The Role of Common-Sense Knowledge in Assessing Semantic Association

(Article begins on next page)

27 April 2024

The Role of Common-Sense Knowledge in Assessing Semantic Association

- PREPRINT version -

Alice Ruggeri · Luigi Di Caro · Guido Boella

Abstract Natural Language Processing techniques often aim at automatically extracting semantics from texts. However, they usually need some available semantic knowledge contained in dictionaries and resources such as WordNet, Wikipedia, FrameNet, and so forth. In this respect, there is a large literature about the creation of novel semantic resources as well as attempts to integrate existing ones. In this context, we here focus on *common-sense* knowledge, which shows to have interesting characteristics as well as challenging issues such as ambiguity, vagueness, and inconsistency. In this paper, we make use of a large-scale and crowdsourced common-sense knowledge base, i.e., ConceptNet, to qualitatively evaluate its role in the perception of semantic association among words. We then propose an unsupervised method to disambiguate and integrate ConceptNet instances into WordNet, demonstrating how the enriched resource improves the recognition of semantic association. Finally, we describe a novel approach to label semantically-associated words by exploiting the functional and behavioural information usually contained in common-sense, demonstrating how this enhances the explanation (and the use) of relatedness and similarity with non-numeric information.

Corresponding author: Luigi Di Caro
University of Turin
E-mail: dicaro@di.unito.it

1 Introduction

One of the dreams of Artificial Intelligence (AI) is the integration of conceptual and behavioural knowledge in machines so as to bridge the gap between humans and computers in solving problems. In particular, natural language represents a fundamental channel and type of data which provides many challenges for AI systems. Understanding language is for computers one of the hardest task due to the huge space of complexity underlying languages features, from lexical ambiguities to sentence- and discourse-level modeling of meaning.

Lexical semantic resources are nowadays the key for most intelligent and advanced processing of language. The reason relies on the paradox that semantic extraction often needs existing semantics as input knowledge. However, there exists plenty of types of semantics that can be encoded in knowledge bases, each one answering specific needs or supporting specific language processing tasks, such as computational lexicons (e.g., [22,24]), frames (e.g., [10]), common-sense knowledge (e.g., [21,31]), geometric approaches (e.g., [8, 12]), probabilistic models (e.g., [17,1]), and many others.

In this paper, we investigate the role and the impact of common-sense knowledge in the context of semantic association between lexicalized concepts. Semantic association may be seen as the most generic semantic connection between words which can in turn subsume simple relatedness (e.g., *baby* vs *cradle*), rather than similarity (e.g., *smart* vs *intelligent*). In detail, the aim of this contribution is threefold:

1. to discover how humans rely on common-sense for assessing semantic association between highly-associated words (taken from manually-annotated resources) at different levels of analysis;
2. to propose a collaborative filtering-based approach to align common-sense facts to WordNet, a well-known electronic dictionary of word senses, evaluating the resulting enrichment on a semantic association classification experiment;
3. to propose a common sense-based method to automatically label semantically-associated words with *semantic explanations* which can better identify the origin of the association.

This paper extends and integrates preliminary works on different aspects such as the perception of similarity with lexical entities [26] and the disambiguation of ConceptNet instances [4]. In particular:

– we integrated and used the cognitive experiment in [26] as a way to motivate the need of CSK in a computational approach to measure and label semantic association. Further experiments have been made and all experimental data are now released;
– we extended the work presented in [4] for the alignment of ConceptNet instances to WordNet synsets and revisited as a collaborative filtering approach, simplifying its description;

– we propose a new method for the labeling of semantically-associated words that improves state-of-the-art systems in identifying the origin of the association;

This paper also represents a significant work of integration of individual modules to form a unique and novel scientific contribution.

2 Research Questions

While the Introduction reports the general context and idea behind the work, in this section we detail the three specific research questions addressed by our contribution.

2.1 Understanding the role of CSK for Assessing Semantic Association

The main characteristic of CSK with respect to classic semantic resources such as WordNet [22] and BabelNet [24] is the kind of represented information. While WordNet-like resources are usually focused on the definition of concepts in terms of descriptions (or glosses) and taxonomical structures, CSKs also contain world knowledge about functional and behavioural facts. In the light of this, we formulated the following research question:

> [**RQ1**] *Is functional/behavioural information (which is often the kind of information contained in CSK) relevant in the perception of meaning and semantic association? And if so, how?*

We tried to answer this question through a cognitive experiment where 4 participants were asked to evaluate the semantic association between 24 cases of lexicalized concept pairs with respect to taxonomic/encyclopedic (i.e., computational lexicons data) rather than functional/behavioural information (i.e., common-sense knowledge). In particular, we found that functional / behavioural facts strongly influence the perception of meaning even when dealing with textual data.

