Encoding Commonsense Lexical Knowledge into WordNet

Gianluca E. Lebani[†‡]
[†]Center for Mind/Brain Sciences
University of Trento, Italy
gianluca.lebani@unitn.it

Emanuele Pianta[†‡]
[‡]HLT Group
Fondazione Bruno Kessler, Italy
pianta@fbk.eu

Abstract

In this paper, we propose an extension of the WordNet conceptual model, with the final purpose of encoding the common sense lexical knowledge associated to words used in everyday life. The extended model has been defined starting from the short descriptions generated by naïve speakers in relation to target concepts (i.e. feature norms). Even if this proposal has been developed primarily for therapeutic purposes, it can be seen as a generalization of the original WordNet model that takes into account a much wider and systematic set of semantic relations. The extended model is also an enhancement of the psycholinguistic vocation of the WordNet model. A featural representation of concepts is nowadays assumed by most models of the human semantic memory.
For testing our proposal, we conducted a feature elicitation experiment and collected descriptions of 50 concepts from 60 participants. Problematic issues related to the encoding of this information into WordNet are discussed and preliminary results are presented.

1 Introduction

WordNet (WN: Fellbaum, 1998) is the largest and most systematic lexical database in electronic format available nowadays. Nevertheless, it is hard to maintain that such a successful and widely used resource contains a complete (or near-to-complete) representation of the information that is encoded in the mental lexicon of English speakers. The lack of completeness is not only referred to the coverage of lexical unites, or to the population of the already defined lexical and semantic relations (see for instance the sparse instantiation of the meronym relation), but also to structural aspects, such as: the number and type of the encoded relations; the encoding of the strength (or any similar quantitative notion) of relations, in order to represent, for instance, prototypicality effects; the encoding of quantifiers and logical operators, as an important aspect of the knowledge associated to concepts; the encoding of syntagmatic information, e.g. collocations and selectional preferences or restrictions.

In order to overcome such limitations, in the last twenty years, many extensions of the WN conceptual model have been proposed by the creators of the original Princeton WordNet (PWN) and by other scholars in projects such as EuroWordNet (EWN: Alonge et al, 1998), MultiWordNet (MWN: Pianta et al, 2002), WordNet Domains (Bentivogli et al, 2004) and BalkaNet (Tufiş, 2004). However, none of such proposals has tried to define, on the basis of psycholinguistic evidence, a close set of semantic relations that are expected to be able to represent all (or most) of the meaning aspects conveyed by a concept. In this paper we make such an attempt, starting from the requirements of a very specific application scenario, in which an electronic lexical database is used to support speech therapists in their daily work with aphasic patients.

2 Background and Motivation

Anomia is a common symptom associated with aphasia. Most patients affected by an acquired linguistic disorder due to a brain damage experience some difficulty in retrieving or producing words. In this context, computers can be helpful in many ways: from assisting the therapist in the rehabilitation, to helping the patients in his/her everyday life (Petheram, 2004). Given the great variability of forms and severity in which anomia can manifest itself, a requirement that any assistive tool has to meet is to be flexible enough to fit the needs of different classes of patients.

STaRS.sys (Semantic Task Rehabilitation Support system) is the outcome of a joint effort between Fondazione Bruno Kessler and the CIMeC Center for Neuropsychological Rehabilitation (CeRiN). The aim of this project is the creation of a tool for supporting the therapist in the preparation of rehabilitative tasks for Italian-speaking patients affected by anomia.

Typically, the information exploited in semantic rehabilitation tasks can be represented as concept-description pairs like <chair> has four legs, <airplane> flies[1]. Notably, this is the same kind of information that is collected by scholars who study the characteristics of conceptual knowledge by running *feature generation* experiments, that is by asking speakers to describe concepts (cfr. Murphy, 2002). We've argued elsewhere (Lebani and Pianta, 2010b) that the WordNet conceptual model fits well the STaRS.sys requirements, so that we chose to build the STaRS.sys semantic knowledge starting from the Italian MWN lexicon (iMWN).

We believe that only a lexicon organized on the basis of psycholinguistic evidence can be flexible enough to meet the STaRS.sys requirements. As a matter of fact, many psycholinguistic assumptions lay at the basis of the WN model (e.g. Miller, 1998), and its psychological validity has been tested explicitly or implicitly by several scholars (e.g. Fellbaum, 1998b; Izquierdo et al, 2007; Barbu and Poesio, 2008). However, just few of the many WN extensions proposed in the last two decades seem to be based on psycholinguistic hypotheses and methodologies.

