proceedings of the 18th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2005)*, Clearwater Beach, Florida: 560-565, AAAI Press.

Poesio, M. (2004). 'The MATE/GNOME Proposals for Anaphoric Annotation, Revisited'. In *Proceedings of the 5th SIGdial Workshop*.

Pustejovsky, J., J. Castaño, R. Ingria, R. Saurí, R. Gaizauskas, A. Setzer & G. Katz (2003). 'TimeML: Robust specification of event and temporal expressions in text'. In *Proceedings of the Fifth International Conference on Computational Semantics*, Tilburg.

Salmon-Alt, S. & L. Romary (2004a). 'RAF – towards a reference annotation framework'. In *Proceedings of LREC 2004*.

Salmon-Alt, S. & L. Romary (2004b). 'Data Categories for a Normalized Reference Annotation Scheme'. In *5th International Conference on Discourse Anaphora and Anaphor Resolution* - DAARC'2004, São Miguel (Azores), Portugal. 2004.

Saurí, R. & J. Littman, B. Knippen, R. Gaizaukas, A. Setzer & J. Pustejovsky (2005). *TimeML Annotation Guidelines* (published online: http://www.timeml.org/-timemldocs.html, accessed October 2005)

van Deemter, K. & R. Kibble (2000). 'On Coreferring: Coreference in MUC and related annotation schemes'. *Computational Linguistics* 26(4).