2.2 Integrating CSK into Computational Lexicons

From the results of the cognitive experiment, we then focused our attention to the integration of CSK into WordNet to test whether common-sense facts may improve an automatic approach to classify semantic association. The research question may be summed up as follows:

> [**RQ2**] *Is it possible to integrate CSK into WordNet? And if so, to what extent? Does such integration improve the accuracy of a semantic association classifier?*

We present an approach inspired by a collaborative-filtering method that allowed us to disambiguate and inject around 600,000 common-sense facts into WordNet synsets. This knowledge has been used in a supervised scenario to demonstrate the improvement of the accuracy levels in the classification of semantic association between words.

2.3 Labeling Semantic Association with CSK

Finally, we started from the concept of Distributional Semantics (DS), which represents a recent and successful approach in Computational Linguistics to see *meanings* of words as distributional profiles over linguistic contexts and to compute semantic association scores by making comparisons over such distributional information. DS, originally inspired by the Distributional Hypothesis [14], has been demonstrated to be very practical and useful, but it has also shown to have intrinsic limitations in the way it deals with the explanation of the nature of semantic association, which can be further used to address formal semantics aspects such as compositionality, quantification and inference. In the light of this, we formulated the following research question:

> [**RQ3**] *May the use of common-sense knowledge help label and explain semantic association?*

We tried to answer this question by proposing a novel method for aligning words on the basis of common-sense facts that improves standard techniques for automatic labeling of semantically-associated words.

3 Assessing Common-Sense Semantic Association

In this section, we present the first contribution of the paper which aims at understanding the role of conceptual rather than behavioral/functional features in the perception of the similarity between words through a cognitive experiment. This part represents the answer for research question RQ1.

3.1 Definitions

Before going through the details of the experiment, we here define the meaning of two aspects which are central in this work: conceptual/encyclopedic rather than behavioral/functional information that can be ascribed to entities. Entities can be either objects, living entities, and abstract concepts.

– *Conceptual.* A conceptual or encyclopedic information is a semantic fact which defines what a concept is. Examples are physical properties such as color, shape, form, and size, but also distinctive properties, components, locations, etc.

– *Functional*. A behavioral or functional information is a semantic fact which is related to actions. Examples are what one can do with an entity, what an entity may do, motivations, prerequisites, goals, expectations, etc.

This vision is directly inspired by studies in Cognitive Science, mainly derived by the concept of *affordances*, introduced by Gibson in 1977 [13], discussing about the existing link between the perceptual aspect of objects and their use in the real world.

3.2 Setting of the Cognitive Experiments

The aim of the following experiments is the understanding of the degree of importance of the above-defined *conceptual* and *functional* features in the perception of semantic association between lexicalized concepts. While there exists a large (and different) literature about "perception" and "association" in the physical world, we here describe a study related to the processing of natural language, and in particular on single lexicalized concepts.

In particular, we designed the test as a comparison between two types of semantically-correlated word pairs, the first one involving *conceptual*-related words and the other one with words linked by *functional* aspects. For example, let us consider the word pairs $A = (cloud, sky)$ and $B = (cloud, rain)$. While A contains two different concepts which are only highly correlated semantically, word pair B clearly entails an action between the two concepts. To generalize, let us consider the words a, b, and c with the *conceptual* word pair a-b and the *functional* word pair a-c. The user was asked to mark the most semantically-associated word (among b and c) to associate with a, and so the most correlated word pairs. The users were not aware of the goal of the test and of the above-mentioned difference between the selected *conceptual* vs *functional* word pairs.

At this point, our experiment required the participants to choose those word pairs they felt to be more correlated. Any significant trend in the results may be taken as an empiric evidence of how the semantic association between lexicalized concepts is influenced by *conceptual* rather than *functional* features, in a controlled dataset with comparable scores of generic semantic association, as currently approached by standard numeric evaluations of relatedness.

The experiment has been presented to 4 participants, having different ages and professions, without cognitive/linguistic disorders. Words have been given to the participants without any disambiguation, since this process is embodied in the human cognition and it has been left to the participants to autonomously decided the sense to associate with the words under comparison.

To let the empirical results of the experiment more accurate and reliable as possible, we put particular attention to the choice of the word pairs, according to the following principles (which are similarly defined in the MRC Psycholinguistic norms [33]) :

– **Conceptual granularity.** If we think at the words *object* and *thing*, we probably do not have enough information to make significant comparisons

due to their large and undefined conceptual boundaries. The same happens in cases when two words represent very specific concepts such as *lactose* and *amino acid*. The word pairs of the proposed test have been selected by considering this constraint (and so they include words which are not too specific nor too general). This principle could be considered similar to the one proposed by [33] called *familiarity*.

– **Concreteness.** Words may have direct links with concrete objects such as "*table*" and "*dog*". In other cases, words such as "*justice*" and "*thought*" represent abstract concepts. Since the literature does not present studies on the impact of concreteness/abstract on the perception of concepts and their similarity through text processing, we decided to focus our analysis on *concrete* concepts only. This principle is intended as in [33].