An outstanding exception to this trend is the evocation relation by Boyd-Graber et al (2006), who proposed the introduction of weighted, oriented arcs between pairs of synsets, e.g. from {car} to {road}, representing how much a concept evokes the other. The relation has been populated by collecting judgments from speakers (Boyd-Graber et al, 2006; Nikolova et al, 2011).

There are many similarities between our work and that by Boyd-Graber and colleagues. In both proposals, WN is enriched with speaker generated semantic information, and the encoding of this information requires an extension of the WN model. Also, both proposals are exploited for assistive purposes. The resource by Boyd-Graber and colleagues has been adopted as the semantic knowledge base behind the tool ViVa (Nikolova et al, 2009), a visual vocabulary designed for aiding anomic patients in their everyday life. In a similar way, STaRS.sys will be part of a computer aided therapy tool designed for supporting therapists in their daily work with patients. In spite of the commonalities between the two projects, we observe that a generic evocation relation seems not to meet all the requirements of speech therapists, which need instead a more fine-grained classification of semantic relations. For the STaRS.sys purposes, we need to encode structured lexical information that is more similar to what can be obtained by exploiting a feature generation paradigm, than to what can be obtained through free associations.

Due to the great variability of impairment shown by anomic patients and to the lack of resources, the preparation of a therapeutic task for anomia rehabilitation is a manual work on behalf of the therapist. STaRS.sys is a system thought for being helpful in this preparatory phase by helping the therapist to (1) retrieve concepts, (2) retrieve information associated to concepts and (3) compare concepts. In the knowledge base underlying this system, the following kinds of information have to be available for every concept: its position in a conceptual taxonomy; a set of featural descriptions (FDs) classified according to the types of knowledge conveyed; a value of prototypicality and of word frequency.

As argued in Lebani and Pianta (2010b), the WN conceptual model fits well our needs, because of its cognitive plausibility, for its ease of use and because it is based on a fully specified is-a hierarchy. Moreover, it is powerful enough, with some modifications, to represent the information contained in featural descriptions. FDs like <cup> is used for drinking can be represented in WN as a relation (say *is Used for*) holding between the described (or "source") synset {cup} and the most prominent synset of the description, i.e. the ("target") synset {drink}.

A similar assumption has been used by Barbu and Poesio (2008), who analyze the overlap between the semantic information encoded in PWN and in the collections by McRae et al (2005) and Garrard et al (2001). In their analysis, the authors, who also considered information contained in glosses, estimated that the overlap between PWN and existing norms collections can vary between 22 and 40% (depending on the collection and on the method used to calculate the overlap). The same analysis showed that the WN coverage with regards to FDs is highly skewed (e.g. categorical information is highly present, whereas functional information is missing).

To overcome some of the limitations of the current WN model, Lebani and Pianta (2010b)

[1] Concepts and features are printed in courier new font. When reporting a concept-feature pair, the concept is further enclosed by <angled brackets>. WordNet synsets are enclosed by {curly brackets}. Feature types, relations and concept categories are reported in *italics times roman*.

proposed to add a set of 25 semantic relations in a dedicated version of iMWN called StarsMultiWordNet (sMWN), with the final objective of finding a complete set of intuitive and cognitively plausible relations representing lexical meaning. This extension has been built by combining experimental evidence from existing feature norms with theoretical proposals developed in lexicography, linguistics and cognitive psychology (for details, see Lebani and Pianta, 2010a).

This paper presents the results of a pilot study aiming at populating the extended set of WN relations by collecting FDs from subjects in a controlled setting, and encoding them into sMWN. Section 3 will present available feature norms collections; Section 4 will illustrate the results of the collection experiment and Section 5 will comment on the issues faced when actually mapping FDs into WN relations.

3 Available feature norms collections

Since the early times of the cognitive psychology enterprise, the feature norm paradigm has been widely employed in the investigation of the human's conceptual representation and computation (cfr. Murphy, 2002). Despite this wide use, to date there are few freely available collections (Garrard et al, 2001; McRae et al, 2005; Vinson and Vigliocco, 2008; De Deyne et al, 2008; Kremer and Baroni, 2011). These resources are strongly influenced by the goals and theoretical framework of the connected studies, so that they differ substantially on the quantity and kind of described concepts, on the procedure adopted for collecting and processing features and on the classification adopted for classifying them.

