Valido: a Visual Tool for Validating Sense Annotations

Roberto Navigli

Dipartimento di Informatica
Università di Roma "La Sapienza"
Roma, Italy
`navigli@di.uniroma1.it`

Abstract

In this paper we present Valido, a tool that supports the difficult task of validating sense choices produced by a set of annotators. The validator can analyse the semantic graphs resulting from each sense choice and decide which sense is more coherent with respect to the structure of the adopted lexicon. We describe the interface and report an evaluation of the tool in the validation of manual sense annotations.

1 Introduction

The task of sense annotation consists in the assignment of the appropriate senses to words in context. For each word, the senses are chosen with respect to a sense inventory encoded by a reference dictionary. The free availability and, as a result, the massive adoption of WordNet (Fellbaum, 1998) largely contributed to its status of *de facto* standard in the NLP community. Unfortunately, WordNet is a fine-grained resource, which encodes possibly subtle sense distictions.

Several studies report an inter-annotator agreement around 70% when using WordNet as a reference sense inventory. For instance, the agreement in the Open Mind Word Expert project (Chklovski and Mihalcea, 2002) was 67.3%. Such a low agreement is only in part due to the inexperience of sense annotators (e.g. volunteers on the web). Rather, to a large part it is due to the difficulty in making clear which are the real distinctions between close word senses in the WordNet inventory.

Adjudicating sense choices, i.e. the task of validating word senses, is therefore critical in building a high-quality data set. The validation task can be defined as follows: let w be a word in a sentence σ, previously annotated by a set of annotators $A = \{a_1, a_2, ..., a_n\}$ each providing a sense for w, and let $S_A = \{s_1, s_2, ..., s_m\} \subseteq Senses(w)$ be the set of senses chosen for w by the annotators in A, where $Senses(w)$ is the set of senses of w in the reference inventory (e.g. WordNet). A validator is asked to validate, that is to adjudicate a sense $s \in Senses(w)$ for a word w over the others. Notice that s is a word sense for w in the sense inventory, but is not necessarily in S_A, although it is likely to be. Also note that the annotators in A can be either human or automatic, depending upon the purpose of the exercise.

2 Semantic Interconnections

Semantic graphs are a notation developed to represent knowledge explicitly as a set of conceptual entities and their interrelationships. Fields like the analysis of the lexical text cohesion (Morris and Hirst, 1991), word sense disambiguation (Agirre and Rigau, 1996; Mihalcea and Moldovan, 2001), ontology learning (Navigli and Velardi, 2005), etc. have certainly benefited from the availability of wide-coverage computational lexicons like WordNet (Fellbaum, 1998), as well as semantically annotated corpora like SemCor (Miller et al., 1993).

Recently, a knowledge-based algorithm for Word Sense Disambiguation, called *Structural Semantic Interconnections*[1] (SSI) (Navigli and Velardi, 2004), has been shown to provide interesting insights into the choice of word senses by providing structural justifications in terms of semantic graphs.

SSI exploits an extensive lexical knowledge base, built upon the WordNet lexicon and enriched with collocation information representing seman-

---

[1]SSI is available online at http://lcl.di.uniroma1.it/ssi.

tic relatedness between sense pairs. Collocations are acquired from existing resources (like the Oxford Collocations, the Longman Language Activator, collocation web sites, etc.). Each collocation is mapped to the WordNet sense inventory in a semi-automatic manner and transformed into a *relatedness* edge (Navigli and Velardi, 2005).

Given a word context $C = \{w_1, ..., w_k\}$, SSI builds a graph $G = (V, E)$ such that $V = \bigcup_{i=1}^{k} Senses_{\text{WN}}(w_i)$ and $(s, s') \in E$ if there is at least one semantic interconnection between s and s' in the lexical knowledge base. A *semantic interconnection pattern* is a relevant sequence of edges selected according to a manually-created context-free grammar, i.e. a path connecting a pair of word senses, possibly including a number of intermediate concepts. The grammar consists of a small number of rules, inspired by the notion of lexical chains (Morris and Hirst, 1991). An excerpt of the context-free grammar encoding semantic interconnection patterns for the WordNet lexicon is reported in Table 1. For the full set of interconnections the reader can refer to Navigli and Velardi (2004).

