From Glosses to Qualia: Qualia Extraction from Senso Comune

Tommaso Caselli
Trento Rise / Via Sommarive 18
IT-38122 Povo
t.caselli@trentorise.eu

Irene Russo
ILC-CNR / Via G. Moruzzi, 1
IT-56124 Pisa
irene.russo@ilc.cnr.it

Abstract

This paper describes a case study on methods for automatically extracting qualia relations from dictionary glosses in Italian, namely the Senso Comune De Mauro Dictionary (SCDM). The qualia extraction has been addressed by means of a pattern-based approach and lexical match with an Italian generative lexicon based language resource, PAROLE-SIMPLE-CLIPS (PSC). The evaluation of the extraction approaches has been performed with respect to a manually built Gold Standard containig 174 different qualia. The results obtained are encouraging (P = 0.84, R = 0.08 for the pattern extraction approach and P=0.73 and R=0.16 for the merging of pattern extraction and lexical match) and suggest that the information contained in the SCDM glosses is complementary with that in PSC.

1 Introduction

This paper describes a case study on methods for automatically extracting qualia relations (Pustejovsky, 1995) information from lexicographic dictionary glosses in Italian, namely the Senso Comune De Mauro (SCDM henceforth) Dictionary[1] for a specific semantic class, i.e. the ARTIFACT class in the Senso Comune ontology.

Qualia structure is a distinctive feature of the Generative Lexicon (GL) theory (Pustejovsky, 1995). It is a simple and powerful structure which contribute to the representation of the meaning of the nouns. The Qualia Structure consists of four roles:

- Formal role: the conceptual super category from which the object inherits its properties;

- Agentive role: the origin of the object, its coming into being into the world;

- Telic role: the purpose, or typical function of an object;

- Constitutive role: the internal constituents (parts, material, weight etc.) of an object.

The actual realization of each role is also dependent on the associated semantic type (or ontological class(es)) of the entity analyzed. For instance, an entity denoting a Natural Object (e.g. *tree*, *flower*, *fruit*, etc.) will never have information for the Agentive role. On the contrary, this information is relevant for Artifacts (*wheel*, *pen*, *table*, etc.).

The qualia extraction task has been mainly addressed in NLP by means of pattern-based approaches on corpora and from dictionaries. Pattern-based approaches for extracting semantic relations are well known in literature (Calzolari (1991); Montemagni and Vanderwende (1992); Hearst (1992); Bouillon et al. (2002); Cimiano and Wenderoth (2005); Pantel and Pennacchiotti (2006), among others) and have proved highly reliable, namely in terms of precision, for extracting the different types of qualia. One the advantages of our work is that the extracted qualia are associated with both a word sense and an ontological class (see Section 3 for details on the SCDM Dictionary). Furthermore, the SCDM dictionary glosses are richer and more descriptive than the WordNet glosses. The data collected can be exploited in different ways, namely:

- to reduce the complexity of the lexicographic entries, thus facilitating dictionary entry merging and sense alignment with other lexica like, for instance, WordNet (even in languages other than Italian);

---

[1]www.sensocomune.it

- to enrich already existing lexica such as PAROLE-SIMPLE-CLIPS (PSC henceforth) (Ruimy et al., 2003);

- to improve the performance of Natural Language Processing tools for complex tasks involving encyclopedic knowledge such as Question Answering and Textual Entailment, among others.

The remainder of this paper is structured as follows: Section 2 will briefly describe related works on the automatic extraction of qualia information. In Section 3 we will highlight the characteristics of the two lexica, namely the SCDM Dictionary and the PSC lexicon. A detailed description of the methodology used to identify the linguistic patterns coding the qualia information and their evaluation with respect to a manually built gold standard is reported in Section 4. In addition to this, we have also carried out experiments i.) to exploit the qualia information in the PSC lexicon to identify additional qualia which were not extracted by means of the patterns; and ii.) to evaluate the coverage of the extracted qualia with respect to the entries in the PSC lexicon in order to enrich it. Finally, Section 5 reports on conclusions and highlights on-going and future research directions.

