Supporting Large-Scale Knowledge Acquisition with Structural Semantic Interconnections

Roberto Navigli

Dipartimento di Informatica
Università di Roma "La Sapienza"
Via Salaria, 113 - 00198 Roma Italy
navigli@di.uniroma1.it

Abstract

In this paper, we discuss the use of a semantic disambiguation algorithm, Structural Semantic Interconnections (SSI), as a tool to support the process of semantic knowledge collection from a web community of volunteers. Starting from implicit knowledge in the form of sentences, terminology or collocations, SSI provides suggestions for sense selection in the form of semantic graphs that volunteers in a distributed environment can individually access. If the suggestion conveys a strong meaning, the majority of users working on that instance is expected to accept it, thus smoothing possible divergences and supporting consistent decisions. Otherwise, the volunteer can still employ SSI as a visual support for comparing other sense choices and make the most appropriate selection.

As a result, in both cases, the use of semantic interconnections as a support for sense selection should guide the intuition of human volunteers and reduce the number of inconsistencies.

Valido, an interface based on the employment of SSI for semantic knowledge acquisition, currently being developed in our laboratory, is described in the last part of the paper.

1. Introduction

Many knowledge-intensive tasks in the field of Artificial Intelligence (word sense disambiguation, commonsense reasoning, question answering, etc.) would benefit from the availability of a wide-coverage, structured or partly structured knowledge (in the form of ontologies, knowledge bases, glossaries, linguistically annotated corpora, and so on).

In the age of the Internet, the so called knowledge acquisition bottleneck can be overcome by collecting such knowledge in either an automatic (e.g. from the web) or a manual way.

While the classical manual approach is costly and time-consuming when performed by experts (e.g. lexicographers), human volunteers spread all over the network or grouped in a community could be involved in the distributed, large-scale collection of knowledge. Such a contribution would drastically speed up the acquisition process, but should be carefully planned in order to avoid problems of inconsistency between users, collection of useless or incorrect knowledge, etc. Furthermore, we cannot expect a user to be an expert in knowledge crafting or semantic annotation of texts.

The development of a user-friendly interface or some kind of visual support would certainly help human volunteers without specific competences provide valuable contributions.

The choice of involving the general public in the process of knowledge collection raises a number of additional interesting problems like the quality assessment of the knowledge collected, the detection and rejection of low-quality contributions, the control and collaboration of a distributed community of contributing users, and so forth.

Previous works, like the *Open Mind Common Sense* (Singh, 2002) and the *Open Mind Word Expert* projects (Chklovski and Mihalcea, 2002) focused respectively on the acquisition of commonsense knowledge and semantically annotated corpora from people. In the latter project *active learning* techniques are used to select for annotation only those examples that are the most informative.

Another interesting approach is *Learner2* (Chklovski, 2005), a system for collecting knowledge about common real-world objects, focusing on the collection of information about their parts and uses.

Argumentation systems (Conklin and Begeman, 1988; Lowrance et al., 2001; Suthers, 2004; Chklovski, Ratnakar, and Gil, 2005) are also related to knowledge collection, in that they aim at acquiring (possibly structured) information about the users' opinions. They usually present information in a textual or graphical fashion with the aim of minimizing the user effort in the formalization of knowledge during argumentation.

In this paper we discuss the use of a semantic disambiguation algorithm, Structural Semantic Interconnections (SSI) (Navigli and Velardi, 2004; Navigli and Velardi, 2004b), as a tool to support knowledge collection from a community of volunteers in a loosely supervised environment. Starting from implicit knowledge in the form of sentences, terminology or collocations, SSI provides suggestions for sense selection in the form of semantic graphs that volunteers can individually access. If the suggestion conveys a strong meaning, the majority of users working on that instance is expected to accept it, thus smoothing possible divergences and supporting consistent decisions. Otherwise, the volunteer can still employ SSI as a graphical support for comparing other sense choices.

In both cases, the use of semantic interconnections as a support for sense selection should guide the intuition of human volunteers and reduce the number of inconsistencies.

The paper is organized as follows: in Section 2 we introduce the Structural Semantic Interconnection algorithm. In Section 3 we present its uses as a support for knowledge acquisition. In Section 4 we describe a tool based on SSI for the support of semantic knowledge collection from knowledge volunteers. Section 5 concludes with a discussion of our contribution and future work.