– **Semantic coherence.** Another criterion used for the selection of the words has been the level of semantic association between the word pairs under comparison. To better analyze whether the functional aspect plays a significant role in the perception of semantic association, we extracted *conceptual* and *functional* word pairs with similar scores of semantic association. We wanted to be sure that the users selection of one among the two was influenced only by the nature of the association (conceptual rather than functional), and not by its strength. Without this constraint, a case with a highly-related word pair A and a poorly-related word pair B would have been likely to produce A-centered user preferences only. Contrariwise, by having similar semantic correlations both in A and B, the users preference can principally reflect the perception of the association. In the light of this, we used a Latent Semantic Space calculated over almost 1 million of documents coming from the collection of literary text contained in the project Gutenberg page[1]. The selected word pairs that the participants were asked to compare had the property of having a very close semantic relatedness score (i.e., less than 0.1 of difference in a [0,1] range). This means that each *conceptual* $a-b$ pair with each *functional* $a-c$ pair in the controlled dataset is complaint with the following constraint: $|relatedness(a, b) - relatedness(a, c)| < 0.1$.

Finally, we then considered distinct degrees of importance of the *functional* aspect. Our assumption is that the relative importance of the *functional* link between two concepts has a direct link with its perceived association value (and thus with the preferences of the participants in contrast with the *conceptual* word pairs). For this reason we added a final criterion for the selection of the dataset:

– **Increasing level of the *functional* aspect.** To estimate the importance of the functional aspect that relates two lexicalized concepts we analyzed the number of actions (or verbs) in which they are usually involved with.

---

[1] http://promo.net/pg/

For each word pair (a, b), we counted the number K of existing verbs connecting a with b and the number N of verbs involving only a. Then, we calculated the ratio K/N as an approximation of the relevance of word b within the whole functional roles played by a. For example, considering the *functional* word pairs *salt-water*, *nail-polish*, and *ring-finger*, they have a functional ratio of 0.003^2, 0.0101^3, and 0.0625^4 respectively, thus representing three different degrees of importance of the functional aspects. To complete the controlled dataset, we selected and clustered word pairs according to these values, forming three groups: *low*, *medium*, and *high* scores of the functional ratio. This additional constraint has been introduced to measure the linear impact of the functional link in the perception of semantic association.

The selection process started from the identification of all word senses in WordNet having a depth in the taxonomy between 4 and 12, discarding top-level and bottom-level concepts such as for example *entity* and *limousine* (*conceptual granularity criterion*). Then, we iterated the following procedure: 1) random selection of a word sense S_a; 2) retrieval of pairs $<< S_a, S_b >$, $< S_a, S_c >$ having similar scores of semantic association (*semantic coherence criterion*); 3) computation of their functional score (*functional aspect criterion*); 4) manual validation of the resulting pairs according to the *concreteness criterion*.

3.3 Results

The results of the association experiment is shown in Table 1 along with all the choices made by the participants in the test.

The need of *functional* features for capturing human-like perceived meaning of the concepts is made evident by the proposed experiment, as these aspects resulted to play an important role for the assessment of semantic association. In particular, some concepts have shown a stronger associative perceptive relation with respect to others ones in those cases they were linked by significant *functional* features.

In addition to this, the experiment indicates a correlation between the obtained association scores and the exclusive verbal phrases connecting the concepts, i.e., the higher the calculated *functional* level, the higher the obtained *functional* preferences. We tried to approximate this concept by looking at noun-verb co-occurrences, but this should deserve further work in future research.

Finally, the fact that low-*functional* and medium-*functional* word pairs show minor preferences with respect to high-*functional* pairs is in line with

---

[2] For the verbs "*to put*", "*to add*" and "*to get*"

[3] For the verbs "*to apply*" and "*to use*"

[4] For the verbs "*to put*" and "*to wear*"

what stated by [7], i.e., words that have a functionality-based relationship may have a more complex visual component that makes such correlation weaker.

A *functional* type of semantic knowledge can be typically found in Common Sense-based resources such as ConceptNet. For this reason, we proposed a first method for integrating ConceptNet into WordNet in order to improve a classic natural language processing task: the evaluation of semantic relatedness. The next sections will detail this further contribution.

Table 1 Results on the word pairs association experiment with 24 different cases (8 low-functional level, 8 medium-functional level and 8 high-functional level). The last column represents the percentage of the times participants have chosen the functional word pair (BF) rather than the conceptual one (CE) in the test, according to each functional level. Each case has been presented to 4 people (answers A1, A2, A3 and A4 in the table).