2   Related Works

In recent years there has been a continuous interest in the NLP community on discovering novel instances of semantic relations. Most of this earlier work was based on surface pattern matching (Hearst (1998); Cimiano and Wenderoth (2005); Yamada and Baldwin (2004) among others). Other works start from matches extracted with this method and then use supervised training data to learn semantic constraints to improve precision (Girju et al. (2003); Katrenko and Adriaans (2010)). Much of previous works concentrated on extracting hypernyms (Snow et al. (2005); Sang and Hofmann (2009). Other works have applied pattern classification approaches to extract larger set of relations. The results obtained proved that extracting a pattern distribution between occurrences and performing supervised classification based on this distribution is a viable and promising solution for extending the range of semantic relations beyond hyperonymy (Ó Séaghdha and Copstake (2007); Herdağdelen and Baroni (2009)). With respect to previous research and similarly

to what was done in the ACQUILEX Project, we tackle the task of extracting qualia relation from dictionary glosses. However, the SCDM glosses are augmented with ontological information whereby each sense is associated with a top level ontology class. This allows us to have at disposal qualia associated with specific senses and ontological classes.

3   The Senso Comune Lexicon and PAROLE-SIMPLE-CLIPS

The SCDM lexicon is part of a larger research initiative, *Senso Comune* (Vetere et al. (2011) Oltramari et al. (2013)). Senso Comune aims at building an open knowledge base for the Italian language, designed as a crowd-sourced initiative that stands on the solid ground of an ontological formalization and well-established lexical resources. The lexicon entries have been obtained from the De Mauro GRADIT (DMG) dictionary and consists in the 2,071 most frequent Italian words. In SCDM, word senses are encoded following lexicographic principles and are associated with lexicographic examples of usage. Senso Comune comprises three modules: i.) a top level module for basic ontological concepts; ii.) a lexical module for linguistic and lexicographic structures; and iii.) a frame module for modeling the predicative structure of verbs and nouns. The top level ontology is inspired by DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) (Masolo et al., 2002), which has been developed in order to address core cognitive and linguistic features of common sense knowledge. 4,586 word senses from De Mauro Dictionary, corresponding to 1,111 fundamental noun lemmas and covering about 80% of the occurrences in texts (Oltramari et al., 2013), have been manually classified according to the ontological concepts.
PSC (Ruimy et al., 2003) is an Italian syntactic-semantic lexicon based on the GL theory. Lexical units are structured in terms of a semantic type system and are characterized by means of a rich set of semantic features and relations. The type system consists of 157 language- and domain-independent semantic types designed for the multilingual lexical encoding of concrete and abstract entities, events and properties. The type system of the resource reflects the GL assumption that lexical items are multidimensional entities. Multidimensionality is encoded by means of

the Extended Qualia Structure, a revisited version of the GL representational tool which extended each of the qualia roles with subtypes (e.g "*Concerns*" is a subtype of the Constitutive qualia). The PSC lexicon has been connected with ItalWordNet (Ruimy et al., 2008), an Italian version of WordNet based on the EuroWordNet principles, and contains 31619 nominal lemmas, for a total of 38092 senses, 38153 associated semantic type (ontological category) and 65539 qualia.

Although the structure of the two lexica is different, there are some common aspects (e.g. the ontological classes associated with word senses) which suggest both the possibility of merging them and respectively enriching their entries with the encoded information. Finally, the GL theory is based on the inclusion of basic encyclopedic information about nouns to model compositionality, and lexicographic glosses offer this kind of knowledge.

4 Experiments

In order to identify reliable patterns expressing qualia relations on the basis of the glosses in the SCDM lexicon, we developed a specific dataset. We first restricted the exploration of the SCDM entry to nouns which have been assigned the ontological class ARTIFACT in the Senso Comune ontology. We then extracted 35 lemmas with a total of 97 different senses as a development set. We manually explored both glosses and lexicographic examples and identified a set of 48 different syntagmatic patterns expressing the four qualia roles in a unique way. In particular, we identified 23 patterns for the telic role, 13 for the constitutive role, 5 for the formal role and 7 for the agentive role. In Table 1 we report some pattern examples and their associated qualia. In the templates N, V and ADJ refer to the target noun, verb and adjective expressing the qualia, respectively and "det" refers to the presence of articles (partitive, definite and indefinite ones). The item expressing the qualia is in bold in the pattern template and in the example.