SSI performs disambiguation in an iterative fashion, by maintaining a set C of senses as a semantic context. Initially, $C = V$ (the entire set of senses of words in C). At each step, for each sense s in C, the algorithm calculates a score of the degree of connectivity between s and the other senses in C:

$$Score_{SSI}(s, C) = \frac{\sum_{s' \in C \setminus \{s\}} \sum_{i \in IC(s,s')} \frac{1}{length(i)}}{\sum_{s' \in C \setminus \{s\}} |IC(s,s')|}$$

where $IC(s, s')$ is the set of interconnections between senses s and s'. The contribution of a single interconnection is given by the reciprocal of its length, calculated as the number of edges connecting its ends. The overall degree of connectivity is then normalized by the number of contributing interconnections. The highest ranking sense s of word w is chosen and the senses of w are removed from the semantic context C. The algorithm terminates when either $C = \emptyset$ or there is no sense such that its score exceeds a fixed threshold.

3 The Tool: Valido

Based on SSI, we developed a visual tool, *Valido*[2], to visually support the validator in the difficult task

---

²Valido is available at http://lcl.di.uniroma1.it/valido.

$S \rightarrow S'S_1|S'S_2|S'S_3$
  (start rule)
$S' \rightarrow e_{nominalization}|e_{pertainymy}|\epsilon$
  (part-of-speech jump)
$S_1 \rightarrow e_{kind-of}S_1|e_{part-of}S_1|e_{kind-of}|e_{part-of}$
  (hyperonymy/meronymy)
$S_2 \rightarrow e_{kind-of}S_2|e_{relatedness}S_2|e_{kind-of}|e_{relatedness}$
  (hypernymy/relatedness)
$S_3 \rightarrow e_{similarity}S_3|e_{antonymy}S_3|e_{similarity}|e_{antonymy}$
  (adjectives)

Table 1: An excerpt of the context-free grammar for the recognition of semantic interconnections.

of assessing the quality and suitability of sense annotations. The tool takes as input a corpus of documents whose sentences were previously tagged by one or more annotators with word senses from the WordNet inventory. The corpus can be input in xml format, as specified in the initial page.

The user can browse the sentences, and adjudicate a choice over the others in case of disagreement among the annotators. To the end of assisting the user in the validation task, the tool highlights each word in a sentence with different colors, namely: *green* for words having a full agreement, *red* for words where no agreement can be found, *orange* for those words on which a validation policy can be applied.

A validation policy is a strategy for suggesting a default sense choice to the validator in case of disagreement. Initially, the validator can choose one of four validation policies to be applied to those words with disagreement on which sense to assign:

(α) **majority voting**: if there exists a sense $s \in S_A$ (the set of senses chosen by the annotators in A) such that $\frac{|\{a \in A \mid a\ annotated\ w\ with\ s\}|}{|A|} \geq \frac{1}{2}$, s is proposed as the preferred sense for w;

(β) **majority voting + SSI**: the same as the previous policy, with the addition that if there exists no sense chosen by a majority of annotators, SSI is applied to w, and the sense chosen by the algorithm, if any, is proposed to the validator;

(γ) **SSI**: the SSI algorithm is applied to w, and the chosen sense, if any, is proposed to the validator;

(δ) **no validation**: w is left untagged.

Notice that for policies (β) and (γ) Valido applies the SSI algorithm to w in the context of its

sentence σ by taking into account for disambiguation only the senses in s (i.e. the set of senses chosen by the annotators). In general, given a set of words with disagreement $W \subseteq \sigma$, SSI is applied to W using as a fixed context the agreed senses chosen for the words in $\sigma \setminus W$.

Also note that the suggestion of a sense choice, marked in orange based on the validation policy, is just a proposal and can freely modified by the validator, as explained hereafter.