In the canonical paradigm, speakers are simply asked to describe a concept. On the one side, this approach has shown his utility for investigating which concepts and/or properties are easier to recall. On the other side, however, it produces a very sparse population of the various components of lexical meaning. As an example, consider that 75.44% of the descriptions of the McRae dataset belongs to just 7 types out of 27. Many factors may contribute to this sparseness, among which the organization of the human semantic memory itself. It is also probable, however, that part of this disproportion is due to the methodology exploited for eliciting and normalizing descriptions. Because of the sparseness of property types, it turned out that none of the available collections can be efficiently exploited for our purposes, as we need to cover the largest and most varied set of semantic aspects as possible. We coped with this issue by adopting a question answering paradigm for the elicitation experiment, as described in Section 4.

Another problematic issue in existing collections concerns the normalization of raw descriptions. Even if this practice is claimed to be as much conservative as possible, the ways in which it is usually carried out leads, from our point of view, to a loss of knowledge. Furthermore, our feeling is that too much is left to the interpretation of the persons in charge of the normalization. As an example, in the Kremer norms, the description of the pair <garage> can be used as a utility room is paraphrased as used for storing. However the original description could be used to encode the information that garage and utility room are similar concepts, encoded by the coordination relation in our relation scheme. The paraphrase, on the other side, leads to missing this information. We will show in Section 5 how iMWN can be used to alleviate such problems.

4 A new norms collection

Given the limitation of existing norms collections, we decided to conduct an elicitation experiment adopting the stimulus set by Kremer and Baroni (2011) and a comparable number of participants, with a slightly different methodology. This allows for the comparison of our dataset with the only freely available norms in Italian.

4.1 Experimental Setup

Participants: 60 Italian speakers participated in the the experiment. Their age ranged from 19 to 55 years (mean: 28.9, s.d. 9.27). All subjects were recruited in the university environment.

Materials: The stimulus set was composed by 50 concepts belonging to the following 10 categories: *bird*, *body part*, *building*, *clothing*, *fruit*, *furniture*, *implement*, *mammal*, *vegetable* and *vehicle*. Kremer and Baroni (2011) selected these same 50 concepts for the reasonable unambiguousness of their lexical realizations.

Procedure: The descriptions have been collected through an on-line experiment. 12 groups of 5 tasks were prepared, each task composed of 10 randomly ordered concepts, one for each category. In this way, every concept has been described by 12 subjects, and no participants received a questionnaire that was previously assigned to another participant.

The semantics of each relation has been paraphrased as a question of the form: "*what are the portions of a [concept]?*" (for the *has Portion* relation). This allowed us to populate as much as possible all feature types, and to reduce need for interpretation in the normalization process.

Every subject has been presented a concept per web page, followed by a set of relevant questions. For each question, examples were available in the online documentation, accessible by clicking on the question text. Subjects were instructed not to report any anecdotal or technical knowledge, and they were allowed to leave a field empty if they didn't come up with any answer. Participants were trained on two example concepts (`cat`, `knife`) for which some suggestions were supplied in different ways (pre-filled fields, auto-completion).

4.2 Results

We collected 18,884 raw FDs, that is a mean of 377.68 descriptions (s.d. 60.71) per concept. Every subject, on average, produced 314.73 (s.d. 115.68) descriptions over 10 concepts and 31.47 descriptions per concept (s.d. 13.71).

In a pre-processing phase every FD has been analyzed as an instance of one the feature types proposed in Lebani and Pianta (2010b). In doing so, we exploited the fact that all FDs have been produced as an answer to a specific question that was formulated on the basis of one of these feature types. Nevertheless, the appropriateness of the descriptions produced by the subjects with regard to the expected feature type was manually checked by one of the authors. This led to the deletion of 1,023 raw descriptions because they were conveying technical, autobiographical or patently wrong information. Given the remaining descriptions, in 2,247 cases we re-categorized the FD, and associated it to a feature type different from that associated to the question. Summing up, a total of 3270 features (17.3% of the total) underwent some change in this phase.