The possibility of restricting the qualia extraction to word senses with explicit ontological classification is an advantage of using the SCDM lexicon, as this allows to disambiguate inherently ambiguous patterns. For instance, the pattern "*prodotto da (det) N*" [produced by (det) N] can express both the Constitutive quale, if it applies to the class of Natural Object such as fruit names, or the Agen-

tive quale, if it applies to the class of Artifact such as man-made objects.

In order to evaluate the quality of the qualia extracted by the identified set of patterns in terms of coverage and to identify limitations of this methodology, such as the presence of qualia which cannot be collected by means of pattern templates and additional missing patterns, we developed a manually annotated gold standard. We selected 46 nominal entries in the SCDM lexicon with at least one sense associated with the ontological type ARTIFACT. This has provided us with a set of 50 different senses and a total of 173 different qualia, namely 79 for the constitutive role; 3 for the agentive role; 46 for the telic role; and 45 for the formal role. None of the entries in the Gold Standard is part of the development set described above. We automatically analyzed part-of-speech and lemmas in the glosses by means of the TextPro tool suite (Pianta et al., 2008), applied the pattern extraction script and then evaluated with respect to the Gold Standard. The results are reported in Table 2; all measures have been computed in terms of Precision (P), Recall (R) and F-measure (F1). We evaluated the reliability of the patterns both globally (Overall Evaluation) and for each qualia.

Evaluation Type	P	R	F1
Overall Evaluation	**0.84**	**0.08**	**0.14**
Agentive	1	0.5	0.66
Formal	1	0.01	0.02
Telic	0.92	0.16	0.27
Constitutive	0.73	0.07	0.12

Table 2: Evaluation of the patterns with respect to the Gold Standard.

The results obtained are quite satisfactory. Precision is extremely high but this has a cost in terms of recall, both for the overall evaluation and for each single qualia. A detailed error analysis (namely false positives and false negatives) has shown: i.) that some (additional) patterns were missing, thus preventing the extraction of qualia fillers which have been manually identified, namely for the constitutive qualia. This also explains the very low level of recall for the constitutive quale; and ii.) that some qualia are expressed in the glosses by using general expressions or as arguments of specific verbs which cannot be codified into general pattern structures at the moment. The recall (and f-measure) for the formal

Template Pattern	Example	Qualia
usare per (fare—mettere) **V** [used to (make—put) V]	usato per **cacciare** [used for hunting]	telic
costituito da (det adj—det) **N** [made of N]	costituito da **metallo** [made of metal]	constitutive
di colore **ADJ** [of ADJ color]	di colore **grigio** [of grey colour]	constitutive
prodotto da [det] **N** [produced by N]	prodotto dalla **lavorazione**	agentive
un tipo di **N** [a kind of N]	un tipo di **strumento** [a kind of instrument]	formal

Table 1: Qualia and patterns extracted from the development noun set.

qualia is the lowest. This is due to the fact that the SCDM dictionary very rarely uses explicit definitional patterns for indicating the supertype (e.g. "*è un N*" [(it) is a N]) but tends to directly use the hypernym item (e.g. as the first noun in the gloss). However, not all glosses exploit hypernyms for the sense descriptions, they sometimes contain synonyms. We thus reduced the identification of formal qualia only to instances expressed in well formed patterns during the pattern template development phase. Although the quantity of correct qualia is not large (we extracted 32 qualia relations from our data), their reliability and quality is extremely high. In the following section, we will describe i.) the methodology we adopted in order to extend the extraction of the qualia from the SCDM glosses by exploiting the information encoded in the PSC lexicon; and ii.) an evaluation of the coverage of the extracted qualia with respect to those in the PSC lexicon to enrich it with them.

4.1 Extending Qualia Matches with PSC

In order to extend the qualia extracted from the SCDM lexicon, we decided to exploit the information encoded in the PSC lexicon. Although the two lexica have different structures, they have common aspects, as described in Section 3. The ontological models, which inform the semantic typing of the word senses and contribute to keep distinct the linguistic and the conceptual levels of representation, can be exploited in order to start merging the two resources. The working hypothesis is that the ontologies of the two lexica can be merged together by means of equivalence relations and subsumption. This will allow us to have sets of ontologically compatible entries which can be further aligned for word senses (word sense alignment; WSA) by means of different methods, such as lexical match on the glosses (Niemann and Gurevych (2011); Meyer and Gurevych (2011)), exploita-

tion of qualia information and graph-based approaches (Matuschek and Gurevych (2013); Navigli and Ponzetto (2012)).

