Semantic Network Manual Annotation and its Evaluation

Václav Novák

Abstract

The present contribution is a brief extract of (Novák, 2008). The Prague Dependency Treebank (PDT) is a valuable resource of linguistic information annotated on several layers. These layers range from morphemic to deep and they should contain all the linguistic information about the text. The natural extension is to add a semantic layer suitable as a knowledge base for tasks like question answering, information extraction etc. In this paper I set up criteria for this representation, explore the possible formalisms for this task and discuss their properties. One of them, Multilayered Extended Semantic Networks (Multi-Net), is chosen for further investigation. Its properties are described and an annotation process set up. I discuss some practical modifications of MultiNet for the purpose of manual annotation. MultiNet elements are compared to the elements of the deep linguistic layer of PDT. The tools and problems of the annotation process are presented and initial annotation data evaluated.

1. Motivation

The longterm goal of the research in the field of Artificial Intelligence has been to create a machine which would understand natural language input and be able to perform the reasoning necessary to perform the desired actions. It is obvious that such a machine must be capable of storing the acquired information in its memory in a form suitable for the necessary reasoning. We will call this form the *knowledge representation*. Let's discuss the criteria which should be imposed upon the form of the information representation, and the existing systems for knowledge representation and their properties with respect to the given criteria.

There are several reasons why Tectogrammatical Representation (TR) may not be sufficient in a question answering system or machine translation:

1. There is no information about sorts of concepts represented by TR nodes. Sorts (the upper conceptual ontology) are an important source of constraints for semantic relations. Every relation has its signature which in turn reduces ambiguity in the process of text analysis and inferencing.

2. The syntactic functors Actor and Patient disallow creating inference rules for cognitive roles like *Affected object* or *State carrier*. For example, the axiom stating that an affected object is changed by the event cannot be feasibly expressed in the TR framework. However, if needed, this information can be stored in the lexicon for individual verb frames.

3. Lexemes of TR have no hierarchy; this limits especially the search for an answer in a question answering system. In TR there is no counterpart of SUB, SUBR, and SUBS MultiNet relations, which connect subordinate concepts to superordinate ones and individual object representatives to corresponding generic concepts.

4. In TR, each sentence is isolated from the rest of the text, except for coreference arrows connected to preceding sentences. This, in effect, complicates inferences combining knowledge from multiple sentences in one inference rule.

5. Nodes in TR always correspond to a word or a group of words in the surface form of a sentence or to a structure which is deleted on the surface (e.g., obligatory verb argument, coordination member). There are no means for representing knowledge generated during the inference process, if the knowledge does not have the form of a TR. For example, consider the axiom of temporal precedence transitivity (1):

$$(a \text{ ANTE } b) \land (b \text{ ANTE } c) \rightarrow (a \text{ ANTE } c) \tag{1}$$

In TR, we cannot add an edge denoting $(a \text{ ANTE } c)$. We would have to include a proposition like "a precedes c" as a whole new clause.

For all these reasons we need to extend our text annotation to a form suitable to more advanced tasks. It is shown in (Helbig, 2006) that MultiNet is capable of solving all the above mentioned issues.

2. Criteria

In order to efficiently retrieve and process the knowledge acquired in the form of natural language input, these criteria should be fulfilled by the internal knowledge representation format:[1]

  I. **Associativity**: The knowledge concerning a concept should be available without the necessity to iterate over the whole knowledge base. A representation lacking this property would not be scalable to real problems.

 II. **Local interpretability**: The knowledge necessary for interpretation of an object should be limited to an easily identifiable local neighborhood of the concept (the knowledge may include a contextual embedding which is crucial for the concept interpretation).

III. **Inference friendliness**: The knowledge data format should allow for further inclusion of new facts, acquired both by new texts and by automatic inferencing. A practical system should be robust with respect to contradictions to avoid a situation where every proposition is true.

---

[1]Criteria II., A., B. and C. are modifications of some of the criteria imposed by (Helbig, 2006). I formulated criteria I. and III.