BF level	CE word pair		BF word pair		A1	A2	A3	A4	BF Preference
low	salt	sugar	salt	water	CE	CE	CE	CE	**28%**
	table	chair	table	food	BF	CE	CE	CE	
	beach	island	beach	holiday	BF	CE	BF	BF	
	book	text	book	knowledge	BF	CE	CE	BF	
	cloud	sky	cloud	rain	CE	BF	BF	CE	
	wine	liquor	wine	bottle	CE	CE	CE	CE	
	bed	couch	bed	hospital	CE	CE	CE	CE	
	story	topic	bed	kid	BF	CE	CE	CE	
medium	bread	pasta	bread	kitchen	BF	CE	BF	CE	**41%**
	car	vehicle	car	horne	CE	CE	CE	BF	
	band	singer	band	music	CE	BF	BF	BF	
	feet	hands	feet	stairs	BF	BF	BF	BF	
	flowers	leaves	flowers	apes	CE	BF	CE	CE	
	mouse	cat	mouse	cheese	BF	CE	BF	CE	
	insect	bee	insect	plant	CE	CE	CE	CE	
	nail	polish	finger	polish	CE	CE	CE	CE	
high	mouth	tooth	mouth	voice	BF	BF	BF	BF	**91%**
	eye	nose	eye	tear	BF	BF	BF	BF	
	ring	necklace	ring	finger	BF	BF	BF	BF	
	door	gate	door	handle	BF	BF	CE	BF	
	bottle	glass	bottle	cap	CE	BF	BF	CE	
	broom	rag	broom	dust	BF	BF	BF	BF	
	scissors	knife	scissors	paper	BF	BF	BF	BF	
	bomb	dynamite	bomb	blast	BF	BF	BF	BF	

4 Measuring Semantic Association through CSK Integration

In this section, we answer to the research question **RQ2**. In particular, we propose a novel approach for the alignment of linguistic and common-sense semantics based on the exploitation of their intrinsic characteristics: while the former represents a reliable (but strict in terms of semantic scope) knowledge, the latter contains an incredible wide but ambiguous set of semantic information. In the light of this, we assigned the role of *hinge* to WordNet, that guides a trusty, multiple and simultaneous retrieval of data from ConceptNet which are then intersected with themselves through a set of mechanisms to produce automatically-disambiguated knowledge. ConceptNet [30] is a semantic graph that has been directly created from the Open Mind Common Sense[5] project developed by MIT, which collected unstructured common-sense knowledge by asking people to contribute over the Web.

4.1 Basic Idea

The idea of the proposed enrichment approach relies on a fundamental principle, which makes it novel and more robust with respect to the state of the art. Indeed, our extension is not based on a similarity computation between words for the estimation of correct alignments. On the contrary, it aims at enriching WordNet with semantics containing direct relations- and words overlapping, preventing associations of semantic knowledge on the unique basis of similarity scores (which may be also dependent on algorithms, similarity measures, and training corpora). This point makes this proposal completely different from what proposed by [6], where the authors created word sense profiles to compare with ConceptNet terms using semantic similarity metrics. Section 5 further describes and deepens this type of non-numeric similarity calculation. Finally, graph-based alignments of lexical resources such as BabelNet [25] can not be directly applied on ConceptNet due to the specificity of the method on the nature of the Wikipedia resource and its features (internal links, categories, inter-language links, disambiguation pages, and redirect pages). ConceptNet instead presents a much weaker set of contextual features to be used for the semantic alignment.

4.2 Definitions

Let us consider a WordNet word sense

$$S_i = < T_i, g_i, E_i >$$

where T_i is the set of synonym terms $t_1, t_2, ..., t_k$ while g_i and E_i represent its gloss and the available examples respectively. Each word sense represents a

---

[5] http://commons.media.mit.edu/

meaning ascribed to the terms in T_i in a specific context (described by g_i and E_i). Then, for each sense S_i we consider a set of semantic properties

$$P_{wordnet}(S_i)$$

coming from the structure around S_i in WordNet. For example, $hypernym(S_i)$ represents the direct hypernym word sense while $meronyms(S_i)$ is the set of senses which compose (in terms of a *made-of* relation) the concept represented by S_i. The above-mentioned complete set of semantic properties $P_{wordnet}(S_i)$ of a sense S_i contains a set of pairs

$$< rel - word >$$

where rel is the relation of S_i with the other senses (e.g., *is-a*) and $word$ is one of the lemmas of such linked senses. For example, given the word sense $S_{cat}\#1$:

$S_{cat}\#1$: *cat, true cat (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)*,

a resulting $< rel - word >$ pair that comes from $hypernym(S_{cat})$ will be:

$$< isA - feline >$$

since $feline$ is a lemma of the hypernym word sense $S_{feline,felid}\#1$:

$S_{feline,felid}\#1$: *feline, felid (any of various lithe-bodied roundheaded fissiped mammals, many with retractile claws)*.

Note that in case of multiple synonym words in the related synsets, there will be multiple $< rel - word >$ pairs. Then, ConceptNet can be seen as a large set of semantic triples of the form

$$NP_i - rel_k - NP_j$$

where NP_i and NP_j are simply non-disambiguated noun phrases whereas rel_k is one of the semantic relationships in ConceptNet.