Comparison with the Kremer norms: A preliminary quantitative evaluation of our dataset shows that we have collected 18,884 descriptions against 8,250 descriptions in Kremer dataset. Other meaningful comparisons concern the number of descriptions per subjects (314.73 vs. 123.48), the number of descriptions per concept (377.68 vs. 170.4) and the average of feature per concept produced by every subject (31.47 vs. 4.96). These data suggest that our strategy paid off, by providing a richer and more systematic set of feature descriptions for each concept.

5 Encoding descriptions into WN

The second step of our pilot study consisted in manually populating sMWN with the *normalized* version of the 1,785 raw descriptions collected for the following five concepts: `seagull`, `finger`, `chair`, `corn` and `airplane`.

5.1 The encoding procedure

The manual encoding of the FDs content in sMWN is based on two main. First, although a certain amount of interpretation cannot be avoided, this should be kept at a minimum. Second, the reduction of the informative content of a description should be used only as a "last resort" strategy.

Normalization: In works belonging to the feature generation paradigm, the collection of the descriptions is always followed by a normalization step, in which FDs conveying the same semantic information are merged. In the simplest case, two linguistically identical FDs uttered by different subjects are clustered together and encoded as a single concept-description pair, by keeping track of the absolute frequency.

The most compelling situation, however, involves those FDs that express the same property in linguistically different ways. In most cases, the linguistic forms assumed by these equivalent FDs are quite similar, like in the following answers to the question "*who does typically use a [concept]?*": `is used by the pilots`, `is used by a pilot`, `a pilot`, `pilots`.

In other cases, however, different FDs expressing the same property can assume very different linguistic forms, even involving different lemmas or different syntactic structures. A major problem of most works belonging to the feature generation paradigm is that a clear explanation of how equivalent descriptions are identified is often missing. As an example, raw descriptions like `is a quadruped` and `has four legs` can be seen as exemplars of the same feature (e.g. `has four legs`) and merged (cfr. Vinson and Vigliocco, 2008). It is questionable, however, that these expressions convey the same information. A quadruped is "an animal that moves by using four legs", and reducing its definition to "having four legs" may be reductive.

In our approach equivalent descriptions are more explicitly defined as descriptions sharing the same semantic relation and the same source and target synsets. Accordingly, then, we consider the two FDs `<wheel> is a component of a car` and `<wheel> is an auto part` equivalent because they can be both mapped into a meronymic relation linking {wheel} and {car, auto}.

Ambiguity: In a number of cases the FD contained an ambiguous word, so we had to choose an appropriate synset for it. We identified two variants of this situation.

If the concurrent synsets are in a hyponym relation, and the property is possessed by all the hyponyms of the more general synset, this is selected. As an example, the target concept of the FD <coltello> è usato dal cuoco (<knife> is used by the cook/chef) can be represented in sMWN as the Italian equivalent of either {cook} or {chef}, where the first is a hypernym of the second. In this situation, given that the property of "using a knife" is possessed by all hyponyms of {cook}, our choice falls on the more general synset.

Instead, when the property cannot be predicated of all the hyponyms of the more general synset, we opt for the more specific. Consider the pair <ciliegia> cresce in giardino ("<cherry> grows in gardens/grounds"). The target concept, in this case, can be encoded with the Italian translations of both {grounds} and {garden}. However, since cherry trees do not usually grow in a {parvis} or in other sMWN hyponyms of {grounds}, we encoded this feature as a relation holding only between {cherry} and {garden}.

In most cases the synsets corresponding to the ambiguous words are not in hyponymy relation. As an example, given the FD <corn> can be found in a cellar, the target concept cellar can be encoded as either {basement, cellar} or {root_cellar, cellar}. Given that both synsets look plausible, we chose to double the concept-description pair in the database.

Loose Talk: Speakers may ignore some terms or they simply may not recall them in a certain moment. As a consequence, some phrases express concepts that could be expressed by an existing term, as in the FD is used by people who cook.

In the standard feature generation paradigm, descriptions like these can be interpreted in many ways. They may even be re-phrased as features of a different kind, such as is used for cooking. In our approach, the rephrasing is guided by the synsets and glosses available in WN. In our case, we encode the phrase people who cook with the synset {cook} given the gloss "*someone who cooks food*".