2. Structural Semantic Interconnections

SSI (Structural Semantic Interconnections) is a word sense disambiguation algorithm based on structural pattern matching (Bunke and Sanfeliu, 1990). It disambiguates words in contexts using a “core” semantic knowledge base including WordNet (Fellbaum, 1998) and other available semantic resources. The algorithm relies on a context-free grammar describing “basic” lexico-semantic patterns, used to detect semantic relations among senses of the words appearing in a context (Navigli and Velardi, 2004; Navigli and Velardi, 2004b). Typical semantic patterns, inspired by (Hirst and St-Onge, 1998; Mihalcea and Moldovan, 2001) and subsequent works, are meaningful sequences of edges, e.g. chains of *hypernymy* (kind-of) and *meronymy* (has-part) edges, *hypernymy* edges following a *gloss* or *example* edge (i.e. an edge connecting a concept to another concept mentioned in its textual definition or in a dictionary example, respectively), and so on. Some examples of semantic patterns are shown in Figure 1(a)[1].

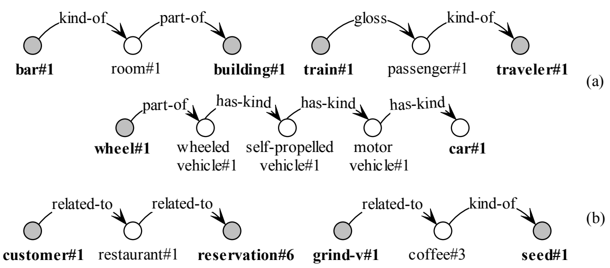

Figure 1. Examples of semantic patterns of knowledge.

In a recent work (Navigli, 2005), we enriched the SSI knowledge base with relatedness relations connecting concepts otherwise unrelated within WordNet (e.g. *newspaper#1* and *advertisement#1*, *fruit#1* and *tree#1*, *cloud#2* and *sky#1*), and we improved the grammar with new patterns involving such relations (some examples are reported in Figure 1(b)).

The SSI algorithm consists of an initialisation and an iterative step. In a generic iteration of the algorithm the input is a list of co-occurring terms $T = [\ t_1, \ldots, t_n\ ]$ and a list of associated senses $I = [S^{t_1},...,S^{t_n}]$, i.e. the semantic interpretation of *T*, where S^{t_i} is either the chosen sense for t_i (i.e., the result of a previous disambiguation step) or the *null* element (i.e., the term is not yet disambiguated).

A set of pending terms is also maintained, $P = \{ t_i \mid S^{t_i} = null \}$. *I* is named the semantic context of *T* and is used, at each step, to disambiguate new terms in *P*.

The algorithm works in an iterative way, so that at each stage either at least one term is removed from *P* (i.e., at least a pending term is disambiguated) or the procedure stops because no more terms can be disambiguated. The output is the updated list *I* of senses associated with the input terms *T*.

Initially, the list *I* includes the senses of monosemous terms in *T*. If no monosemous terms are found, the algorithm makes an initial guess based on the most probable sense of the less ambiguous term. The initialisation policy is further adjusted depending upon the specific task considered.

During a generic iteration, the algorithm selects those terms *t* in *P* showing an interconnection between at least one sense *S* of *t* and one or more senses in *I*. Relevant interconnections are encoded in a context-free grammar describing meaningful lexico-semantic patterns. The likelihood for a sense *S* of being the correct interpretation of *t* is given by a function of the weights of patterns connecting *S* to other synsets in *I*.

As an example, consider the following initial context *T*, given by [*taxi*, *fare*, *passenger*, *car*, *license*, *driver*, *medallion*, *minicab*]. *I* is initialised to [*taxi#1*, -, *passenger#1*, -, -, -, -, *minicab#1*] (*taxi*, *passenger* and *minicab* are monosemous). The outcome of SSI is the set *I* = [*taxi#1*, *fare#3*, *passenger#1*, *car#1*, *license#1*, *driver#1*, *medallion#3*, *minicab#1*]. The semantic patterns identified by SSI are illustrated in Figure 2. For an extensive running example, the interested reader can refer to the bibliography.

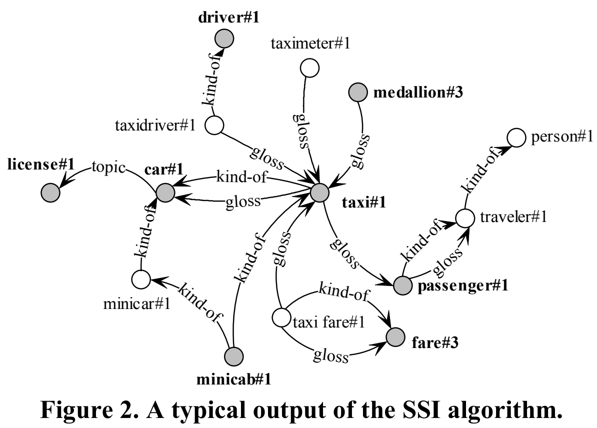

Figure 2. A typical output of the SSI algorithm.