4.3 Method

The solution for aligning ConceptNet triples with WordNet synsets is described in this section. Initially, for each synset S_i, we compute the set of all candidate semantic ConceptNet triples

$$P_{conceptnet}(S_i)$$

as the union of the triples that contain at least one of the terms in T_i. The inner cycle iterates over the candidate triples to identify those that can enrich the synset under consideration. We proposed 4 mechanisms to align each ConceptNet triple c_k (of the form $NP - rel - NP$) to each synset S_i. The first one, i.e., *gloss-based*, can be viewed as a baseline approach that only considers a matching ConceptNet term within a WordNet synset gloss.

Gloss-based (baseline) [*Condition*: If a lemma of an NP of the triple c_k is contained in the lemmatized gloss g_i of the synset S_i]. This would mean that ConceptNet contains a relation between a term in T_i and a term in the description g_i, making explicit some semantics contained in the gloss. Note that the systematic inclusion of *related-to* relations with all the terms in the gloss g_i would carry to many incorrect enrichments, so this mechanism is necessary to identify only correct alignments.

Structure-based [*Condition*: If a lemma of an NP of c_k is also contained in $P_{wordnet}$]. By traversing the WordNet structure, it is possible to link words of related synsets to S_i by exploiting existing semantics in ConceptNet.

Gloss^2-based [*Condition*: If a lemma of an NP of c_k is contained in the lemmatized glosses of the most probable synset associated to the words in g_i]. The word sense disambiguation algorithm used for disambiguating the text of g_i is a simple match between the words in the triple with the words in the glosses. In case of empty intersections, the most frequent sense is selected.

Collaborative Filtering . After taking all the hypernyms of S_i, we queried ConceptNet with their lemmas obtaining different sets of triples (one for each hypernym lemma). [*Condition*: If the final part $* - rel - word$ of the triple c_k is also contained in one of these sets]. We associate c_k to S_i.

The idea is to intersect different sets of ambiguous common-sense knowledge to make a sort of hypernym-based collaborative filtering of the triples. For example, let S_i be

$$S_{burn,burning} : \textit{pain that feels hot as if it were on fire}$$

and the two candidate ConceptNet triples

$$c_1 = burning - relatedto - suffer$$

and

$$c_2 = burn - relatedto - melt.$$

Once retrieved

$$hypernyms(S_{burn,burning}) = \{pain, hurting\}$$

from WordNet, we query ConceptNet with both *pain* and *hurting*, obtaining two resulting sets

$$P_{conceptnet}(pain)$$

and

$$P_{conceptnet}(hurting).$$

Given that the end of the candidate triple c_1 is contained in $P_{conceptnet}(pain)$, the triple is added to synset $S_{burn,burning}$. On the contrary, the triple c_2 is not added to $S_{burn,burning}$ since $relatedto - melt$ is not contained neither in $P_{conceptnet}(pain)$ and $P_{conceptnet}(hurting)$.

Fig. 1 Example of resulting WordNet+ConceptNet alignment on a standard lexical type of knowledge (*partOf* relation on the concept *car*).

4.4 Results and Evaluation

The proposed method is able to link (and disambiguate) a total of 98,122 individual ConceptNet instances to 102,055 WordNet synsets. Note that a single ConceptNet instance is sometimes mapped to more than one synset (e.g., the semantic relation *hasproperty-red* has been added to multiple synsets such as [*pomegranate, …*] and [*pepper, …*]). Therefore, the total number of ConceptNet-to-WordNet alignments is 582,467. Note that we only kept those instances which were not present in WordNet (i.e., we removed redundant relations from the output).

To have an idea of the complementary nature of WordNet and ConceptNet, Figures 1 and 2 show the result of the proposed automatic alignment on two examples: 1) the extracted knowledge related to what can be a part of the concept *car* (standard lexical knowledge); and 2) the extraction of what can *hurt* (common-sense type of knowledge).
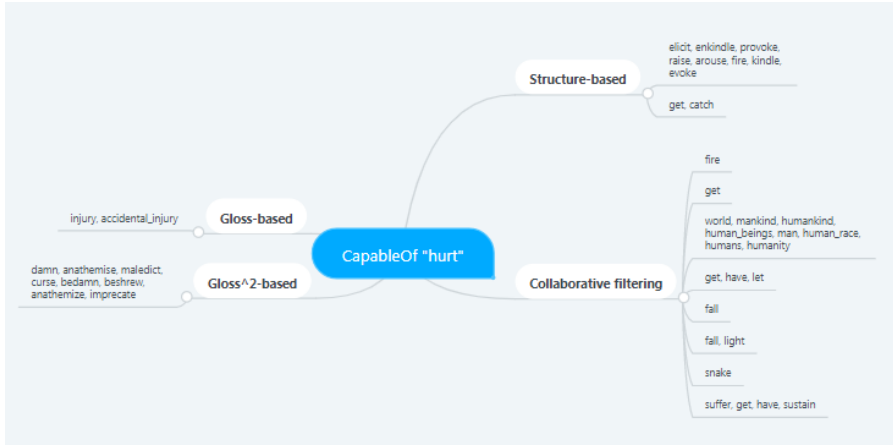
Fig. 2 Example of resulting WordNet+ConceptNet alignment on a common-sense type of lexical knowledge (*capable* relation on the concept *hurt*).