Compositionality: One of the most complex issues faced in the encoding of FDs into sMWN is given by complex linguistic descriptions like in the FD <seagull> has an orange beak. Complex target concepts such as orange beak cannot be represented as WN synsets, in that in this model synsets are bound to be lexical units.

The solution has been to exploit the notion of *phraset*, introduced in MWN for coping with cross-language lexical gaps and with complex ways to express a concept for which a synset already exists (cfr. Bentivogli and Pianta, 2004). In this way, a free combination of words like coltello da pane (the Italian translation for breadknife) is encoded as a phraset {GAP}{coltello_da_pane} linked by the lexical relation *composed-of* to the synsets {coltello} ('knife') and {pane} ('bread'), and by the semantic relation *hypernym* to the synset {coltello} ('knife').

In Lebani and Pianta (2010b) we proposed to exploit the same structure for representing complex descriptions, with the important addition, shown in figure 1, that, given the extended set of relations available in sMWN, we can now provide a precise characterization of the semantics of the modifier (in our example orange). In other words we can link the phraset to the "modifying" synset also with a semantic relation. This allows us to keep track of properties of the described concept that would be otherwise lost.

The set of normalized features: The outcome of the encoding phase has been the insertion into sMWN of 871 relation instances for 5 concepts. On average, every concept received 174.6 relation instances (s.d. 33.44). The results of this encoding confirm that the WN model is apt to represent the kind of commonsense knowledge carried by featural descriptions.

The simplest encoding procedure has been, as a matter of fact, powerful enough for represent-

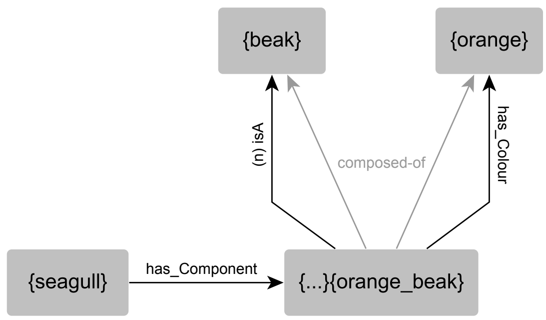

Figure 1: Representation of the FD <seagull> has an orange beak

ing the vast majority of the collected descriptions. The semantics of 795 normalized FDs (91.3% of the total) could indeed be fully encoded as a semantic relation between two simple synsets. In 137 cases (15.7%) a synset for the focal concept of the description was missing. By exploiting iMWN, 59 equivalent descriptions have been merged together into 29 relation instances.

The encoding of 71 normalized features required the creation of one or more phrasets. This process lead to the creation of 76 new phrasets, 5 of which are concept-feature doublings encoded for coping with cases of ambiguity. As an example, the FD <dito> è usato per afferrare oggetti ("<finger> is used to grab/ grasp objects") has been represented in sMWN by encoding two instance relations between the source concept {finger} and the Italian equivalents of the phrasets {GAP}{grab_ _object} and {GAP}{grasp_object}.

In the disambiguation of words we faced an average ambiguity of 3.2 synsets per lemma (s.d. 2.87), and 64 descriptions (7.3% of the sample) have been encoded with more than one relation instance. In 32 cases it was impossible to represent all the information expressed by a FD, so that just a portion of it has been encoded into WordNet. Accordingly, FDs like <seagull> uses the beak for fishing and <finger> is used with the other fingers, has been reduced to the FDs <seagull> uses the beak and <finger> is used with fingers, respectively.

Finally, only 5 raw descriptions were discarded because an efficient way to encode them was not found. Examples of FDs that couldn't be encoded at all into sMWN are <seagull> is partially black and <chair> is about half the size of a man.

5.2 Modifying the WN model

Even if the bulk of the design of sMWN is the WN model implemented in iMWN, some minor modifications have been necessary to cope with some recurrent problematic kinds of descriptions.

Apart from the exploitation of the phraset structure, we used *relation features*, that is features (labels) associated to relation instances, in order to refine the semantics of a specific relation-concept pair, along the lines of the proposal advanced by Alonge et al (1998) in the context of the EuroWordNet project.

Negation: In some norms collections, e.g. the McRae database, negative statements are treated as a class on their own, so that FDs like <bike> doesn't have an engine and <chicken> cannot fly are treated as conveying the same type of information. However, for our purposes, it is important to encode not only that a concept does not possess some property, but also the property it does not possess.