Apart from the overall necessary requirements, there are also further criteria necessary for a representation if it is to be annotated manually:

A. **Consistency**: Analogous facts should be treated analogously.
B. **Cognitive Adequacy**: The representation must be understandable to the annotators and easy to visualize and review.
C. **Communicability**: The instructions should contain applicable operational criteria (Hajičová and Sgall, 1980), definitions, and standards.

The next requirement for the representational formalism is to integrate smoothly into the layered nature of the PDT (Karcevskij, 1929, Callmeier et al., 2004).

Why are these requirements crucial?

Without associativity (I.), the query for information would always require a search through the whole knowledge base. Furthermore, for queries which cannot be answered using only one sentence, one would have to create a kind of associative structure on the fly to make use of disambiguation, coreferences etc.

Local interpretability (II.) is needed for concepts embedded in a way that changes their mode of existence. Consider the clause "*If I were you*". We do not want to extract the information that *I* refers to the same person as *you*. However, this is what we would infer if we ignored the contextual embedding associated with the word *if*. Therefore the knowledge representation must ensure this information is readily available for every piece of information without the need to iterate through the whole knowledge base.

Inference friendliness (III.) allows us to enrich the acquired knowledge by applying inference rules. If we know that "*Mrs. Hill is the current vice president finance*", we can infer for instance that "*The current vice president finance is Mrs. Hill*". An inference friendly representation will allow a compact representation of such an inference. Without this compactness (e.g., in the case where the inference must be included as a whole new sentence) the scale of practical inferences would be very limited.

Without consistency (A.) the annotation process is unimaginable, because annotators are able to use only a limited set of instructions and they always treat the new sentences by analogy. If this were not the correct way to annotate, they could not produce meaningful results.

Cognitive adequacy (B.) is practical when the annotators must deal with complicated sentences. There are few people who understand modal operators and first order logic axioms, but there are many people who understand the sentences in *The Wall Street Journal*. Ideally, the complexity of annotating a sentence should be 100% correlated with the complexity of understanding its meaning. Without cognitive adequacy of the representation, the annotation cannot leave the realm of toy sentences.

Communicability (C.) is another key to the success of annotation. A mere learning by example can prove to be useful, but it fails in the case of border cases. Unfortunately, however contradictory this may sound, border cases make up a significant percentage of decisions and can be found in every Wall Street Journal sentence.

3. Existing Meaning Representations

In this section we will discuss various formalisms of knowledge representation and their conformance to the criteria presented in Section 2.

3.1. Representations Based on First Order Logic

The first attempts to formalize natural language were made using the predicate calculus (Frege, 1892). Since then various approaches have been trying to fix the problems of using first order logic purely extensional interpretation of the meaning. First, intensional semantics was developed (Montague, 1972) to introduce the notion of conceivable worlds. This theory was further developed in several directions:

- TIL: Transparent Intensional Logic (Tichý, 1988) aimed at further elaboration of the semantics of conceivable worlds
- Description Logic (Donini et al., 1996) focused on the computational aspects of meaning representation.
- DRT: Discourse Representation Theory (Eijk and Kamp, 1996) focused on the treatment of coreferences, quantifiers, and their interplay.
- Hybrid Modal Logic (Areces and Blackburn, 2001, Areces, Blackburn, and Marx, 2004, Blackburn, 2000, Blackburn, 2001) applied the framework of modal logic to natural language semantics.

All these formalisms have been used to represent real-life sentences. There has been a successful attempt to automatically create DRT structures proposed in (Bos, 2005). Hybrid Modal Logic has been investigated from the linguistic viewpoint in (Kruijff, 2001, Novák, 2004, Novák and Hajič, 2006). The TIL has been subject to automatic transduction (Horák, 2001), but not to manual annotation.

How do these systems fit into our criteria? They are very strong in associativity (I.): every concept is represented by one or more variables and these variables can be looked up easily. Inference friendliness (III.) is guaranteed as to the ease of addition of new knowledge: it can be added by simply adding predicates. On the other hand the robustness with respect to contradictions is addressed only in some of these systems and in general requires non-monotonicity of the reasoning.