Table 4.4 shows an analytical overview of the resulting WordNet enrichment according to the used mechanisms. The baseline method was able to infer less than 40% of the total extracted alignments.

Mechanism	# of Alignments
Gloss-based (baseline)	222,544
Structure-based	109,212
Gloss^2-based	19,769
Collaborative Filtering	230,942

Table 2 Overview of the WordNet-to-ConceptNet alignments according to the used mechanisms. The baseline method was able to infer less than 40% of the total extracted alignments.

In order to obtain a first and indicative evaluation of the approach, we manually annotated a set of 505 randomly-picked individual synset enrichments. In detail, given a random synset S_i which has been enriched with at least one ConceptNet triple $c_k = < NP - rel - NP >$, we verified the semantic correctness of c_k when added to the meaning expressed by S_i, considering the synonym words in T_i as well as its gloss g_i and examples E_i. Table 4.4 shows the results.

The manual validation revealed a high accuracy of the automatic enrichment. While the total accuracy is 88.31% (note that higher levels of accuracy are generally difficult to reach even by inter-annotation agreements), the extension seems to be highly accurate for relations such as *capable-of* and *has-property*. On the contrary, *is-a* and *related-to* relations have shown a lower performance (around 85%). However, this is in line with the type of used resources: on the one hand, WordNet represents a quite complete taxonomical structure of lexical entities; on the other hand, ConceptNet contains a

Relation	# correct	# incorrect	Accuracy
related-to	121	22	84.62%
is-a	99	17	85.34%
at-location	39	5	88.84%
capable-of	36	1	97.29%
has-property	29	2	93.55%
antonym	27	4	87.10%
derived-from	25	1	96.15%
...	...	...	...
Total	446	59	88.31%

very large semantic basis related to objects behaviours and properties. Finally, *related-to* relations are more easily identifiable through statistical analyisis of co-occurrences in large corpora and advanced topic modeling built on top of LSA [8], LDA [1] and others. Extending WordNet with non-disambiguated common-sense knowledge may by challenging, also considering the very limited contextual information at disposal. However, such an alignment is feasible due to the few presence of common-sense knowledge related to very specific word senses (e.g., the term "*cat*" is very improbable to have common-sense facts related to the unfrequent word sense $S_{cat}\#8$:

> $S_{cat}\#8$: *A method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis.*

4.5 Impact of the Integration

The leading idea of the paper is to contructively criticize the way semantic similarity or relatedness is calculated and used in its numerical form (both as human perception and automatic estimation).

In the first place, we asked the participants to guess scores of similarity and relatedness for already-annotated word pairs contained in the SimLex-999 corpus. 6 people participated at the test, having no information about the topics and the goals of the analysis. In detail, they have been asked to score the perceived similarity of terms in 50 randomly-selected word pairs, using the range [0,10]. Table 3 gives an overview of the results, demonstrating how, generally, people significantly disagree in assessing similarity by using numbers. This is because numbers do not encode meaning while subjective and unstable perceptions of it. Thus, regression tests are not reflecting our intention of evaluation, while clustering ranges of scores into bins for classification seemed to us the best choice.

To validate this assumption, we then further discretised the obtained scores into two bins (as in the later classification task described in the following sections). Pairs with a value lower than 5.01 have been labelled as *non-similar or lowly-similar*, whereas the others with the label *quite similar or highly-similar*). We then computed the Fleiss' kappa score [29], finding a significant

Table 3 Correlation table of the six participants (from P1 to P6) similarity scores. The average correlation value resulted to be 0.3024, demonstrating the high human disagreement in assessing similarity through numbers (and as a consequence, the low significance of evaluations of the automatic approach through regression tests).

	P1	P2	P3	P4	P5	P6	
P1	1	0.3849	0.6671	0.5573	0.5105	-0.0395	
P2	0.3849	1	0.3416	0.4650	0.1404	0.1804	
P3	0.6671	0.3416	1	0.3783	0.5743	-0.0167	
P4	0.5573	0.4650	0.3783	1	0.3572	0.0263	
P5	0.5105	0.1404	0.5743	0.3572	1	0.0089	
P6	-0.0395	0.1804	-0.0167	0.0263	0.0089	1	
Avg	**0.4161**	**0.3025**	**0.3889**	**0.3569**	**0.3183**	**0.0319**	**0.3024**

agreement of 0.419, which can be considered as a good degree of aligned classification over that which would be expected by chance. This, alone, is in contrast with the low correlation values obtained through their original numeric form.