Our solution is the exploitation, in sMWN, of a *negative* operator analogue to that implemented in the EWN database. In this way, a FD like <chicken> cannot fly is encoded as a relation of type *is Involved in* between {chicken} and {fly} and the relation is marked with the *negation* relation feature.

In accordance with the rationale behind the implementation of the *negation* operator in EWN, we noticed that the properties negated by our speakers can be seen as blocking "expected" undesired implications. In our example, indeed, the negated property fly is a distinctive property possessed by birds, the general category to which the described concept belongs.

Cardinality: This issue affects every work belonging to the feature generation paradigm. Many different solutions have been proposed, but none of them is useful for our purposes. As an example, in Vinson and Vigliocco (2008), descriptions such as has 4 wheels are split into the two concepts 4 and wheels. However, what is predicated in the pair <bus> has 4 wheels cannot be equivalent to what is encoded by associating the concepts 4 and wheel to the concept bus. McRae and colleagues, on the other side, treated these cases by splitting them in two features (has wheels and has four wheels), thus introducing some redundancy in their data.

Our proposal is to encode cardinality by means of a *has cardinality* relation feature that specifies the number, or range of numbers, of the elements of the set referred to in the description. Accordingly, pairs like <bus> has 4 wheels have been encoded as a *has Component* relation, marked with a "*has cardinality:4*" label, holding between the synsets {bus} and {wheel}. When encoding FDs involving the same synsets with different cardinalities (e.g. <truck> has wheels, may have 4 wheels, may have 6 wheels), we clustered them by marking the range or set of different cardinalities (in our example, "*has cardinality:4,6*").

Certainty features: Another common problem for the building of norms collections is the

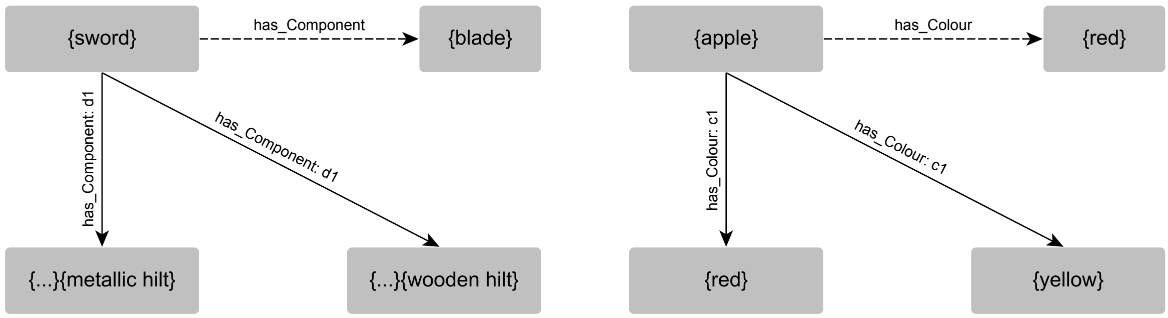

Figure 2: Representation of the FDs <sword> has a metallic or wooden hilt (left) and <apple> can be red and yellow (right).

treatment of modifiers like “generally”, “most of the times” and “sometimes”. Standard approaches to feature norms collection remove such expressions in the normalization phase. Also standard WN encoding of semantic relations ignores any kind of qualification of the probability or strength of semantic relations between concepts.

However we think that by ignoring this kind of information an important aspect of lexical meaning gets lost. In the same vein, Boyd-Graber et al (2006) argue for the usefulness of adding to the WN model a relation (evocation) that can be characterized also from the point of view of its strength.

We propose to add a relation feature, called Certainty, representing the intuition of the language speaker about how strong is his/her expectation that a certain relation holds between the instances of two concepts. We distinguish four levels of expectation:

- *True by definition*: the speaker thinks that the relation between two concept instances holds because of how the concepts are conventionally defined; no exceptions are admitted: <cat> is a feline.
- *Certain*: the speaker expects the relation to hold unless an anomaly occurs, which needs a causal explanation: <man> has arms, <socks> always come in a couple.
- *Probable*: the speaker expects the relation to hold most of the times; however if this does not occur it is not perceived as an anomaly. This feature is associated to pairs like <wardrobe> is typically made of wood.
- *Possible*: the speaker expects the relation to occur sometimes, but not most of the times. This feature is associated to FDs like: <wardrobe> can be made of plastic.