Local interpretability (II.) is addressed only in DRT, where the relevant contextual embedding should be present only in the current box. Cognitive adequacy (B.) is the most difficult obstacle which prevents these systems from being manually annotated. The model-theoretic way of thinking and use of quantifiers are largely unintuitive. This is not apparent for sentences which are usually addressed in the relevant literature (e.g., "*Every farmer owns a donkey*"). Nevertheless, it emerges when we try to come up with a predicate calculus representation of an ordinary sentence like "*The U.S. trade representative, Carla Hills, announced …*" It seems unintuitive to think about *trade* as a function from possible worlds to a set of objects, which is the typical treatment for nouns.

3.2. Representations Based on Linguistic Structures

The meaning representations based on linguistic structures emerged as an extension of dependency syntax (Tesnière, 1934). There are various formalisms, which all share some common features: they start with the text or speech and transform it into formalized layers of representation, where the last layer should be the most suitable for the knowledge representation tasks. They are:

- Functional Generative Description (Sgall, Hajičová, and Panevová, 1986), where the highest layer of description is the Tectogrammatical Representation (Hajičová, Panevová, and Sgall, 2000)
- Robust Minimal Recursion Semantics (Copestake et al., 2005) as a pluggable layer of the framework of (Callmeier et al., 2004)
- Meaning-Text Theory (Mel'čuk, 1988, Bolshakov and Gelbukh, 2000), which is in many respects similar to the FGD framework (Žabokrtský, 2005).

These approaches have difficulties with respect to the inference friendliness (III.): to include a piece of inferred knowledge, we often have to add a whole new sentence which describes the fact. For example if we are to apply a rule stating the symmetry of a predicate in a logic-based system, we simply add one predicative statement for every instance. In a linguistics-based system, we have to copy the whole statement and transform it into the inverse form.

The next obstacle concerns the cognitive adequacy: the tree constraints force the annotators to choose only one connection where more of them could be applied: in "*They met during the concert on Tuesday.*" the above mentioned systems require the annotator to decide whether *on Tuesday* is connected to *met* or *concert*, although from the knowledge base viewpoint it would be ideal if both *met* and *concert* were connected with the temporal specification under consideration.

3.3. Semantic Networks

Semantic networks, as different from the logic-based systems as they may seem, have much in common with them. The semantic network, being a directed graph, can usually be turned into a set of formulae of predicate calculus. The main difference lies in the fact that the relationship between the predicates and the knowledge is not direct: the predicates encode information about the network. The elements of the network then carry their own meaning.

The main advantage of semantic networks is their concept-centeredness. As noted on page 4 of (Helbig, 2006), the difference is similar to the difference between a logical programming language (e.g., Prolog) and an object oriented programming language (e.g., Java). Every concept should correspond to a cognitive concept and it is assumed that two distinct concepts do not represent the same object, unless there is a piece of information indicating the opposite. On the other hand, in a model-theoretic framework, the model builders tend to create a model as small as possible, therefore collapsing the referents of all variables where possible. This, in effect, often leads to a wrong conclusion.

Individual semantic network formalisms differ in their repertoire of formal means. In prac-

tice, two systems have been used for purposes of natural language processing:

- KL-ONE: knowledge representation system (Brachman and Schmolze, 1985)
- MultiNet: Multilayered Extended Semantic Networks (Helbig, 2006)

They satisfy all the criteria presented in Section 2 and therefore they are discussed in the remaining chapters.

3.4. Semantic Web

A Semantic web is sometimes considered yet another semantic representation. However, it is more a framework allowing us to standardize the representations and exchange the data in a structured format. It is therefore not possible to simply create a semantic web corpus. The technologies being used are the Web Ontology Language (Horrocks and Patel-Schneider, 2004), which allows for standardization and exchange of ontologies, and Resource Description Framework (RDF Core Working Group, 2007), which is an XML-based data format for exchanging predicate-like structures.