However, we thought that discretising values which have been acquired through numerical assessments could potentially capture different behaviours w.r.t. direct requests for categorical labels. In this respect, we asked the same participants[6] to re-annotate the same word pairs using a binary decision (*non-similar/lowly-similar* or *quite similar/highly-similar*). What we found is that the agreement on such second-stage evaluation with direct categorical assessments reached a Fleiss' kappa of 0.616. This can be interpreted as a better perception and ease of decision over categories against numerical judgements.

In the light of these experiments, the next sections will describe a supervised scenario for the automatic classification of semantic similarity of input word pairs. In the first test, we only used WordNet and ConceptNet separately, without enrichment. Then, we experimented with their conjunctive use (without common-sense disambiguation) and finally with a novel proposed integration method.

4.5.1 Methodology

To validate the impact of CSK in our computational experiments, we made use of a simple methodology which replaces target words with available semantic information. The idea is to evaluate the informativeness of the semantic features by testing their discriminatory power in a supervised setting. In other words, we asked a standard Machine Learning method to learn from semantically-associated words which are the relevant features and if they can generalize over unseen word pairs.

4.5.2 Description of the Experiment

The experiment starts from the transformation of a word-word-score similarity dataset into a context-based dataset in which the words are replaced by sets

---

[6] Note that this has been done after six months since their first annotation task, without having knowledge about partial results of the work.

of semantic information taken from ConceptNet. The aim is to understand if common-sense knowledge may represent a useful basis for capturing the similarity between words, and if it may outperform systems based on linguistic resources such as WordNet.

4.5.3 Data

We used two similarity datasets: SimLex-999 [16] and MEN [3]. The former contains one thousand word pairs that were manually annotated with similarity scores. The inter-annotation agreement was 0.67 (Spearman correlation). The MEN dataset contains 3,000 pairs of randomly selected words scored on a [0,50]-normalized semantic relatedness scale via ratings obtained by crowd-sourcing on the Amazon Mechanical Turk.

As described in Section 4.5.1 we leveraged (1) Wordnet, (2) ConceptNet (3) their union, and (4) our proposed integration (through disambiguation and alignment) to build the sets of semantic features to replace to the words of each pair.

4.5.4 Running Example

Let us consider the pair *rice-bean*. ConceptNet returns the following set of semantic information for the term *rice*:

> [hasproperty-edible, isa-starch, memberof-oryza, atlocation-refrigerator, usedfor-survival, **atlocation-atgrocerystore**, isa-food, relatedto-grain, madeof-sake, isa-grain, isa-traditionally, receivesaction-eatfromdish, isa-often, **usedfor-cook**, relatedto-kimchi, **atlocation-pantry**, atlocation-ricecrisp, **isa-domesticateplant**, relatedto-sidedish, atlocation-supermarket, receivesaction-stir, isa-staplefoodinorient, hasproperty-cookbeforeitbeeat, madeof-ricegrain, partof-cultivaterice, **receivesaction-eat**, derivedfrom-rice, isa-cereal, relatedto-white, hasproperty-white, **isa-vegetable** hascontext-cook, relatedto-whitegrain, **relatedto-food**]

Then, the semantic information for the word *bean* are:

> [usedfor-fillbeanbagchair, atlocation-infield, atlocation-can, usedfor-nutrition, **usedfor-cook**, memberof-leguminosae, usedfor-makefurniture, usedfor-grow, atlocation-foodstore, isa-legume, usedfor-count, hasproperty-easytogrow, partof-bean, **receivesaction-eat** atlocation-cookpot, isa-vegetableorpersonbrain, atlocation-beansoup, **isa-domesticateplant**, atlocation-soup, **atlocation-pantry**, usedfor-plant, **isa-vegetable**, atlocation-container, usedfor-supplyprotein, atlocation-jar, **atlocation-atgrocerystore**, usedfor-useasmarker, atlocation-field, derivedfrom-beanball, atlocation-coffee, usedfor-fillbag, **relatedto-food** receivesaction-grindinthis, usedfor-beanandgarlicsauce, atlocation-beancan, usedfor-makebeanbag]

Finally, the intersection produces the following set (*semantic intersection*, marked in bold in the previous individual instances):

$<semantic\ intersection>$ = [atlocation-atgrocerystore, isa-vegetable, isa-domesticateplant, atlocation-pantry, relatedto-food, usedfor-cook, receivesaction-eat]

Then, for each non-empty intersection, we created one instance of the type:

$<semantic\ intersection>,\ <similarity\ score>$

building a standard *term-document* matrix, where the *term* is a semantic term within the set of semantic intersections (e.g., usedfor-cook) and the *document* dimension represents the instances (i.e., the semantic intersection of the features of the word pairs) of the original dataset. After this preprocessing phase, the *score* attribute is discretized into two bins:

− *non-similar* class - range in the dataset [0, (max / 2)]
− *similar* class - range in the dataset ((max / 2), max]

where max represents the maximum score in the annotated similary dataset (10 for SimLex-999, 50 for MEN).