It should be stressed that in the above definitions we are interested in representing a subjective, speaker-oriented, notion of possibility/probability instead of the corresponding formally oriented notions defined in modal logic (Hughes and Cresswell, 1996). Note also that when a FD does not include any type of modifier, it is impossible to decide which of the four classes above it belongs to. Because of this, we represent the Certainty feature only when an explicit linguistic clue allows us to infer a value for it. In all other cases the value of the feature is undefined. We experiments aiming at systematically collecting the value of the certainty feature for all relations, see Nikolova et al (2011).

Conjunction and disjunction: the last set of relation features introduced in sMWN are an implementation of the *conjunction/disjunction* labels introduced in EWN for marking the relation holding between relations of the same type that have been predicated of a certain concept.

In sMWN, we set a default value for every semantic relation. As an example, by default the *has Component* descriptions stand in a conjunctive relation, while the *has Colour* ones are disjunctive. As in EWN, moreover, special cases are marked by adding labels to the semantic relations. In this way, the two descriptions <sword> can have a wooden hilt and <sword> can have a metallic hilt have been encoded in sMWN as shown in figure 2 (left), while figure 2 (right) shows how we encoded conjunctive FDs such as <apple> can be red and yellow, given the disjunctive default for the *has Colour* relation. In this figure, “d_i”/”c_i” stands for “disjunction”/“conjunction” and the index points to the other relations standing in a disjunctive/conjunctive relation.

5.3 Comparison with the Kremer Sample

We can get some indications of the goodness of our methodology also from a quick comparison with the information collected for the same 5 concepts in the Kremer dataset. For these concepts Kremer and colleagues collected 832 raw descriptions. We annotated their dataset with our feature types, obtaining 231 distinct properties, that is, a mean of 46.2 properties per concept (s.d. 7.95). A chi-square analysis failed to highlight a significant difference in the distribution of raw descriptions across concepts in the two samples ($p > .5$). However, the difference in the average number of FDs per concept is significant ($W = 25$, $p < .01$).

On the basis of their semantics, most of the properties of our classification can be grouped into types classes (for details, see Lebani & Pianta, 2010a). A chi-squared analysis highlighted a significant difference in the distribution of descriptions among the different relation classes ($\chi^2 = 43.97$, $df = 6$, $p < .001$). In details, the results of a Pearson residual test showed a significant medium size that the *part-of* and the *associated events and attributes* relation classes have been relatively more represented in the Kremer sample, while the weight the *perceptual* relations is bigger in our sample.

While in our sample there are on average 30.1 descriptions for the 29 represented feature types (s.d. 22.24), in the re-tagged Kremer sample the 23 represented feature types received, on average, 10.04 descriptions (s.d. 9.88).

Our sample, finally, seems to suffer a little less from the problem of disproportionate representation of certain types over others reported by Kremer and Baroni (2011). In the sample from their dataset, indeed, the 6 most frequent relations account for the 62.8% of the whole set of descriptions, while in our sample this measure reduces itself to the 45.1%.

6 Conclusions and future work

In this paper we presented our reflections and preliminary work for the creation of a WordNet that can be exploited for therapeutic purposes. Even if created with a specific applicative use in mind, we conceived this resource as to be able to represent every kind of knowledge that can be associated with a concrete concept.

By modifying the WN model, we have been able to represent a subset of the descriptions we collected from 60 Italian speakers. Even if we concentrated only on a subset of our collection, we feel safe to claim that we demonstrated that it is possible to represent in a WN-like resource almost all of the semantic information that can be collected through a description elicitation experiment.

There are, still, many steps left to go. We are currently mapping all the remaining features of our collection and we are testing the reliability and the intuitiveness of our feature type classification. Given that building a norms collection is a time consuming task (McRae and colleagues begun working on their collection in the 90s), an issue that we will face in the immediate future is how to automatically mine and annotate the commonsense knowledge to encode into WN.

Furthermore, being our resource based on a multilingual version of WN, i.e. MWN, another issue we're going to pursue is the evaluation of the portability of the information we elicited from our participants to languages other than Italian.