4. Evaluation Metrics

Human annotations are usually evaluated against each other to measure the consistency of the annotation. The most common measures of agreement are accuracy (number of correct decisions divided by the number of all decisions) and F-measure (harmonic mean between the recall and the precision). However, these approaches suffer from the fact that some annotation agreement is present simply by chance. This fact was the reason to propose annotation agreement metrics corrected for the agreement by chance. First, Scott's π (Scott, 1955) and Cohen's κ (Cohen, 1960) were introduced. They were later generalized to the K coefficient of agreement (Carletta, 1996).

I do not use any of these corrections for three reasons:

1. The agreement metrics itself is difficult to develop and to obtain the most appropriate agreement score there is still much to do.
2. The agreement by chance is difficult to compute in such a complex situation. The probability that two annotators will produce exactly the same oriented graph with the same size and all the attributes is virtually zero.
3. The measures have no clear probabilistic interpretation (Artstein and Poesio, 2007).

When a stable level of annotator agreement is achieved and maintained, and the agreement measure is robust with respect to equivalent annotations, the metrics extended for hierarchical values should be used. An example is Krippendorf's α (Krippendorff, 1980).

5. Evaluation Data

The initial evaluation presented in this section has been carried out on a portion of *The Wall Street Journal* articles from the Penn Treebank (Marcus, Marcinkiewicz, and Santorini, 1993), which have been annotated on all the FGD layers and are available as the Prague English Dependency Treebank (Hajič et al., est. 2009). Initially, some sentences were used during the

training of annotators. These sentences were removed from the evaluation sample. The evaluation sample contains 67 annotated sentences (1793 words), annotated by two annotators, of which 46 sentences (1236 words) were annotated by three independent annotators. All annotators are native English speakers.

6. Structural Agreement

The structural agreement is measured for every sentence in isolation in two steps. First, the best match between the two annotators' graphs is found. Most of the graph nodes are connected to the tectogrammatical tree and for the remaining nodes, all possible one-to-one mappings are constructed and the optimal mapping w.r.t. the F-measure is selected. Second, the optimal mapping is used to compute the agreement.

Formally, we start with a set of tectogrammatical trees containing a set of nodes N. The annotation is a tuple $G = (V, E, T, A)$, where V are the vertices, $E \subseteq V \times V \times P$ are the directed edges and their labels (e.g., agent of an action: $\text{AGT} \in P$), $T \subseteq V \times N$ is the mapping from vertices to the tectogrammatical nodes, and finally A are attributes of the nodes. We simplified the problem by ignoring the mapping from edges to tectogrammatical nodes, the metaedges, and the MultiNet edge attribute *knowledge type*. Analogously, $G' = (V', E', T', A')$ is another annotation of the same sentence and our goal is to measure the similarity $s(G, G') \in [0, 1]$ of G and G'.

To measure the similarity we need a set Φ of admissible one to one mappings between vertices in the two annotations. A mapping is admissible if it connects vertices which are indicated by the annotators as representing the same tectogrammatical node:

$$\Phi = \left\{ \phi \subseteq V \times V' \middle| \right.$$

$$\forall_{\substack{n \in N \\ v \in V \\ v' \in V'}} \left(((v, n) \in T \wedge (v', n) \in T') \rightarrow (v, v') \in \phi \right)$$

$$\wedge \quad \forall_{\substack{v \in V \\ v', w' \in V'}} \left(((v, v') \in \phi \wedge (v, w') \in \phi) \rightarrow (v' = w') \right)$$

$$\left. \wedge \quad \forall_{\substack{v, w \in V \\ v' \in V'}} \left(((v, v') \in \phi \wedge (w, v') \in \phi) \rightarrow (v = w) \right) \right\} \qquad (2)$$

In Equation 2, the first condition ensures that Φ is constrained by the mapping induced by the links to the tectogrammatical layer. The remaining two conditions guarantee that Φ is a one-to-one mapping.