[1] With *w#i*, *w-v#i*, *w-a#i* we denote, respectively, the *i*-th sense of the noun, verb and adjective *w* in WordNet.

As a second example, consider the semantic network in Figure 3. The two concepts in gray, *computer#1* and *terminal#3*, are interconnected by a number of edge sequences, i.e. knowledge patterns, of different weight. This graph represents the justification for selecting senses 1 and 3 of *computer* and *terminal*, respectively, as a conceptualization of the multiword expression *computer terminal*.

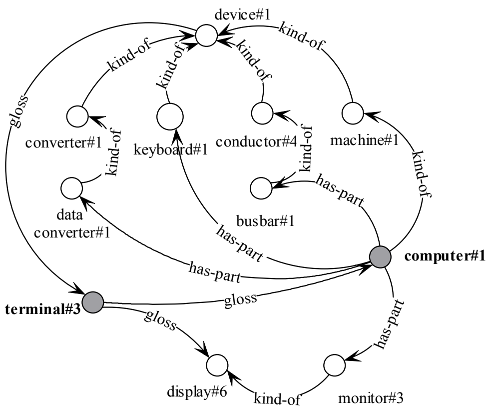

Figure 3. Semantic interconnections between *terminal#3* and *computer#1*.

SSI produces a justification of its sense choices in terms of the detected semantic patterns and their weights, since not all the patterns equally contribute to the choice of a specific sense.
Patterns like those shown in Figures 2 and 3 provide, in our view, relevant support to human annotators in that they constitute a visual aid that the volunteer can exploit to reason on the possible sense choices.

3. SSI as a Support for Knowledge Acquisition

In this section we describe the employment of SSI as a support for human volunteers in the following knowledge acquisition tasks: semantic annotation (Section 3.1), glossary construction (Section 3.2), and enrichment of knowledge bases (Section 3.3). These applications, detailed in the subsequent sections, can all be applied in a loosely supervised environment, so that on one side human volunteers can provide their own contribution independently, but on the other side their intuition is guided by the suggestions proposed by the SSI algorithm, thus reducing the possibility of inconsistencies (Section 3.4).
Finally, we describe an application of SSI to the detection of multiple levels of interpretation of sentences (Section 3.5).

3.1. Supporting Semantic Annotations

The availability of large semantically annotated corpora is critical for the success of Word Sense Disambiguation systems. Among the large-scale hand tagging efforts we mention SemCor (Miller et al., 1993), the DSO corpus (Ng and Lee, 1996) and, lately, the training and test sets provided during the Senseval WSD competitions[2].
In this section we propose an approach to the distributed collection of semantic annotations on a large scale, providing the human annotators with a visual support consisting of semantic graphs that connect the senses chosen for the words contained in a sentence. Given an arbitrary sentence, the SSI algorithm selects word senses based on the strongest semantic interconnections (scored with the aid of pattern weights). For instance, consider the following sentence from the Senseval-3 English All-Words task: “Brakes howled and a horn blared furiously”. The application of SSI to this sentence produces a number of interesting interconnections (reported in Figure 4) between the appropriate senses of *brake*, *horn*, *howl* and *blare*: *brake#1* and *horn#10* are both devices, *howl-v#3* refers to vehicles (and cars have horns), the gloss of *blare-v#2* mentions horns, and so on.

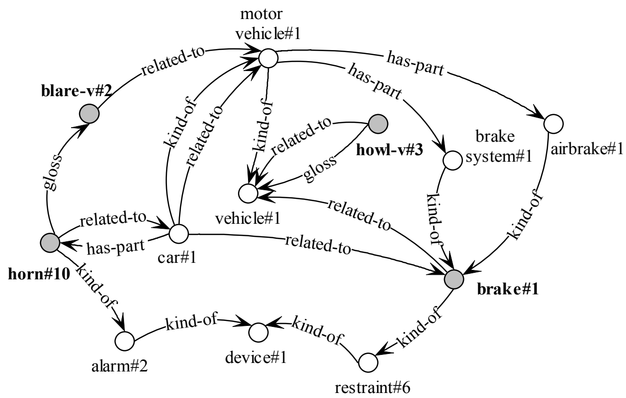

Figure 4. Semantic interconnections between word senses in a conceptually rich sentence.

Clearly, not all the sentences convey enough semantic richness for SSI to find meaningful interconnections. For example, consider the following sentence, again from Senseval-3: “The stranger was thanking Haney profusely and had one arm around his shoulders as if he were an old friend”. Here SSI is unable to find semantic interconnections between *stranger* and *thank*, or between *old* and *friend*, while it succeeds in relating the physical senses of both *arm* and *shoulder*, supported by the patterns in Figure 5.

