

Representation and Inference for Open-Domain QA: Strength and Limits of two Italian Semantic Lexicons

Francesca Bertagna

Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche
Via Moruzzi, 1, 56100 Pisa
Francesca.bertagna@ilc.cnr.it

Abstract

The paper reports on the results of the exploitation of two Italian lexicons (ItalWordNet and SIMPLE-CLIPS) in an Open-Domain Question Answering application for Italian. The intent is to analyse the behavior of the lexicons in application in order to understand what are their limits and points of strength. The final aim of the paper is contributing to the debate about usefulness of computational lexicons in NLP by providing evidence from the point of view of a particular application.

1. Introduction

The paper reports on the results of the exploitation of lexico-semantic language resources (LRs) in a Question Answering (QA) application for Italian. Often, lexicon developers have very different perspectives from application developers. As lexicon developers, our intent is to analyse the behavior of the lexicons in application in order to understand what are their limits and points of strength. The lexicons that constitute the focus of our inquiry are ItalWordNet (Roventini *et al.*, 2003) and SIMPLE-CLIPS (Ruimy *et al.*, 2003). An Italian QA prototype (Bertagna *et al.*, 2005) has been built with the aim to provide an experimental setting for our research, while the collection of questions and documents used in the QA tracks of the CLEF evaluation campaign constitutes the test-bed of the experiment.

The overall system is organized following the classic three-module architecture, i.e. the question analysis, the search engine and the answer extraction modules. In order to better explain the impact of the lexico-semantic feedback on this type of application, we organized the system as a two-layer architecture, consisting in what we called the *baseline* and *enhanced prototypes*. In the baseline prototype, no lexico-semantic information is exploited and its results are considered a baseline of the performance of the system. The enhanced prototype exploits instead the information available in the IWN and SIMPLE-CLIPS.

2. Enhanced Prototype

IWN and SIMPLE-CLIPS databases are used by the enhanced prototype in four fundamental phases:

- i) assessment of keyword relevance,
- ii) determination of the expected answer type,
- iii) query expansion,
- iv) strategies for answer detection.

2.1. Assessment of keyword relevance

In the baseline prototype, a method was adopted to assign a *relevance score* to each keyword of the question, mainly based on the recognition of the PoS. In the enhanced prototype, in order to understand which the most

important keywords in the question are, we evaluated the impact of the exploitation of information of a semantic nature by taking into consideration the *specificity* of nouns. A manual analysis of the CLEF2004 questions shows that is particularly important to understand when the *answer type term*¹ has to be sent to the Search Engine, and this is especially crucial for questions introduced by the pronoun *Quale* (What). If we analyse this type of question in the CLEF2004 test-bed, we see that two tendencies seem to emerge:

i) meta-linguistic ATTs should never be present in the query (*nome* (name), *titolo* (title), *sigla*, *abbreviazione* (abbreviation) etc.).

ii) generic, vague ATTs often do not appear in the answer. Intuitively, terms like *ingrediente* (ingredient), *professione* (profession), *unità di misura* (unit of measurement) etc. can be considered generic terms, because we expect them to categorize a certain number of things and should also be quite high in the hierarchies. The basic idea is that very specific keywords should not be dropped from the query. The enhanced prototype determines the specificity of the keyword by assessing two measures: i) the number of hyponyms of the corresponding concept (we considered a threshold of at least 10 hyponyms), ii) the number of levels in the hyperonymic chain above the concept (threshold: 4 levels).

As regards the first tendency, we can exploit LRs to recognize the “meta-linguistic” ATTs that are categorized: i) under the node {unità_linguistica} (linguistic unit) in IWN and ii) under the Semantic Type METALANGUAGE in SIMPLE-CLIPS.