Then we can define the annotation agreement s as

$$s_{(G, G', m)} = F_m(G, G', \phi^*) \qquad (3)$$

where ϕ^* is the optimal mapping between nodes of alternative annotations:

$$\phi^* = \operatorname{argmax}_{\phi \in \Phi}(F_m(G, G', \phi)) \tag{4}$$

and F_m is the F1-measure:

$$F_m(G, G', \phi) = \frac{2 \cdot m(\phi)}{|E| + |E'|} \tag{5}$$

where $m(\phi)$ is the number of edges that match given the mapping ϕ. We use four variants of m, which gives us four variants of F and consequently four scores for every sentence:

Directed unlabeled:

$$m_{du}(\phi) = \left| \left\{ (v, w, \rho) \in E \middle| \exists_{v', w' \in V', \rho' \in P} \left((v', w', \rho') \in E' \right. \right. \right.$$
$$\left. \left. \left. \land (v, v') \in \phi \land (w, w') \in \phi \right) \right\} \right| \tag{6}$$

Undirected unlabeled:

$$m_{uu}(\phi) = \left| \left\{ (v, w, \rho) \in E \middle| \exists_{v', w' \in V', \rho' \in P} \left(\right. \right. \right.$$
$$\left((v', w', \rho') \in E' \lor (w', v', \rho') \in E' \right) \tag{7}$$
$$\left. \left. \left. \land (v, v') \in \phi \land (w, w') \in \phi \right) \right\} \right|$$

Directed labeled:

$$m_{dl}(\phi) = \left| \left\{ (v, w, \rho) \in E \middle| \exists_{v', w' \in V'} \left((v', w', \rho) \in E' \right. \right. \right.$$
$$\left. \left. \left. \land (v, v') \in \phi \land (w, w') \in \phi \right) \right\} \right| \tag{8}$$

Undirected labeled:

$$m_{ul}(\phi) = \left| \left\{ (v, w, \rho) \in E \middle| \exists_{v', w' \in V'} \left(\right. \right. \right.$$
$$\left((v', w', \rho) \in E' \lor (w', v', \rho) \in E' \right) \tag{9}$$
$$\left. \left. \left. \land (v, v') \in \phi \land (w, w') \in \phi \right) \right\} \right|$$
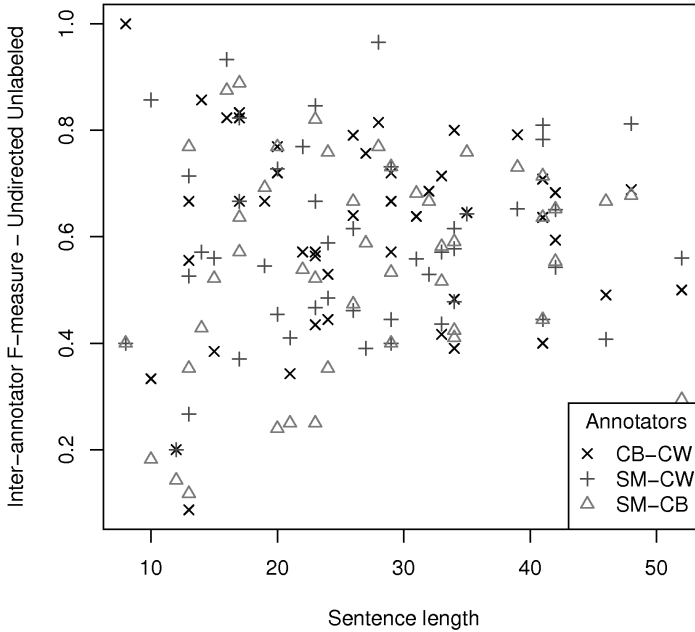
Figure 1. Inter-annotator agreement depending on the sentence length.

These four $m(\phi)$ functions give us four possible F_m measures, which allows us to have four scores for every sentence: s_{du}, s_{uu}, s_{dl} and s_{ul}.