3.2. Supporting Glossary Construction

When a community decides to construct a glossary, domain experts must face the problem of providing textual definitions for new terms that are not included in a general-purpose lexicon. As the experts are not necessarily

[2] http://www.senseval.org

lexicographers, they need to be helped in the complex task of defining multiword expressions. Furthermore, individual contributors can produce different definitions for the same term, thus requiring a procedure for establishing a consensus between the proposals (Navigli and Velardi, 2005).

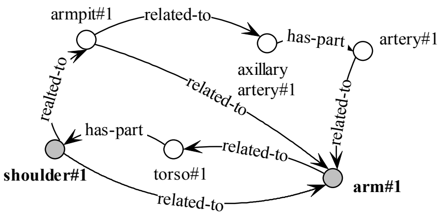

Figure 5. Semantic patterns connecting *shoulder#1* to *arm#1*.

In our previous work on automated ontology learning (Navigli and Velardi 2004; Cucchiarelli et al. 2004) and in a recent glossary construction experiment (Navigli and Velardi, 2005), SSI was used to detect semantic relations between the component words of complex multiword expressions (MWEs). In the fully automated version of the system, this information is used only to establish taxonomic relations among terms (e.g. the *kind-of* relation between *rule-based system* and *inference engine*).

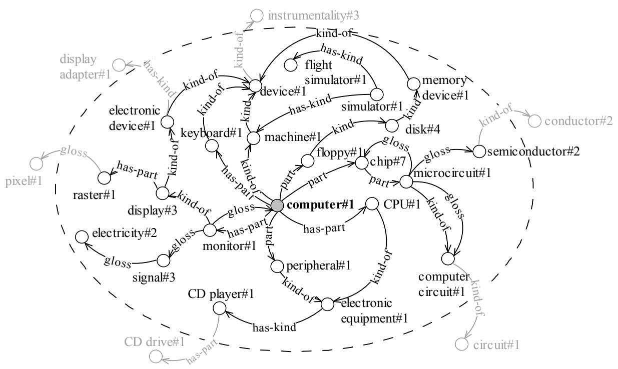

Figure 6. The semantic network for *computer#1*.

However, as shown by the semantic graph of Figure 6, the information provided is much richer. A human volunteer may select certain patterns and use them to produce a formal definition of the concept representing the multiword expression. For example, the definition "a *computer keyboard* is a *keyboard device* for *computers*" can be obtained by traversing the patterns connecting *computer#1* with *keyboard#1* in the semantic network, as shown in Figure 7.

More complex compositional definitions for multiword expressions can be created by volunteers with the support of knowledge patterns highlighted by SSI. As an example, consider the term *food ingredient industry*. The structural interconnections presented to the user (shown in Figure 8) chiefly concern the strict link between *food#2* and *ingredient#3* ("ingredient for (making) food"), with the addition of a relatedness relation connecting *industry#1* to *food#2* ("the industry of food").

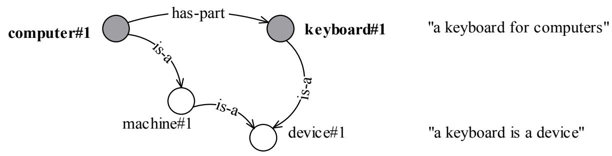

Figure 7. A compositional definition of a multiword expression supported by knowledge patterns.

As a result, the volunteer can easily infer a definition for the multiword expression by combining the two distinct implicit meanings: "the industry of ingredients for making food". Furthermore, most words in the definition being created can be semantically annotated with the appropriate senses in the graph ("the *industry#1* of *ingredients#3* for making *food#2*), thus providing a useful connection with the WordNet sense inventory. Finally, knowledge patterns supporting the definition can be selected and included in the glossary being developed as a formal counterpart of the textual gloss.

Unfortunately, not all MWEs can be defined in a decomposable way (i.e. by analysing its components). MWEs can also be *non decomposable* (e.g. *room night*) or *idiosyncratically decomposable* (i.e. some of their parts may assume non-standard senses, e.g. *queen room*).

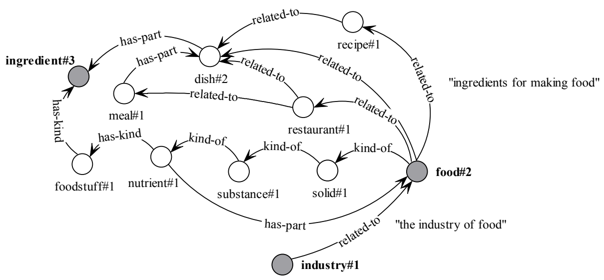

Figure 8. A compositional definition of *food ingredient industry*.