2.2. The Answer Type Taxonomy

In the determination of the expected answer type the system exploits the so-called Answer Type Taxonomy (ATTax), i.e. the hierarchical structure that organizes the expected answer types. The ATTax can be described as composed by two modules: the first one is constituted solely by clusters of lexical-syntactic patterns typical of specific types of question. These clusters are conceived to map different syntactic realizations onto the same answer

¹ The term preceded by ambiguous question stems that allows the derivation of the expected answer type.

type. Examples of clusters are the patterns *Quanto + essere + alto* (How tall is...) and *Quale + essere + altezza* (What is the height of ..), both mapped to the same Answer Type HEIGHT. This first layer of the ATTax is the only one exploited in the baseline prototype. In the second layer, used within the enhanced prototype, answer type terms contained in the syntactic patterns are linked to the senses of the two lexicons, in this way becoming the roots of the taxonomies that collect senses revealing specific Answer Types. The aim of such a data structure is allowing the system to recognize a higher number of expected answer types, in order to adopt more sophisticated strategies for answer recognition. The following figure shows the part of the taxonomy dedicated to the Answer Type LOCATION:

```

<ANSWER_TYPE VALUE="LOCATION">
  <SYN_PATTERNS>
    <SYN_PATTERN>
      <CHUNK>
        <CC>ADV_C</CC>
        <CHUNK_POTGOV>DOVE#B</CHUNK_POTGOV>
      </CHUNK>
    </SYN_PATTERN>
    <SYN_PATTERN>
      <CHUNK_PREAMODIF="QUALE#A#MS">
        <CC>P_C</CC>
        <CC>N_C</CC>
        <CHUNK_POTGOV>
          <IWN_SYNSETS>
            <WORD_MEANING ID="#10671" PART_OF_SPEECH="N">
              <VARIANTS>
                <LITERAL LEMMA="ambito" SENSE="2"/>
                <LITERAL LEMMA="luogo" SENSE="1"/>
              </VARIANTS>
            </WORD_MEANING>
          </IWN_SYNSETS>
          <SIMPLE_CLIPS_SEMU_TEMPLATE>
            <SEMU ID="USEm2066" naming="luogo"/>
            <SEMU ID="USEm2065" naming="posto"/>
            <TEMPLATE ID="Location"/>
          </SIMPLE_CLIPS_SEMU_TEMPLATE>
        </CHUNK_POTGOV>
      </CHUNK>
    </SYN_PATTERN>
  </SYN_PATTERNS>
  <HAS_SUPERFOCUS VALUE="ENTITY"/>
</ANSWER_TYPE>

```

Fig. 1: The node LOCATION in the Answer Type Taxonomy

In case of questions of type LOCATION, for example, this formalization allows the recognition of the right Answer Type also in questions where the words *luogo* (place) or *dove* (where) are not present (for example questions like *In quale oceano sono le Isole Canarie?*², *In quale contea della California si trova Modesto?*³, *Qual è la capitale dello Zimbabwe?*⁴, etc.). On the 200 questions of the CLEF2004 test-bed, both resources allow the system to increase the number of identified expected answer types from the 126 of the baseline prototype to i) the 171 recognized thanks to IWN and ii) the 166 recognized by exploiting SIMPLE-CLIPS. Nevertheless, some ATs were incorrectly identified and there are still about 30 questions for which the system was not able to derive any answer type.

2.3. Query expansion

We carried out an experiment by expanding all the nouns, verbs and adjectives with not maximum relevance score. In the experiment we took into consideration different types of information available in SIMPLE-CLIPS and IWN: synonyms, cross-pos synonyms, role relations,

relations between nouns and adjectives. Our approach consists in expanding the query only of one level, by taking into consideration only the target of the relations having as its source the keywords of the query. Nevertheless, when exploiting SIMPLE-CLIPS, we decided to add, to the SemUs of the first level of relation, their synonyms, in an attempt to create the same conditions we have when we exploit the IWN synsets. Synonyms and other expansions of the basic terms are composed in Boolean expressions by using an OR connector. The results of the experiment show that only a small improvement is obtained by expanding the query: the average precision increases only by few points going from 0 to 0.005 when exploiting IWN and to 0.002 when using SIMPLE-CLIPS.