At this stage of development, we partially tackled the task of aligning the ontological models of the two lexica by restricting our analysis of the PSC semantic types to those which are compatible with or equivalent to the SCDM semantic type ARTIFACT, namely *Instrument* and *Artifact*. Notice that we excluded the PSC types *Artifactual_food* and *Artifactual_drink* which in SCDM are assigned to the type SUBSTANCE.

We then extracted all qualia information for the 46 lemmas in the Gold Standard which have a corresponding lemma entry and ontological class *Artifact* or *Instrument* in the PSC lexicon. In this way we obtained 333 couples lemma - qualia. We then applied a baseline method for extracting additional qualia from the glosses based on token match. For each matched lemma in the PSC lexicon we grouped all its associated qualia and looked for an exact match in the gloss tokens of the corresponding lemmas in the SCDM lexicon. To avoid repetitions and to get also a preliminary evaluation both of the coverage of the PSC qualia and of the richness of the SCDM glosses in terms of qualia, we excluded from the SCDM glosses all qualia which had been extracted by means of the patterns. We then merged together the data obtained from the PSC lexicon with those obtained from the pattern extraction and evaluated this new data set against the manually built Gold Standard (DirectMatch_Extracted). Additionally, we also evaluated the data obtained from the direct match only against the Gold (DirectMatch_only) so to get a preliminary estimate of the coverage of the PSC qualia. The figures obtained are reported in Table 3.

By comparing the results of the DirectMatch_Extracted with those obtained from the pat-

Evaluation Type	P	R	F1
DirectMatch_Extracted	**0.73**	**0.16**	**0.26**
DirectMatch_only	0.60	0.06	0.12

Table 3: Evaluation of extraction by direct match of the PSC qualia in the SCDM glosses.

tern extraction method only (see Table 2), we can notice that the decrease in precision (0.84 *vs*. 0.73) is balanced by an increase in recall (0.08 *vs*. 0.16). Furthermore, by observing the figures of the results of DirectMatch_only, the lower level of precision (i.e. high number of false positives; 0.60 *vs*.0.84) suggest that the qualia information in PSC, for the great part, contains information which is additional and complementary with respect to the qualia which can be extracted from and are actually contained in the SCDM glosses, namely for the constitutive and formal roles.

To identify further support to this observation, we conducted three further evaluation assessments. We computed precision and recall between: i.) the manual gold standard and the PSC qualia (Gold-PSC), with the PSC data as the key and the manual gold standard as the response; ii.) the results of the pattern extraction and the PSC data (Extracted-PSC), with the PSC data as the key and the pattern extracted data as the response; iii.) the false positives items from the Extracted-PSC evaluation and the manual gold standard (FP_PSC-Extracted), with the manual gold as the key and the false positives as the response. We report in Table 4 the results obtained.

Evaluation Type	P	R
Gold-PSC	0.16	0.04
Extracted-PSC	0.12	0.006
FP_PSC-Extracted	0.85	0.07

Table 4: PSC qualia, manual Gold qualia and extracted qualia coverage.

It is interesting to notice that both for the Gold-PSC and the Extracted-PSC data the recall and precision values are extremely low. On the other hand, the results for the PF_PSC-Extracted are in line, both for precision and recall, with those obtained for the pattern extraction against the Gold. On the basis of these figures we claim that: i.) the pattern-based extraction method has some issues in coverage but it provides highly reliable data;

ii.) the data in PSC are complementary to those which can be extracted from the SCDM glosses and that, in the perspective of merging the two resources, we can obtain a richer lexicon with a better coverage of qualia information. However, the process of integration of the SCDM extracted qualia into the PSC is not straightforward as qualia in PSC are assigned with a semantic type and a specific semantic unit (i.e. a sense). So far, we have provided a partial enrichment of the PSC entries with ontological type *Instrument* and *Artifact* for the type of relation (telic, agentive, formal or constitutive) and lemma(s) of the relation filler(s). For instance, the PSC entry for "*arco*" [bow], type *Instrument*, has the following qualia information: formal "*strumento*" [instrument]; agentive "*fabbricare*" [to make]; telic *tirare*" [to throw]. We have integrated the entry with the following additional qualia obtained from the gloss: telic "*caccia*" [hunt]; constitutive "*asta*" [rod], "*flessibile*" [flexible]. With respect to these latter aspects, we are currently experimenting with similarity measures between words and planning a crowd-sourcing task for word sense disambiguation (WSD) through the Senso Comune platform.