Therefore, for the collection of term definitions, the human volunteer can be driven by SSI only for the subset of decomposable terms, in most cases those for which the algorithm produces a meaningful output.

3.3. Supporting the Enrichment of Lexical Knowledge Bases

Large-scale knowledge bases are crucial for knowledge-intensive AI tasks. Building a broad-coverage resource from scratch is a hard and time-consuming task, requiring expertise and coordination. A viable solution is the extension of existing knowledge repositories (e.g. WordNet). In a recent, ongoing work (Navigli, 2005) we propose a semi-automatic methodology for the enrichment of WordNet with relatedness relations between concepts, with the result of (at least partially) overcoming the lack of semantic relations in the Princeton resource, not considering *hypernymy* (kind-of) and *meronymy* (has-part). The extension is being performed through the use of automatic techniques and manual integration of lexical resources (e.g. the Longman Language Activator, the Oxford Collocations, collocation web sites, etc.).

For such an extension to be reliable, human validation is required. This is a case where, again, the contribution of distributed volunteers can be productively employed. Due to the large quantity of knowledge to be collected and validated, the process is indeed iterative and requires the progressive intervention of human annotators to validate the newly acquired information.

Consider the following example: we collected a number of (mostly ambiguous) collocations or relations involving the word *meaning*: *message*, *communication*, *sense*, *modifier*, *word*, *sentence*, *concept*. Figure 9 illustrates the knowledge patterns retrieved by SSI and presented to the human volunteer to the end of validating the senses chosen for the collocated words.

After validation, we can add a relatedness relation between *meaning#1* and each of the word senses chosen for the collocated words. As a result, the enriched lexical knowledge base can be used in future iterations as a support for the validation of new lexical knowledge.

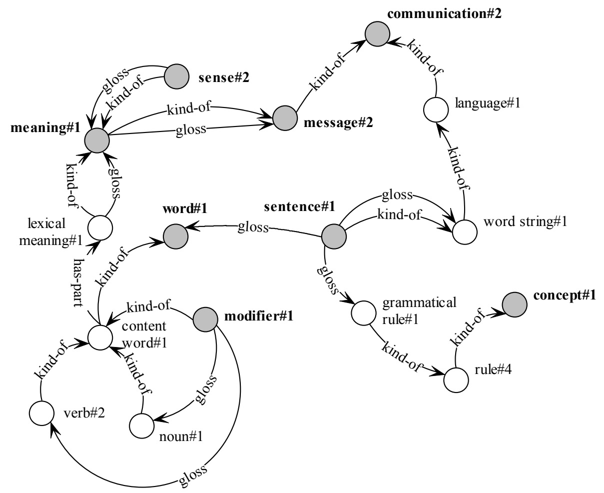

Figure 9. Interconnections between senses of words collocated with *meaning*.

As a second example, refer to the graph in Figure 2 showing the semantic interconnections between senses of words collocated with *taxi#1*. Here again a number of interconnections guide the volunteer in choosing the most appropriate word senses, so that the lexical knowledge base can be enriched with new instances of conceptual relations.

As a final remark notice that, although the enrichment and validation methodology was applied to the SSI lexical knowledge base (including WordNet and other resources), it is also applicable to similar knowledge collection efforts that do not discriminate between word senses (Ravichandran and Hovy, 2002; Chkvloski and Pantel, 2004; Etzioni et al., 2004).

3.4. Detecting Inconsistencies

Consistency and subjectivity are typical problems when manually building semantic resources, even when the task is performed by professional lexicographers, provided with clear guidelines, and directed by a supervising team (e.g. the Princeton team during the development of WordNet (Fellbaum, 1998)).

We may certainly expect that these problems will be emphasised if the work is to be conducted by a community of volunteers acting in a distributed environment (e.g. the Internet).

SSI can be used to detect knowledge patterns as indicators of inconsistencies between the intuition of the builders of a knowledge repository (e.g. WordNet) and the human volunteers aiming at annotating text with respect to the reference resource. The patterns can also be used to point at emerging discordances between annotators.

In the case of generic semantic annotation, the system can identify complex patterns of knowledge occurring many times and suggest the standard choice when the human user inconsistently opts for a different one. For instance, suppose that the phrase "a bunch of flowers" recurs many times in the corpus to be semantically annotated, with some variations, like "a bunch of odorous flowers", "a bunch of scented, star-shaped flowers", etc. Volunteers could certainly choose different senses for each word composing the phrase, but such different intuitions (sometimes wrong) will be smoothed by the SSI knowledge patterns suggested to the users (Figure 10).