Figure 1 shows that the agreement is not correlated with the sentence length. This means that longer sentences are on average no more difficult than short sentences. The variance decreases with the sentence length as expected.

In Figure 2 I present a comparison of directed and labeled evaluations with the undirected unlabeled case. By definition, the undirected unlabeled score is the upper bound for all the other scores. The directed score is well correlated and not very different from the undirected score, indicating that the annotators did not have much trouble with determining the correct direction of the edges. This might be in part due to support from the formalism and the *credit* tool: each relation type is specified by a sort signature; a relation that violates its signature is reported immediately to the annotator. On the other hand, labeled score is significantly lower than the unlabeled score, which suggests that the annotators have difficulties in assigning the

Sample	Annotators	Agreement F-measure			
		s_{uu}	s_{du}	s_{ul}	s_{dl}
Smaller	CB-CW	61.0	56.3	37.1	35.0
Smaller	SM-CB	54.9	48.5	27.1	25.7
Smaller	SM-CW	58.5	50.7	31.3	30.2
Smaller	average	58.1	51.8	31.8	30.3
Larger	CB-CW	64.6	59.8	40.1	38.5

Table 1. Inter-annotator agreement in percents. The results come from the two samples described in the Section 5.

correct relation types. The correlation coefficient between s_{uu} and s_{ul} (approx. 0.75) is also much lower than the correlation coefficient between s_{uu} and s_{du} (approx. 0.95).

Figure 3 compares individual annotator pairs. The scores are similar to each other and also have a similar distribution shape.

A more detailed comparison of individual annotator pairs shows that there is a significant positive correlation between scores, i.e., if two annotators can agree on the annotation, the third annotator is also likely to agree, but this correlation is not a very strong one. The actual correlation coefficient varies between 0.34 and 0.56. All the results are summarized in Table 1.
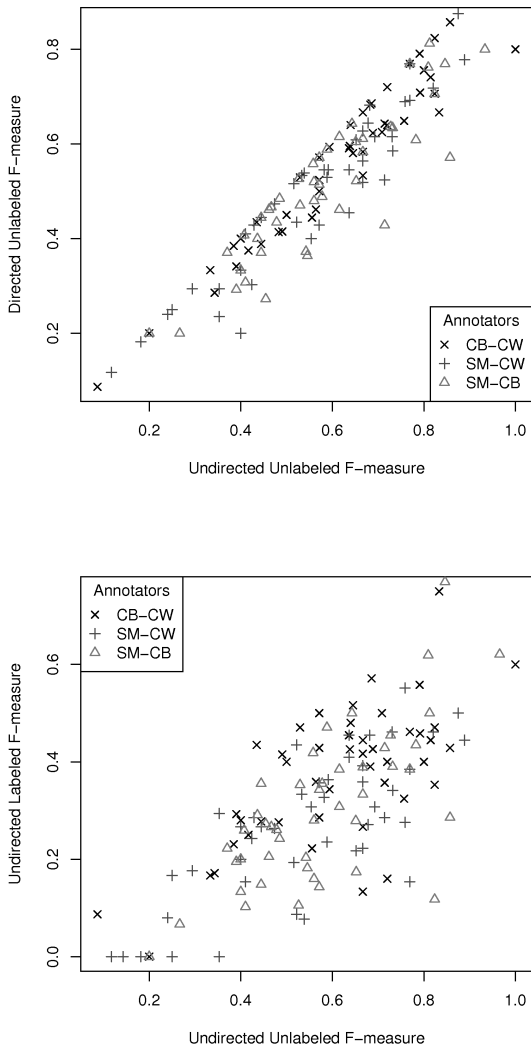
Figure 2. Upper: Directed vs. undirected inter-annotator agreement. Lower: Labeled vs. unlabeled inter-annotator agreement.
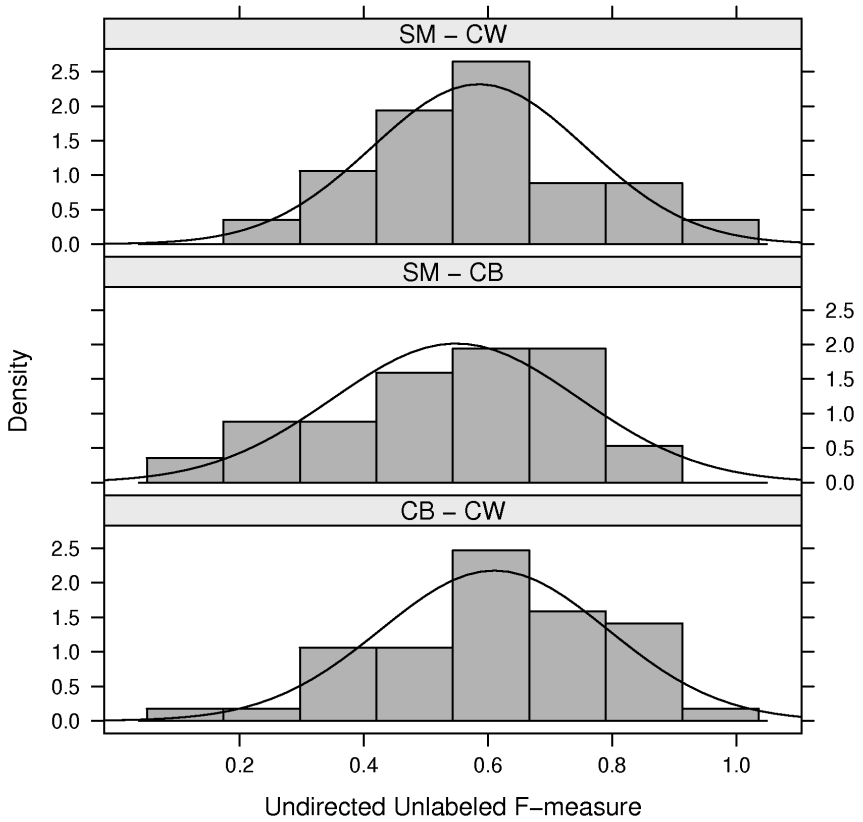
Figure 3. Comparison of individual annotator pairs.