2.4. Answer detection

Answer detection is the last module where lexicons are exploited. The baseline prototype exploits four sources of information to pinpoint the answer among the other candidates: i) syntactic relations, ii) Named Entities, iii) pattern matching on the text of the paragraph, iv) paragraph ranking. Example of the syntax-based rules adopted in the baseline prototype to match question and answer are the ones exploited in handling CLEF2004question#7 (*Quanti membri della scorta sono morti nell'attentato al giudice Falcone?*⁵). In that case, a match is tried, without any success, by exploiting the syntactic relations returned by the dependency analysis, i.e:

```

mod([slot?],(membro))
comp((membro), (scorta))
subj((membro), (morire))
comp (morire, attentato)

```

In the enhanced prototype, an expansion of the syntax-based rules for answer detection was performed by exploiting synonyms and hyperonym of the question keywords. In this case, the exploitation of the IWN hyperonymy relation between the word *membro* (*member*) and *uomo* (*man*) allows the individuation of the answer in the candidate paragraph.

2.5. Quantitative and qualitative analysis of Results

In the appendix of the paper, we provide tables that present the results of the system. Results show an overall improvement determined by the exploitation of the two LRs (Table 1 and 2). The improvement is obvious when one considers the ten percentage points that divide the two prototypes in general accuracy but also when we consider its distribution on the various types of question (Table 3). The types of question whose results improved in the most evident way are the ones for which the system has to analyse the ATT in order to individuate the expected answer type⁶.

² What body of water are the Canary Islands in?

³ What county is Modesto, California in?

⁴ What is the capital of Zimbabwe?

⁵ How many members of the escort died in the attack to Judge Falcone?

⁶ Questions introduced by the pronoun and adjective *Quale* (What-Which), but also by the various imperatives *dammi* (give me), *dimmi* (tell me), *nomina* (name) and by the frequent interrogative form “*come si chiama...?*” (What is the name of..?)

In what follows we try a qualitative analysis of system failures⁷ on the basis of phenomena more directly connected to the methods adopted in the two lexicons to individuate and characterize the *conceptual/semantic content of the lexical item*. These methods concern the following intertwined issues:

- i) *granularity* of the representation of the ambiguity;
- ii) *breadth* of the lexicon, i.e. the number and type of lexemes admitted in the language resource
- iii) *depth* of the lexicon, i.e. number and type of the linguistic phenomena described in the lexical entry; usefulness of such descriptions for supporting reasoning and inference.

Problems connected to these aspects are somehow transversal to all the modules of LR exploitation so we can assume they can be interpreted as structural problems, having general significance.

2.5.1. Granularity Issues

In the system, no actual WSD module is exploited and sense selection is performed only by relying on the “first sense in the corpus” heuristics. In this sense, the overall performance of the system should be conceived as a lower bound that “can only get better”. Nevertheless, no perfect WSD system exists at the moment and the problem of identification of the “right” sense of 100% of occurrences seems nowadays almost irreversible. Problems connected to sense distinction rise in every single interaction between lexicons and application. The first module whose performance is negatively impacted by incorrect sense selection is the Answer Type determination. One of the reasons behind failure in AT identification is the incorrect selection of the word sense. Not in every case, however, incorrect disambiguation has an effect on the AT identification. For example, in the case of CLEF2004question#113 *Come si chiama la compagnia di bandiera tedesca?*⁸ the selected sense of *compagnia* (company) is not the commercial one but the one referring to an *informal gathering of people*. The two cases, however, share the same AT HUMAN GROUP, so the final result is not affected by the erroneous sense attribution.