Before starting the interface, the validator can also choose whether to add a virtual annotator a_{SSI} to the set of annotators A. This virtual annotator tags each word $w \in \sigma$ with the sense chosen by the application of the SSI algorithm to σ. As a result, the selected validation policy will be applied to the new set of annotators $A' = A \cup \{a_{SSI}\}$. This is useful especially when $|A| = 1$ (e.g. in the automatic application of a single word sense disambiguation system), that is when validation policies are of no use.

Figure 1 illustrates the interface of the tool: in the top pane the sentence at hand is shown, marked with colors as explained above. The main pane shows the semantic interconnections between senses for which either there is a full agreement or the chosen validation policy can be applied. When the user clicks on a word w, the left pane reports the sense inventory for w, including information about the hypernym, definition and usage for each sense of w. The validator can then click on a sense and see how the semantic graph shown in the main pane changes after the selection, possibly resulting in a different number and strength of semantic interconnection patterns supporting that sense choice. For each sense in the left pane, the annotators in A who favoured that choice are listed (for instance, in the figure annotator #1 chose sense #1 of *street*, while annotator #2 as well as SSI chose sense #2).

If the validator decides that a certain word sense is more convincing based on its semantic graph, (s)he can select that sense as a final choice by clicking on the *validate* button on top of the left pane. In case the validator wants to validate present sense choices of all the disagreed words, (s)he can press the *validate all* button in the top pane. As a result, the present selection of senses will be chosen as the final configuration for the entire sentence at hand.

In the top pane, an icon beside each disagreed

	Precision	Recall
Nouns	75.80% (329/434)	63.75% (329/516)
Adjectives	74.19% (46/62)	22.33% (46/206)
Verbs	65.64% (107/163)	43.14% (107/248)
Total	73.14% (482/659)	49.69% (482/970)

Table 2: Results on 1,000 sentences from SemCor.

word shows the validation status of the word: a *question mark* indicates that the disagreement has not yet been solved, while a *checkmark* indicates that the validator solved the disagreemeent.

4 Evaluation

We briefly report here an experiment on the validation of manual sense annotations with the aid of Valido. For more detailed experiments the reader can refer to Navigli (2006).

1,000 sentences were uniformly selected from the set of documents in the semantically-tagged SemCor corpus (Miller et al., 1993). For each sentence $\sigma = w_1 w_2 \dots w_k$ annotated in SemCor with the senses $s_{w_1} s_{w_2} \dots s_{w_k}$ ($s_{w_i} \in Senses(w_i), i \in \{1, 2, \dots, k\}$), we randomly identified a word $w_i \in \sigma$, and chose at random a different sense $\bar{s}_{w_i}$ for that word, that is $\bar{s}_{w_i} \in Senses(w_i) \setminus \{s_{w_i}\}$. In other words, we simulated *in vitro* a situation in which an annotator provides an appropriate sense and the other selects a different sense.

We applied Valido with policy (γ) to the annotated sentences and evaluated the performance of the approach in suggesting the appropriate choice for the words with disagreement. The results are reported in Table 2 for nouns, adjectives, and verbs (we neglected adverbs as very few interconnections can be found for them).

The experiment shows that evidences of inconsistency due to inappropriate annotations are provided with good precision. The overall F1 measure is 59.18%. The chance baseline is 50%.

The low recall obtained for verbs, but especially for adjectives, is due to a lack of connectivity in the lexical knowledge base, when dealing with connections across different parts of speech.