5 Conclusions and Future Work

This paper has reported an on-going research on the extraction of semantic information from dictionary glosses and merging of lexica at the semantic level. We have experimented a pattern-based method for qualia extraction from ontologically annotated dictionary entries (SCDM glosses) which has allowed us, on the one hand, to add high quality semantic information to the Senso Comune repository and, on the other hand, has provided a set of data for the automatic enrichment of an existing lexicon containing qualia information, namely PSC. At the current stage of development we are facing an issue related to the recall, i.e. a quantity issue rather than a quality issue. The solution can be only in part addressed by adding missing patterns, as most qualia concerning the constitutive and the formal roles are not always expressed in the SCDM glosses by means of collocation patterns. As a way to boost the extraction of additional qualia we have explored the possibility of using the information in the PSC lexicon. As a preliminary strategy we have adopted the "direct match" solution, i.e. lexical match of the token in the SCDM gloss. The results obtained are quite

surprising as most of the qualia manually identified in the process of creation of the gold standard are not contained in the PSC entries. This signals that the two lexica contains complementary information and both of them could benefit from their merging. In order to accomplish this, as a preliminary task the two ontological models representing the conceptual backbone of the two lexica must be manually aligned. To improve the recall we are planning: i.) to run a new pattern extraction experiment by exploiting full parsing of the glosses and dependency relations, though parsing errors may not contribute to improve recall (Sang and Hofmann, 2009) and ii.) to exploit similarity measures between the qualia data obtained by the patterns and the lexical items in the the gloss which have not been extracted.

The availability of qualia information associated with word senses can be further exploited for achieving word sense alignment (WSA) among different lexica. As a matter of fact, words sharing the same sense, or meaning, must have a common subset of qualia roles. To prove the validity of our hypothesis, we are currently trying to achieve WSA between MutiWordNet synsets (Pianta et al., 2002) and Senso Comune entries by exploiting qualia extraction from the glosses, hypernyms relations and meronyms.

Finally, an analysis of the manually annotated data has highlighted that, at least for the SCDM lexicon, the constitutive and the telic roles have a primary function in describing the meaning of a nominal instance of type ARTIFACT, more than the formal and the agentive roles. This suggests an interesting working hypothesis for the identification of what could be called "core qualia" in order to express and identify the core lexical semantics characteristics of the entities belonging to different ontological classes. On the basis of the results we have obtained from the manual exploration, it seems that as far as the ARTIFACT type is concerned the core qualia are the constitutive and the telic roles.

References

P. Bouillon, V. Claveau, C. Fabre C, and P. Sébillot. 2002. Acquisition of qualia elements from corpora - evaluation of a symbolic learning method. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Spain.

N. Calzolari. 1991. Acquiring and representing semantic information in a lexical knowledge base. In J. Pustejovsky and S. Bergler, editors, *Lexical Semantics and Knowledge Representation. First SIGLEX ACL Workshop. Lecture Notes in Computer Science*, pages 235–244. Springer Verlag.

P. Cimiano and J. Wenderoth. 2005. Automatically learning qualia structures from the web. In *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, pages 28–37, Ann Arbor, Michigan, June. Association for Computational Linguistics.

R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

R. Girju, D. Moldovan, M. Tatu, and D. Antohe. 2005. On the semantics of noun compounds. *Comput. Speech Lang.*, 19(4):479–496, October.

M. A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics - Volume 2*, COLING '92, pages 539–545, Stroudsburg, PA, USA. Association for Computational Linguistics.

M. Hearst. 1998. Automated discovery of wordnet relations. In Christiane Fellbaum, editor, *WordNet: An Electronic Lexical Database*. MIT Press, Amsterdam.

A. Herdağdelen and M. Baroni. 2009. BagPack: A general framework to represent semantic relations. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 33–40, Athens, Greece, March. Association for Computational Linguistics.