A reflection on the nature of the distinctions that drive the sense splitting in semantic lexicons is needed: it seems that, for the majority of the sub-tasks encountered in our application, a coarse granularity in the definition and representation of the lexical items is sufficient to achieve good results. QA is somehow “Named-Entity-Sensitive”: it means that being able to understand that the expected answer is the name of a ship does not have any positive consequences unless the system is also able to individuate the Named Entity class “ships” in the candidate answer. This surely has an effect on the granularity of lexical description that is required by this type of application and this can be observed when we evaluate the connection between the AT Taxonomy and the nodes of the lexical resources: in order to guarantee a successful recognition of the ATT and of other meaningful words of the question, we had to link some ATs to more than one sense of the same word. This happens, for example, for the Answer

Type YEAR>DATE that we decided to link to all the synsets in IWN with variant “*anno*” (year): The senses of *anno* in IWN all share fundamental information: they are all hyponyms of {tempo, periodo} (time, period) and they are all subsumed by the same Top Concepts, i.e. TIME and QUANTITY. This representation and organization of the distinction among the senses is of no use under the computational point of view and, even if in IWN there are four senses of the word *anno*, an automatic procedure will be unlikely to operate on them as separate senses. What we are trying to do here is not to define *the sense in abstract* but rather to understand which is the best sense grouping/organization we can operate on our semantic lexicon in order to facilitate the computational exploitation of the bulk of information stored in the resource. It is obvious that the answer to this question is not universally valid but it highly depends on the various final applications we are thinking of. When we analyse the question *In quale anno Thomas Mann ha ricevuto il premio Nobel?*⁹ we want our application to be able to derive the answer type YEAR>DATE in order to recognize the answer among the textual material returned by the Search Engine. That is why the AT DATE has been connected to all the four synsets, creating in this way a sort of super-sense of the word. The same strategy has also been applied to typical cases of *regular polysemy*, for example linking the node CITY of the AT Taxonomy to the two SIMPLE-CLIPS SemUs of *città* (city) (Città – Semantic Type: GeopoliticalLocation and Città – Semantic Type: HumanGroup). We think it is quite useful to collapse these cases of polysemy in *de facto* unique senses¹⁰, since for our application this kind of polysemy does not seem to have any important impact on the analysis of the question and in the successive steps of retrieval and answer identification. The two questions:

*In quale città la Mosella incontra il Reno?*¹¹

*Quale città è stata insignita della medaglia al valore civile?*¹²

are examples of the two readings of the word *città*: for the application, there is no actual need to distinguish the two readings, since the strategy that should be triggered in the answer detection module is the same: looking in the candidate answer for entities of the type CITY>LOCATION satisfying certain conditions.

Looking at these examples, it seems that QA requires more coarse-grained grouping of word meanings, in some way confirming what Ide and Wilks (forthcomings) state about WSD and NLP tasks. Nevertheless, our results show that this can be considered only a very general observation, not valid in many cases: sometime, the distinction between two very close senses can be appropriate for the exigencies of the application. For example, in the case of the question *Di quale gruppo Teresa Salgueiro e' la cantante?*¹³, when exploiting SIMPLE-CLIPS, the system was not able to derive the AT HUMAN GROUP because the semantics of *gruppo* was

⁹ *In What year Thomas Mann won the Noble Prize?*

¹⁰ In SIMPLE-CLIPS these cases are already connected by means of specific relation, Polysemy, but when exploiting the vertical links to derive the Answer Type it is surely simpler to connect both nodes to the AT Taxonomy.

¹¹ *In what town does the Mosel meet the Rhine?*

¹² *What town receives the medal for civic valor?*

¹³ *Of what band is Teresa Salgueiro the vocalist?*

⁷ We do not take into consideration failures deriving from the erroneous treatment of syntactic or morpho-syntactic information.