Bibliography

Areces, Carlos and Patrick Blackburn. 2001. Bringing them all Together. *Journal of Logic and Computation*, 11(5):657–669.

Areces, Carlos, Patrick Blackburn, and Maarten Marx. 2004. Hybrid Logics: Characterization, Interpolation and Complexity. To appear in the Journal of Symbolic Logic, http://www.loria.fr/projets/hylo/Papers/jsl.pdf.

Artstein, Ron and Massimo Poesio. 2007. Inter-coder agreement for computational linguistics. *Computational Linguistics*, submitted.

Blackburn, Patrick. 2000. Representation, Reasoning, and Relational Structures: A Hybrid Logic Manifesto. *Logic Journal of the IGPL*, 8(3):339–625.

Blackburn, Patrick. 2001. Modal Logic As Dialogical Logic. *Synthese*, 127(1 - 2):57–93, April.

Bolshakov, Igor and Alexander Gelbukh. 2000. The Meaning-Text Model: Thirty Years After. *International Forum on Information and Documentation*, 1:10–16.

Bos, Johan. 2005. Towards Wide-Coverage Semantic Interpretation. In *Proceedings of Sixth International Workshop on Computational Semantics IWCS-6*, pages 42–53.

Brachman, Ronald and James Schmolze. 1985. An Overview of the KL-ONE Knowledge Representation System. *Cognitive Science*, 9:171–216.

Callmeier, Ulrich, Andreas Eisele, Ulrich Schäfer, and Melanie Siegel. 2004. The DeepThought Core Architecture Framework. In *Proceedings of LREC*, May.

Carletta, Jean. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, 22(2):249–254.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Copestake, Ann, Dan Flickinger, Carl Pollard, and Ivan Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation*, 3(4):281–332, December.