Acknowledgments

This work has been partly supported by the *Provincia autonoma di Trento* and the *Fondazione Cassa di Risparmio di Trento e Rovereto*.

References

Antonietta Alonge, Nicoletta Calzolari, Piek Vossen, Laura Bloksma, Irene Castellon, Maria A. Marti and Wim Peters. 1998. The linguistic design of the EuroWordNet database. *Computer and the Humanities*, 32 (2,3): 91-115.

Eduard Barbu and Massimo Poesio. 2008. A comparison of Wordnet and Feature Norms. *Proceedings of GWC 2008*: 56-73.

Luisa Bentivogli and Emanuele Pianta. 2004. Extending WordNet with Syntagmatic Information. *Proceedings of GWC 2004*: 47-53.

Luisa Bentivogli, Pamela Forner, Bernardo Magnini and Emanuele Pianta. 2004. Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. *Proceedings of COLING 2004* :101-108.

Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson and Robert Schapire. 2006. Adding Dense, Weighted, Connections to WordNet. *Proceedings of GWC 2006*: 29-36.

Simon De Deyne, Steven Verheyen, Eef Amel, Wolf Vanpaemel, Matthew J. Dry, Wouter Voorspoels and Gert Storm. 2008. Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, 40 (4): 1030-1048.

Christiane Fellbaum. 1998a. *WordNet. An electronic lexical database*. The MIT Press. Cambridge, MA.

Christiane Fellbaum. 1998b. A Semantic Network of English Verbs. In C. Fellbaum (ed.) *WordNet. An electronic lexical database.* The MIT Press. Cambridge, MA: 69-104.

Peter Garrard, Matthew A. Lambon Ralph, John R. Hodges and Karalyn Patterson. 2001. Prototypicality, distinctiveness and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology* 18: 125-174.

George E. Hughes and Max J. Cresswell. 1996. A New Introduction to Modal Logic. Routledge. London and New York.

Rubén Izquierdo, Armando Suárez and German Rigau. 2007. Exploring the automatic selection of basic level concepts. *Proceedings of RANL 2007*: 298-302.

Gerhard Kremer and Marco Baroni. 2011. A set of semantic norms for German and Italian. *Behavior Research Methods*, 43 (1): 97-109.

Gianluca E. Lebani and Emanuele Pianta. 2010a. A Feature Type Classification for Therapeutic Purposes: a preliminary evaluation with non expert speakers. *Proceedings of ACL-LAW IV*: 157-161.

Gianluca E. Lebani and Emanuele Pianta. 2010b. Exploiting Lexical Resources for Therapeutic Purposes: the Case of WordNet and STaRS.sys. *Proceedings of COLING-CogALex-II*: 94-101.

Ken McRae, George S. Cree, Mark S. Seidenberg and Chris McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments & Computers*, 37 (4): 547-559.

George A. Miller. 1998. Nouns in WordNet. In C. Fellbaum (ed.) *WordNet. An electronic lexical database*. The MIT Press. Cambridge, MA: 23-46.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross and Katherine J. Miller. 1990. Introduction to WordNet. An On-line Lexical Database. *International Journal of Lexicography*, 3 (4): 235-244.

Gregory L. Murphy. 2002. *The big book of concepts*. The MIT Press, Cambridge, MA.

Sonya Nikolova, Jordan Boyd-Graber and Perry R. Cook. 2009. The design of ViVA: a mixed-initiative visual vocabulary for aphasia. *Proceedings of CHI'09*: 415-420.

Sonya S. Nikolova, Jordan Boyd-Graber and Christiane Fellbaum. 2011. Collecting semantic similarity ratings to connect concepts in assistive communication tools. In A. Mehler et al (eds.) *Modeling, Learning and Processing of Text Technological Data Structures*. Springer. Berlin and Heidelberg:81-93.

Brian Petheram. 2004. Special Issue on Computers and Aphasia. *Aphasiology*, 18 (3): 187-282.

Emanuele Pianta, Luisa Bentivogli and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. *Proceedings of GWC 2002*: 292-302.

Dan Tufiş. 2004 ed. Special Issue on BalkaNet. *Romanian Journal on Information Science and Technology*, 7 (1,2): 1-248.

David P. Vinson and Gabriella Vigliocco. 2008. Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40 (1): 183-190.