⁸ *What is the official German airline called?*

too generic to be captured by the portion of lexicon subsumed by the AT. As a matter of fact, while in IWN a specific synset was created just to gather the “social” groups, in SIMPLE-CLIPS no similar concept is available and the “social” groups are collected instead by a node of the Top Ontology. When the ATT is simply *gruppo* (like in this case) it is not recognized as human group since the only SemU available (which covers both groups of people and of things) is directly linked to the Constitutive node. It begs the question of whether it was correct to isolate a sense of *gruppo* as composed only by people, distinguishing it by the more general sense of *gruppo* that is also encoded as its hyperonym. From the point of view of our applicative exigencies, we can say that such a sense distinction is surely worth being encoded: as a matter of fact, the more granular vertical organization allows the system to circumscribe a portion of the lexicon containing similar meanings (the various synsets *association*, *organization*, *team*, *political party*, *commercial enterprise*) and to infer from occurrences of *gruppo* similar to the one in question *Di quale gruppo Teresa Salgueiro e' la cantante?* that the system has to search for the answer among Named Entities of the type companies, teams etc.

2.5.2. Breadth of the Lexicon

The two lexicons provide to the application a reach repository of lexical senses. The vast majority of the words analysed by the system were in fact found in the lexicons, even with some exception due to the fact that SIMPLE-CLIPS is relatively smaller in size than IWN. The two lexicons, however, differ for the support they provide in two specific cases, i.e. multiword recognition and exploitation of reflexive and transitive pronominal verbs.

About 16 question keywords in the CLEF test-bed should be considered not in isolation but rather as parts of multiword expressions (*bomba atomica* (atomic bomb), *campo di sterminio* (death camp), *salto con l'asta* (pole vault) etc.). Most of these MWEs are listed among the lexical entries of the IWN database, while we can state that multiwords are not present in the current version of the SIMPLE-CLIPS lexicon¹⁴. Recognition of poly-lexical units is an important sub-task, foreseen by most of the state-of-the-art QA systems. As far as our system is concerned, MW recognition is important in the module for the assessment of keyword relevance, where *pena di morte* (death sentence) and *genere musicale* (musical genre) have surely a smaller number of hyponyms than the more generic terms *pena* (penalty) and *genere* (genre). But recognizing MWEs is also of crucial importance in the module for AT identification (where analysing *unità di misura* (unit of measurement) is different from isolating the more general *unità*) and during query expansion (where expanding *campo di sterminio* (extermination camps) with the synonyms of *sterminio* (*ecatombe*, *eccidio*, *macello*, *massacro*, *strage*) (massacre, hecatomb etc.) is not productive and creates noise while it would be useful to expand the multiword expression with *campo di concentramento* (concentration camp). The fundamental

¹⁴ In SIMPLE-CLIPS only few, very general mwees were introduced as dummy entries to help categorization of homogeneous sets of senses (*unità di misura* (unit of measurement), *essere umano* (human being) etc).

issue, however, is the possibility of actually exploiting *mwees* that are encoded in IWN: as a matter of fact, in IWN *mwees* are just strings of text with one or more blanks and no information is given on the internal structure of the entry. This prevents the system to easily morphologically analyse the various parts of the entry and in this sense handling the morphological variation of the keyword in the question and in the answer is not straightforward. We think that higher acceptance of mws in the lexicon could have some very positive effects on the performance of the applications. For sure, however, the description of the syntactic structure of the entry is something that should not be missing from the entry: without a complete description of the internal structure of the *mwe* and without any clues about how it can vary in the target text, no full exploitation of this type of information will be really feasible.

While IWN is a useful provider of multiword expressions, the SIMPLE-CLIPS lexicon is more suited to allow the system to analyse and exploit *reflexive* and *transitive pronominal verbs*. As a matter of fact, a substantial difference exists on the treatment of this type of verbs between the linguistic analysis chain exploited by the system and the IWN synsets: in IWN, the transitive pronominal and reflexive forms of the verb have been encoded in distinct synsets; for example *sposare* and *sposarsi* are encoded in different synsets while the output of the *chunker* foresees the recognition of the basic form *sposare* and the encoding of the clitic. On the contrary, in SIMPLE-CLIPS a strategy coherent with the output of the *chunker* has been followed. What makes not feasible the direct exploitation of the IWN entries of this type as such is the fact that in IWN no mechanism is foreseen to represent the “reflexivity” of the verb, and no representation is given of the internal organization of the string “sposarsi”.