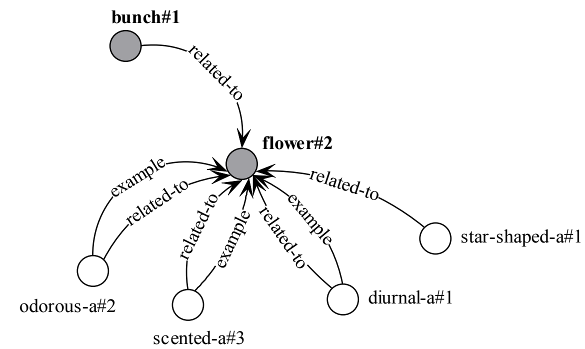

Figure 10. Basic knowledge patterns for the phrase "a bunch of ADJ* flowers".

Clearly, a contributor is free to persist in his/her choice if he/she does not repute the suggestion valid enough, but this guarantees that – in presence of strong knowledge patterns – user choices on the same instance will be much more consistent than in a typical, loosely supervised distributed environment.

Figure 11 reports a second example of a knowledge pattern that can be used as a reference for the detection of tagging inconsistencies.

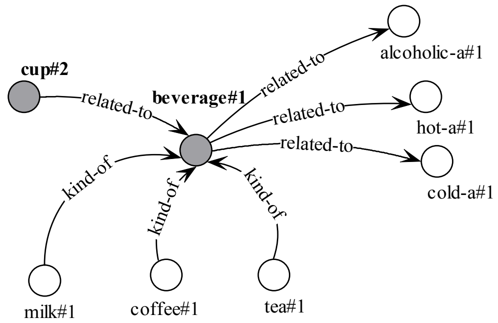

Figure 11. Knowledge patterns involving the concept of *cup#2*.

The pattern involves the concept of *cup*, intended as "the quantity a cup will hold", and identifies beverages as the entities that usually constitute such a quantity. The pattern can drive the human volunteer in tagging phrases like "a cup#2 of tea#1", "a cup#2 of coffee#1", "a cup#2 of milk#1", as *tea*, *coffee* and *milk* are all kinds of *beverages*.

In the more specific case of the semantic annotation of dictionary glosses, where the contributor is asked to assign a sense to each word in a dictionary definition (*gloss*), the structure of the reference knowledge repository often suggests a preferred choice. During the Senseval-3 gloss WSD competition (Litkowski, 2004), knowledge patterns detected by the SSI algorithm were indeed used as indicators of inconsistencies between the intuition of WordNet lexicographers and that of the annotators who provided the set of "golden" glosses used to test the various systems (Navigli and Velardi, 2004b).

As an example, consider the WordNet definition of the fourth sense of *custom* as "habitual patronage". A *hypernym* (*kind-of*) pattern connects *custom#4* with *patronage#5* ("the business given to a commercial establishment by its customers"), while the annotators' choice was the first sense of *patronage* ("accepted or habitual practice"). This divergence does not necessarily mean that the human annotator has chosen the wrong sense, but is a clear indicator of inconsistency raised by SSI.

As a second example, consider the definition of *motorcycle#1*, "a motor vehicle with two wheels and a strong frame". This definition clearly states that a *motorcycle#1* is-a *motor vehicle*, has-part *wheel* and has-part *frame*. All these information are made explicit by specific conceptual relations encoded in the SSI lexical knowledge base (mostly, from WordNet), as shown in Figure 12.

The choice of the first sense of *frame*, as done by the annotators of the Gloss WSD task at Senseval-3, is inconsistent with the resource structure, and such an inconsistency can be raised by SSI during manual gloss annotation.

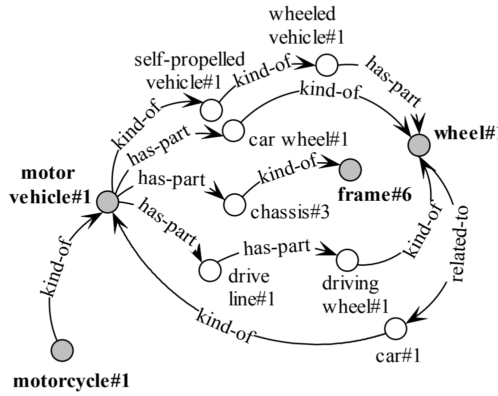

Figure 12. Semantic interconnections between word senses in the dictionary definition of *motorcycle#1*.