5 Conclusions

In this paper we presented Valido, a tool for the validation of manual and automatic sense annotations. Valido allows a validator to analyse the coherency of different sense annotations provided for the same word in terms of the respective semantic interconnections with the other senses in context. We reported an experiment showing that
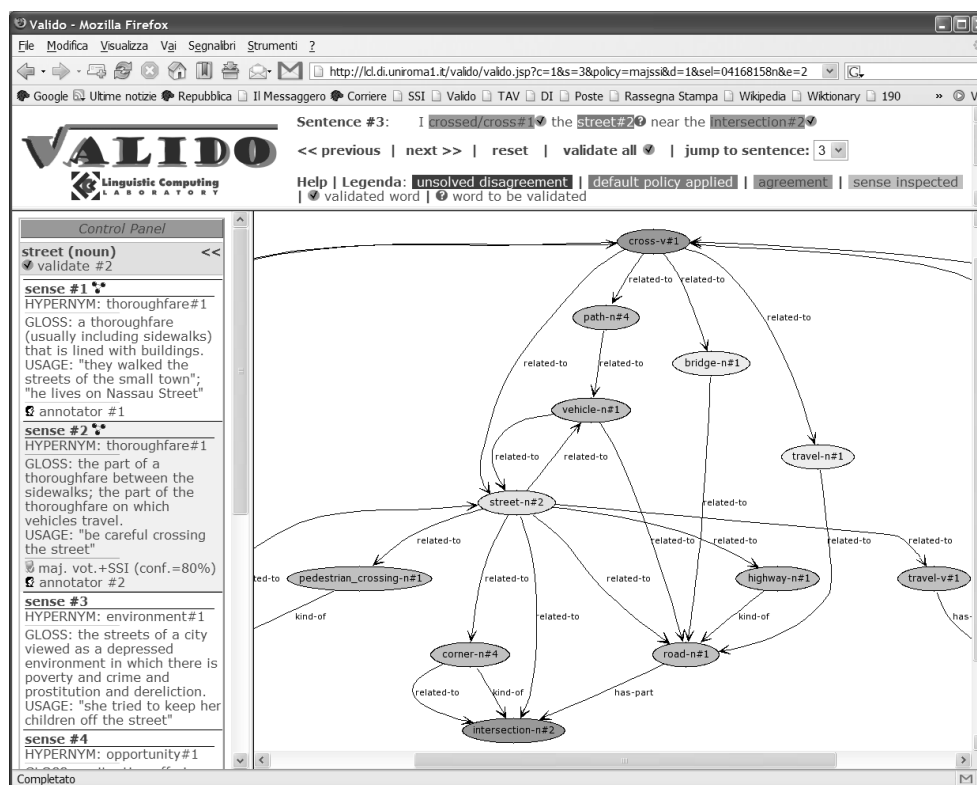
Figure 1: A screenshot of the tool.

the approach provides useful hints. Notice that this experiment concerns the quality of the suggestions, which are not necessarily taken into account by the validator (implying a higher degree of accuracy in the overall validation process).

We foresee an extension of the tool for supporting the sense annotation phase. The tool can indeed provide richer information than interfaces like the Open Mind Word Expert (Chklovski and Mihalcea, 2002), and the annotator can take advantage of the resulting graphs to improve awareness in the decisions to be taken, so as to make consistent choices with respect to the reference lexicon.

Finally, we would like to propose the use of the tool in the preparation of at least one of the test sets for the next Senseval exercise, to be held supposedly next year.

Acknowledgments

References

Eneko Agirre and German Rigau. 1996. Word sense disambiguation using conceptual density. In *Proc. of COLING 1996*. Copenhagen, Denmark.

Tim Chklovski and Rada Mihalcea. 2002. Building a sense tagged corpus with open mind word expert. In *Proc. of ACL 2002 Workshop on WSD: Recent Successes and Future Directions*. Philadelphia, PA.

Christiane Fellbaum, editor. 1998. *WordNet: an Electronic Lexical Database*. MIT Press.

Rada Mihalcea and Dan Moldovan. 2001. Automatic generation of a coarse grained wordnet. In *Proc. of NAACL Workshop on WordNet and Other Lexical Resources*. Pittsburgh, PA.

George Miller, Claudia Leacock, Tengi Randee, and Ross Bunker. 1993. A semantic concordance. In *Proc. 3rd DARPA Workshop on Human Language Technology*. Plainsboro, New Jersey.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1).

Roberto Navigli and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated websites. *Computational Linguistics*, 30(2).

Roberto Navigli and Paola Velardi. 2005. Structural semantic interconnections: a knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(7).

Roberto Navigli. 2006. Experiments on the validation of sense annotations assisted by lexical chains. In *Proc. of the European Chapter of the Annual Meeting of the Association for Computational Linguistics (EACL)*. Trento, Italy.