2.5.3. Depth of the Lexicon

In the prototype, the most exploited type of semantic relation is *hyperonymy*. Nevertheless, the observation of system failures shows that, even if links driven by the *hyperonymy* relation are the most exploited in our prototype, they are however not completely reliable. What seems to happen is that the IS-A relation has become a sort of repository of different aspects of meaning, aspects that collapse into the same label losing their important distinctions. Important reference for this kind of considerations is the work done by the Guarino and Gangemi’s research group and resulted in the OntoClean methodology (Gangemi et al., 2001). One of the problems raised by analysing WordNet with OntoClean is what is called the *ISA overloading* phenomenon. In our computational lexicons there is an over exploitation of the ISA expressive means, used to express *purpose*, *function*, *origin*, *material*, *part-whole information* etc. For example, when the system tries to determinate the specificity of the keyword *ingrediente* (ingredient) in the question *Qual è un ingrediente base della cucina giapponese?*¹⁵, fails both when uses IWN and SIMPLE-CLIPS. As a matter of fact, the two semantic lexicons represent the word meaning *ingrediente* as a synset and a SemU without hyponyms, at the same level with other substances such as *cibo* (food), *insetticida* (insect-powder), *cemento* (cement), etc. In this

¹⁵ *What is a basic ingredient of Japanese cuisine?*

way *ingrediente* is “understood” as a very specific concept while it is not. Probably, the most coherent and logically valid solution would be finding a representational device capable of stating that “all substances can be ingredients if they are used to prepare dishes or medicines” and to precisely recognize the telic and constituency dimension of *ingrediente*. The representation closer to this type of solution is the one proposed in SIMPLE-CLIPS, where *ingrediente* is a SemU without hyperonym, directly connected to the Top Ontology node Constitutive. Unfortunately in SIMPLE-CLIPS nothing links the concept *ingrediente* to the taxonomy of substances. Using Gangemi et al.’s words: we should distinguish between the *type* and the *role*, preserving the ISA vertical structure for the types (the actual types of substance, like *dust*, *cement*, *grease* etc.) and allowing for a horizontal account of their telic dimension by means of specific semantic relations.

The representation the two lexicons give of the concept *ingrediente* is also not useful when the system performs the enrichment of the query with terms hyponyms of the ATT: looking at the answer to our question (*Qual è l'ingrediente base della cucina giapponese?*¹⁶), we learn that the basic ingredients are *pesce* (fish), *tofu* and *verdura* (vegetables). In the end, what we would have needed was something that links *tofu*, *pesce* and *verdura* to the ATT *ingrediente*.

When using ItalWordNet, the exploitation of the all-level hyponyms of the answer type term is often effective, and the system is able to generate the query with the candidate answer that leads to the extraction of the answer paragraph. This is true, for example, for questions like *Qual è l'unità di misura di frequenza?*¹⁷, *Come vengono chiamati i piloti suicidi giapponesi?*¹⁸, *Che lingua si parla in Germania?*¹⁹ etc. Nevertheless, sometime the exploitation of the ISA relation shows its points of weakness. Examples of limitation in the use of hyperonymy are the cases of the ATTs *sintomo* (symptom) and *fattore di rischio* (risk factor): *mal di gola* (sore throat) and *febbre* (fever) are not inherently symptoms but rather pathological conditions, as well as *ipertensione* (hypertension) and *fumo* (smoke) are not inherently *fattori di rischio*. A particularly difficult situation is represented by the impossibility of exploiting the subset of professions in IWN and SIMPLE-CLIPS in the case of questions of the type “*quale è la professione/l'incarico/l'ufficio...?*”²⁰. In the test bed we find the two questions *Quale incarico ricopre Ariel Sharon?* and *Qual è la professione di James Bond?*. The answers are respectively *ministro degli esteri* and *agente segreto*, but we can see that neither in IWN nor in SIMPLE-CLIPS we can find them listed among the hyponyms of *professione-incarico*: in IWN the list of professions is organized (as it happens in WordNet) solely as a taxonomy having as root the synset {*persona, essere umano, uomo, individuo*} (human being), with an intermediate level represented by the fictitious synset {*lavoratore*} (worker). Following the OntoClean recommendations (Gangemi et al., 2001), we should avoid