3.5 Detecting Multiple Interpretations

As a final application of the SSI semantic graph justifications, in this section we discuss the challenging possibility, to be further developed, of detecting multiple interpretation levels of words in a context. We start with an example from the Senseval-3 English all-words disambiguation task: "Eyes that were clear, but also bright with a strange intensity, a sort of cold fire burning behind them". Here, the interesting point raised by SSI is the semantic connection between *fire* and *burn*. Both the literal interpretation (as if a fire were burning) and the metaphorical one (feeling ardour and fervour) are acceptable, depending upon the level of interpretation that we desire to focus on. SSI detects both kinds of interconnections, because the two senses of *fire* (#1 and #6) are respectively coupled with those of *burn* (again #1 and #6) in the SSI lexical knowledge base, and each sense choice gets the same weight so that it is arbitrary whether to choose the literal or the metaphorical interpretation.

For the same reasons, "a shining sun" or the WordNet example "the only cloud in the horizon was the possibility of dissent by the French" can be interpreted on two different levels. Usually, in these cases, the shift of meaning due to the use of a metaphor brings us to prefer the extended meaning (*fire* in the sense of *strong feeling*, *sun* in the sense of *important person*, *cloud* in the sense of *worry*, etc.), but the human volunteer may also prefer the literal sense (or vice-versa), e.g. to the end of adding both interpretations to the knowledge repository.

4. Building the Interface

We performed an internal assessment of the effectiveness of SSI as a support for sense selection, but a real evaluation will be feasible only after the release of *Valido*, a visual tool implementing most of the features described in Section 3. Nevertheless, a practical (and unexpected) assessment of the methodology occurred in the context of the Gloss Word

Sense Disambiguation task at Senseval-3 (Litowski, 2004), where we submitted the inconsistent sense choices detected by SSI to the organizers, as previously detailed.

The development of *Valido* is ongoing. The tool enables the user to visualize semantic interconnections between the senses of words in a context (depending upon the task). Given a task, *Valido* actively chooses which contexts to be submitted to the volunteer based on the quantity and heterogeneity of previous contributions, starting from instances analysed by a lower number of users.

In the task of semantic annotation (Section 3.1), the human volunteer is provided with the senses suggested by SSI for some words in the sentence and is asked to accept (i.e. validate) each of the automatic choices or to make a change. As a result of a modification in a sense choice, the user can view an updated graph connecting the new chosen sense with the other senses in the context. For instance, consider the sentence "The driver stopped swearing at them, turned on his heel and went back to his truck". In this case, SSI detects valid knowledge patterns from the golf domain between *driver* and *heel* that are unfortunately not applicable in this context (Figure 13(a)). When the user changes the senses of both words, he/she gets patterns that are much more convincing with respect to the context at hand (Figure 13(b)).

Notice that SSI can be used as a support for semantic annotation of either a single word (like in the OMWE) or all the words in a sentence (typical of an all-words tagging task).

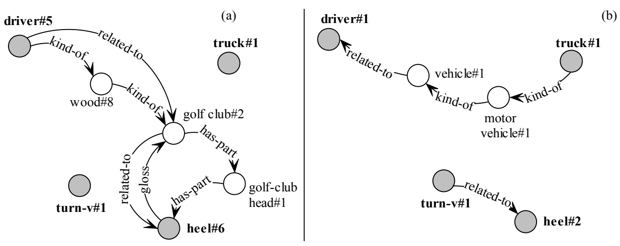

Figure 13. Connections selected by SSI for a sentence (a) and after the user adjustment to the context (b).

In the task of enrichment of knowledge bases (Section 3.3), the context is the set of words collocated with a selected term and the user is presented with a similar scenario.

In the case of glossary construction (Section 3.2), the tool allows the user to associate partial definitions with knowledge patterns and to type a complete definition for the multiword expression at hand starting from the textual fragments assigned to single semantic interconnections.

During each task, the volunteer is assisted by the tool in the detection of inconsistencies. When the sense selection contrasts with the majority of the choices made by other users on the same instance or concerning an established knowledge pattern (as described in section 3.4), the inconsistency is notified to the volunteer who is nevertheless free to persist in a divergent decision.

Finally, we plan to include the notification of possible multiple interpretation levels (as illustrated in section 3.5) in a non intrusive manner.

5. Conclusions

In this paper, we discussed different usages of the Structural Semantic Interconnection algorithm as a support tool to collect semantic knowledge from human volunteers in a distributed environment.

In each task the user is presented with suggestions produced by SSI in the form of semantic graphs, providing guidance for the choice to be made. This contrasts with most of the knowledge collection systems, which do not suggest a preferred option to the volunteer.

Our approach requires non-tagged knowledge to be previously mined in the form of corpora to be linguistically annotated, terminology to be provided with textual definitions, collocations or ambiguous relations for the enrichment of existing knowledge bases.