S. Katrenko and P. Adriaans. 2008. Semantic types of some generic relation arguments: Detection and evaluation. In *Proceedings of ACL-08: HLT, Short Papers*, pages 185–188, Columbus, Ohio, June. Association for Computational Linguistics.

S. Katrenko and P. Adriaans. 2010. Finding constraints for semantic relations via clustering. In *Proceedings of CLIN-2010*.

C. Masolo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider. 2002. Wonderweb deliverable D17: the wonderweb library of foundational ontologies. Technical report.

M Matuschek and I. Gurevych. 2013. Dijkstra-wsa: A graph-based approach to word sense alignment. *Transactions of the Association for Computational Linguistics (TACL)*, 2:to appear.

C. Meyer and I. Gurevych. 2011. What psycholinguists know about chemistry: Aligning wiktionary and wordnet for increased domain coverage. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 883–892, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

S. Montemagni and L. Vanderwende. 1992. Structural patterns vs. string patterns for extracting semantic information from dictionaries. In *Proceedings of the International Conference on Computational Linguistics - COLING-92*, pages 546–552, Nantes, France.

R Navigli and S. Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

E. Niemann and I. Gurevych. 2011. The peoples web meets linguistic knowledge: Automatic sense alignment of wikipedia and wordnet. In *Proceedings of the 9th International Conference on Computational Semantics*, pages 205–214, Singapore, January.

D. Ó Séaghdha and A. Copstake. 2007. Cooccurrence contetxs for noun compound interpretation. In *ACL Workshop on A Broader Perspective on Multiword Expressions*. Association for Computational Linguistics.

D. Ó Séaghdha. 2007. Annotating and learning compound noun semantics. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

A. Oltramari, G. Vetere, I. Chiari, E. Jezek, F.M. Zanzotto, M.Nissim, and A. Gangemi. 2013. Senso Comune: A collaborative knowledge resource for italian. In I. Gurevych and J. Kim, editors, *The Peoples Web Meets NLP, Theory and Applications of Natural Language Processing*, pages 45–67. Springer-Verlag, Berlin Heidelberg.

P. Pantel and M. Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*. The Association for Computer Linguistics, July.

E. Pianta, L. Bentivogli, and C. Girardi. 2002. Multiwordnet: developing an aligned multilingual database. In *First International Conference on Global WordNet*, Mysore, India.

E Pianta, C. Girardi, and R. Zonoli. 2008. Textpro tool suite. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-08)*, volume CD-ROM, Marrakech, Morocco. European Language Resources Association (ELRA).

J. Pustejovsky. 1995. *The Generative Lexicon*. MIT Press.

A. Roventini, N. Ruimy, R. Marinelli, U. Marisa, and M. Michele. 2007. Mapping concrete entities from parole-simple-clips to italwordnet: Methodology and results. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague, Czech Republic, June.

N. Ruimy, M. Monachini, E. Gola, N. Calzolari, M.C. Del Fiorentino, M. Ulivieri, and S. Rossi. 2003. A computational semantic lexicon of italian: *SIMPLE*. *Linguistica Computazionale XVIII-XIX, Pisa*, pages 821–64.

N. Ruimy, A. Roventini, R. Marinelli, and M. Ulivieri. 2008. Linking and integrating two electronic lexicon. In *Proceedings of the First International Conference on Global Interoperability for Language Resources - ICGL 2008*, pages 197 – 204, Hong Kong, January.

E. T. K. Sang and K. Hofmann. 2009. Lexical patterns or dependency patterns: which is better for hypernym extraction? In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, CoNLL '09, pages 174–182, Stroudsburg, PA, USA. Association for Computational Linguistics.

R. Snow, D. Jurafsky, and A. Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In *NIPS 17*.

G. Vetere, A. Oltramari, I. Chiari, E. Jezek, L. Vieu, and F.M. Zanzotto. 2011. Senso Comune, an open knowledge base for italian. *JTraitement Automatique des Langues*, 53(3):217–243.

I. Yamada and T. Baldwin. 2004. Automatic discovery of telic and agentive roles from corpus data. In *Proceedings of the 18th Pacific Asia Conference on Language*, pages 115–141, Tokyo, Japan.