to encode the various professions under the Human node, since a role (the profession) cannot be subsumed by a type (the human being). If this recommendation would have been followed, the system would have been able to exploit the subset of lexicon dedicated to professions to individuate the answer to this type of questions. Nevertheless, we are not persuaded that a simple shift of the taxonomy from the human to the activity node would have been completely resolutive of the problem. As a matter of fact, even if the professions organized under the node *PROFESSIONE>ATTIVITÀ* can be exploited in the answer detection phase, the classification under the HUMAN node is useful when we want to derive the Answer Type (in that case, if the question asks *Quale presidente americano è stato renitente alla leva?*²¹, it is of primary importance to allow the system to *understand* that the answer is probably a person’s name, not an activity). In our perspective, the casting out of nouns constituted by usefulness in application is obviously of great importance to verify the correctness of a choice thus, in this specific case, the fact that two distinct yet specular modules of the same system require two different classification of the ATT is very problematic.

Another point of weakness of the exploitation of LRs is *mereotopology*, i.e. the theory of parts and of wholes. If we look at the taxonomy of the most general sense of *gruppo* (group) in IWN and in SIMPLE-CLIPS we can see that it is an amalgam of very heterogeneous concepts²² such as *imbracatura* (sling), *sciame* (swarm), *bendaggio* (bandage), *contabilità* (bookkeeping), *attrezzatura* (equipping), etc. In this way we lose the fundamental dimension of meaning that constitutes the backbone of such concepts: all these word meanings are not simply sets or groups, but rather they are physical objects (the *sling* and the *bandage*) and activities (the *bookkeeping*). Again, we are probably in front of what Gangemi et al. (2001) describes as a case of *IS-A overloading*, i.e. the phenomenon of *reduction of sense* according to which “the ISA link points to an aspect of the meaning of a given concept that does not fully account for its identity”. In this case, the attribution of the ISA is clearly due to the practice of identifying in the genus term of the definition the hyperonym of the *definiendum*, regardless of the loss of information. Even when the “constitutive” dimension of meaning is very marked, it can be the case that the hyperonymy link to {*insieme, gruppo*} is not what is needed to derive the Answer Type capable of matching question and answer. This is the case, for example, of CLEF2004question#137: *Dammi il nome di una catena di Fast Food*²³. The system was not able to derive the Answer Type because in IWN the ATT *catena* (chain) is represented as a group. A different representation, consisting of ascribing *catena* to the synset {*azienda, ditta, compagnia*} (company) would have easily allowed the recognition of the AT HUMAN GROUP and the

¹⁶ What is a basic ingredient of Japanese cuisine?

¹⁷ What is the frequency unit?

¹⁸ What are Japanese suicide pilots called?

¹⁹ What language is spoken in Germany?

²⁰ What is the profession/office..?

²¹ What American president failed to report for military service?

²² Many lexemes are in these taxonomies only “thanks” to the content and form of their lexicographic definition. As a matter of fact, if we look at the definition of *imbracatura*, we can see that it is defined as “the set of ropes used to sling”, *bendaggio* is “a set of bandage”, *contabilità* is “the set of books and accounts of an organization”, etc.