For its very nature, our approach works on the semantic side of the knowledge acquisition process, while many of the works in the literature do not discriminate between word senses in collecting knowledge. This opens up the possibility to combine non-semantic knowledge collection with our semantic approach, in order to provide the volunteer with a battery of tools for the large-scale acquisition of explicit knowledge.

Finally, we plan to perform a large-scale experiment in order to assess the usability and effectiveness in a distributed environment of our interface based on structural semantic interconnections.

Acknowledgements

The author would like to thank Prof. Paola Velardi for her support and useful discussions. This work is partially funded by the Interop Network of Excellence (508011), 6^{th} European Union FP.

References

Bunke, H. and Sanfeliu, A. eds. 1990. *Syntactic and Structural pattern Recognition: Theory and Applications*. World Scientific, Series in Computer Science, vol. 7.

Chklovski, T. 2005. Designing Interfaces for Guided Collection of Knowledge about Everyday Objects from Volunteers. In *Proceedings of Conference on Intelligent User Interfaces*, San Diego, CA.

Chklovski, T. and Mihalcea, R. 2002. Building a Sense Tagged Corpus with Open Mind Word Expert. In *Proceedings of the ACL 2002 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, PA.

Chklovski, T. and Pantel, P. 2004. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In

Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-04), Barcelona, Spain.

Chklovski, T., Ratnakar, V., and Gil, Y. 2005. User Interfaces with Semi-Formal Representations: a Study of Designing Argumentation Structures In *Proceedings of Conference on Intelligent User Interfaces*, San Diego, CA.

Conklin, J. and Begeman, M. 1988. gIBIS: A hypertext tool for exploratory policy discussion. *ACM Transactions on Office Information Systems*, 6(4):303-331.

Cucchiarelli, A., Navigli, R., Neri, F., and Velardi, P. 2004. Automatic Ontology Learning: Supporting a Per-Concept Evaluation by Domain Experts. In *Proc. of Workshop on Ontology Learning and Population* (*OLP*), in the *16th European Conference on Artificial Intelligence (ECAI)*, Valencia, Spain.

Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D., and Yates, A. 2004. Web-scale Information Extraction in KnowItAll. In *Proceedings of the World Wide Web Conference*, New York, NY.

Fellbaum, C. ed. 1998. *WordNet: an Electronic Lexical Database*. MIT Press.

Hirst, G. and St-Onge, D. 1998. Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. In Fellbaum, C. ed. *WordNet: An Electronic Lexical Database and some of its Applications*, Cambridge, MA: The MIT Press.

Litkowski, K. C. 2004. SENSEVAL-3 Task: Word-Sense Disambiguation of WordNet Glosses. In *Proc. of SENSEVAL-3 Workshop on Sense Evaluation*, in the *42th Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcelona, Spain.

Lowrance, J., Harrison, I., and Rodriguez, A. Capturing Analytic Thought. In *Proceedings of the 1st International Conference on Knowledge Capture* (K-CAP 2001), 2001.

Mihalcea, R. and Moldovan, D. 2001. eXtended WordNet: Progress Report. In *Proceedings of Workshop on WordNet and other lexical resources*, Pittsburgh, PA.

Miller, G., Leacock, C., Randee, T., and Bunker, R. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, Plainsboro, New Jersey.

Navigli, R. 2005. Semi-Automatic Extension of Large-Scale Linguistic Knowledge Bases. To appear in *Proc. of 18th International FLAIRS (FLorida Artificial Intelligence Research Society) Conference*, Clearwater Beach, Florida.

Navigli, R. and Velardi, P. 2005. Automatic Acquisition of a Thesaurus of Interoperability Terms. To appear in *Proc. of 16th IFAC World Congress*, Praha, Czech Republic.

Navigli, R. and Velardi, P. 2004. Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites. *Computational Linguistics*, MIT press, (30)2:151-179.

Navigli, R. and Velardi, P. 2004b. Structural Semantic Interconnection: a Knowledge-Based Approach to Word Sense Disambiguation. In *Proceedings of SENSEVAL-3 Workshop on Sense Evaluation*, in the *42th Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, 179-182. Barcelona, Spain.

Ng, H. T., and Lee, H. B. 1996. Integrating multiple knowledge sources to disambiguate word sense: An examplar-based approach. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, CA.

Ravichandran, D. and Hovy, E. 2002. Learning Surface Text Patterns for a Question Answering System. In *Proceedings of the 40th Association for Computational Linguistics*, Philadelphia, PA.

Singh, P. 2002. The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Palo Alto, CA.

Suthers, D. 2004. Representational Guidance for Collaborative Inquiry. In Andiressen, J., Baker, M., Suthers, D. eds. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Springer.