Donini, Francesco M., Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. 1996. Reasoning in Description Logics. In Gerhard Brewka, editor, *Principles of Knowledge Representation*. CSLI Publications, Stanford, California, pages 191–236.

Eijk, Jan and Hans Kamp. 1996. Representing Discourse in Context. In Johan Benthem and Alice Meulen, editors, *Handbook of Logic and Language*. Elsevier, Amsterdam, pages 179–237.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. In Mark Textor, editor, *Funktion - Begriff - Bedeutung*, volume 4 of *Sammlung Philosophie*. Vandenhoeck & Ruprecht, Göttingen.

Hajič, Jan, Kristýna Čermáková, Lucie Mladová, Anja Nedolužko, Jiří Semecký, Jana Šindlerová, Josef Toman, and Zdeněk Žabokrtský. est. 2009. Prague English Dependency Treebank (in progress).

Hajičová, Eva, Jarmila Panevová, and Petr Sgall. 2000. A Manual for Tectogrammatic Tagging of the Prague Dependency Treebank. Technical Report TR-2000-09, ÚFAL MFF UK, Prague, Czech Republic. in Czech.

Hajičová, Eva and Petr Sgall. 1980. Linguistic meaning and knowledge representation in automatic understanding of natural language. In *Proceedings of the 8th conference on Computational linguistics*, pages 67–75, Morristown, NJ, USA. Association for Computational Linguistics.

Helbig, Hermann. 2006. *Knowledge Representation and the Semantics of Natural Language.* Springer, Berlin, Germany.

Horák, Aleš. 2001. *The Normal Translation Algorithm in Transparent Intensional Logic for Czech.* Ph.D. thesis, Faculty of Informatics, Masaryk University, Brno, Czech Republic.

Horrocks, Ian and Peter F. Patel-Schneider. 2004. Reducing OWL Entailment to Description Logic Satisfiability. *Journal of Web Semantics*, 1(4):345–357.

Karcevskij, Sergei. 1929. Du dualisme asymétrique du signe linguistique. *Travaux du Cercle linguistique de Prague*, 1:88–93.

Krippendorff, Klaus. 1980. *Content analysis: an introduction to its methodology.* Sage Publications, Newbury Park, CA.

Kruijff, Geert Jan. 2001. *A Categorial-Modal Logical Architecture of Informativity.* Ph.D. thesis, Faculty of Mathematics and Physics, Charles University. http://www.coli.uni-sb.de/ gj/dissertation.phtml.

Marcus, Mitchell P., Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics*, 19(2):313–330.

Meľčuk, Igor A. 1988. *Dependecy Syntax: Theory and Practice.* State University of New York Press.

Montague, Richard. 1972. Pragmatics and Intensional Logic. *Semantics of Natural Language.*

Novák, Václav. 2004. Towards Logical Representation of Language Structures. *The Prague Bulletin of Mathematical Linguistics*, 82:5–86.

Novák, Václav. 2008. *Semantic Network Manual Annotation and its Evaluation.* Ph.D. thesis, Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic.

Novák, Václav and Jan Hajič. 2006. Perspectives of Turning Prague Dependency Treebank into a Knowledge Base. In *Proceedings of the LREC Conference*, Genova, Italy.

RDF Core Working Group. 2007. Resource Description Framework (http://www.w3.org/RDF/).

Scott, William A. 1955. Reliability of Content Analysis: the Case of Nominal Scale Coding. *Public Opinion Quarterly*, 19:321–325.

Sgall, Petr, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects.* D. Reidel, Dordrecht, The Netherlands.

Tesnière, Lucien. 1934. Comment construire une Syntaxe. *Bulletin de la Faculté des Lettres de Strasbourg*, pages 219–229.

Tichý, Pavel. 1988. *The Foundations of Frege's Logic.* Walter de Gruyter & Co, Berlin/New York.

Žabokrtský, Zdeněk. 2005. Resemblances between Meaning-Text Theory and Functional Generative Description. In *Proceedings of the 2nd International Conference of Meaning-Text Theory*, pages 549–557.