²³ Name a fast food chain

identification of the answer *McDonald's* (a Named Entity of type Organization).

2.5.4. Synonymy and Xpos relations

Evaluating the impact of the exploitation of xpos relations and synonymy is not simple: adoption of query expansion methods in our prototype gives a very small improvement. One of the things that minimize the impact of the use of LRs is the adoption of the stemming technique in the IR module. As a matter of fact, most of the time, the information conveyed by the xpos semantic relations in IWN and by the predicate object in SIMPLE-CLIPS is not really useful because the stemmer has already correctly identified and extracted the root of the word, thus enabling the retrieval of occurrences not only morphologically but also semantically correlated. Stemming and exploitation of semantic information in this way are concurrent strategies to obtain the same results; the problem is that using stemming techniques is much simpler and more straightforward than navigating through the SemUs and synsets of our resources, collecting correlated items and disambiguating word senses. Query expansion is more useful when it involves the exploitation of *synonyms*.

3. References

Bertagna F., Chiran L., Simi M., *ILC-UniPi Italian QA*, in Peters C. and Clough P. and Gonzalo J. and Jones G. and Kluck M. and Magnini B. (eds.), Fifth Workshop of

the Cross-Language Evaluation Forum (CLEF-2004), Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany, 2005.

Gangemi A., Guarino N., Oltramari A., *Conceptual Analysis of Lexical Taxonomies: the Case of WordNet Top-Level*. In C. Welty, B. Smith (eds.), Proceedings of the 2001 Conference on Formal Ontology and Information Systems, Amsterdam, IOS Press, 2001.

Ide, N., Wilks, Y., *Making Sense About Sense*. In Agirre, E., Edmonds, P. (Eds.), Word Sense Disambiguation: Algorithms and Applications, Springer, forthcoming.

Roventini A., Alonge A., Bertagna F., Calzolari N., Girardi C., Magnini B., Marinelli R., Speranza M., Zampolli A., *ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian*. In Zampolli A., Calzolari N., Cignoni L. (eds.), Computational Linguistics in Pisa, Special Issue of Linguistica Computazionale, Vol. XVIII-XIX, Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma, 2003.

Ruimy N., Monachini M., Gola E., Calzolari N., Del Fiorentino M.C., Ulivieri M., Rossi S., *A Computational Semantic Lexicon of Italian: SIMPLE*. In Zampolli A., Calzolari N., Cignoni L. (eds.), Computational Linguistics in Pisa, Special Issue of Linguistica Computazionale, Vol. XVIII-XIX, Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma, 2003.

#Answer	#Right	#Wrong	#IneXact	Overall Accuracy %
200	91-111	87-71	22-18	45.5-55.5

Table 1: comparison between the results of the baseline and of the IWN-based enhanced prototypes. (on the right side in each column are the baseline results, on the left side are the enhanced results).

#Answer	#Right	#Wrong	#IneXact	Overall Accuracy %
200	91-100	87-81	22-19	45.5-50

Table 2: comparison between the results of the baseline and of the SIMPLE-CLIPS -based enhanced prototypes. (on the right side in each column are the baseline results, on the left side are the enhanced results).

Question Type	Improvement (%) -IWN	Improvement (%) -SIMPLE-CLIPS	Difference in the obtained improvement (%)
Quale (pronoun)	17.6	11.7	5.9
Come si chiama	16.7	0	16.7
Quale/ Che (adj)	11.7	7	4.7
(Che) Cosa (pn)	10.8	10.8	0
Others (dimmi, dammi, nomina)	14.2	14.2	0
Quanto (pn)	11	11	0
Quanto (adj)	5.5	0	5.5
Chi	2.8	0	2.8
Quanto (adv)	0	0	0
Come	0	0	0
Dove	0	0	0
Cosa (DEF)	0	0	0
Quando	0	0	0

Table 3: comparison between the improvement obtained with IWN and the one obtained with SIMPLE-CLIPS