We decided to keep the task simple to only evaluate the importance of the functional aspects, without adding complexity or erroneously capturing other dynamics (related to the applied ML algorithm, for example). However, since instances with a score close to $(max/2)$ could be considered difficult cases (e.g., the word pair $<afraid,\ anxious>$ has a score of 5.07 in the [0,10] range, so it would be marked as *similar*, while $<toe,\ finger>$ as *non-similar* with a score of 4.68), we also experimented a three-bins classification:

− *non-similar* class - range in the dataset [0, (max / 3)]
− *borderline* class - range in the dataset ((max / 3), (2 * max / 3)]
− *similar* class - range in the dataset ((2 * max / 3), max]

then removing the borderline instances. In this case, the results were in line with those of Tables 7, 5, and 6 described in the next section, even if all values were higher of (on average) 0.05 points in Precision, Recall and F-Measure.

4.5.5 Results

The splitting of the data into two clusters allowed us to experiment a classic supervised classification system, where a Machine Learning tool (a Support Vector Machine, in our case) has been used to learn a binary model for automatically classifying *similar* and *non-similar* word pairs.

The result of the experiments is shown in Tables 7, 5 and 6 for Precision, Recall and F-Measure respectively. Noticeably, the use of ConceptNet alone sometimes produced better values than with Wordnet, demonstrating the potentiality of CSK even if not structured and disambiguated as in WordNet.

Note that with ConceptNet, *similar* word pairs are generally easier to identify with respect to *non-similar* ones. On the other side, WordNet resulted to have a high Precision but a very low Recall for *similar* word pairs.

Table 4 Precision values of the experiment with ConceptNet, WordNet, their combination (marked with the **+** operator), and the proposed integration (marked with the ***** operator) between the two resources.

Class	Pr. (CN)	Pr. (WN)	Pr. (WN+CN)	**Pr. (WN*CN)**
non-similar (MEN)	0.594	0.512	0.596	0.612
similar (MEN)	0.590	0.904	0.591	0.814
weighted total (MEN)	0.592	0.712	0.594	**0.715**
non-similar (SimLex)	0.610	0.579	0.598	0.650
similar (SimLex)	0.492	0.772	0.479	0.686
weighted total (SimLex)	0.558	0.664	0.545	**0.666**

Table 5 Recall values of the experiment with ConceptNet, WordNet, their combination (marked with the **+** operator), and the proposed integration (marked with the ***** operator) between the two resources.

Class	Re. (CN)	Re. (WN)	Re. (WN+CN)	**Re. (WN*CN)**
non-similar (MEN)	0.539	0.990	0.537	0.893
similar (MEN)	0.643	0.087	0.648	0.452
weighted total (MEN)	0.592	0.532	0.593	**0.669**
non-similar (SimLex)	0.543	0.977	0.522	0.848
similar (SimLex)	0.560	0.100	0.556	0.422
weighted total (SimLex)	0.551	0.590	0.536	**0.660**

Table 6 F-Measure values of the experiment with ConceptNet, WordNet, their combination (marked with the **+** operator), and the proposed integration (marked with the ***** operator) between the two resources.

Class	FM (CN)	FM (WN)	FM (WN+CN)	**FM (WN*CN)**
non-similar (MEN)	0.565	0.675	0.565	0.726
similar (MEN)	0.616	0.159	0.618	0.581
weighted total (MEN)	0.591	0.413	0.592	**0.653**
non-similar (SimLex)	0.574	0.727	0.557	0.736
similar (SimLex)	0.524	0.177	0.514	0.522
weighted total (SimLex)	0.552	0.484	0.538	**0.642**

Finally, the tables report a clear superiority of the Machine Learning task when applied on the proposed integration of the two semantic resources Word-Net and ConceptNet (marked with WN*CN in the tables). Note that the simple union (marked with WN+CN in the tables) between WordNet and ConceptNet features did not carry to any increase in performance. This is due to the high ambiguity of the terms in ConceptNet. We can conclude from this that ConceptNet may not by directly usable as is. Overall, this experiments demonstrate the need as well as the successful disambiguation process with its impact on the automatic calculation of semantic similarity between lexical entities.

Table 7 General view on Precision, Recall and F-Measure of the proposed integration approach.

Class	Precision	Recall	F-measure
non-similar (MEN)	0.612	0.893	0.726
similar (MEN)	0.814	0.452	0.581
weighted total (MEN)	**0.715**	**0.669**	**0.653**
non-similar (SimLex)	0.650	0.848	0.736
similar (SimLex)	0.686	0.422	0.522
weighted total (SimLex)	**0.666**	**0.660**	**0.642**

5 Labeling Semantic Association with CSK

In this section, we answer to the research question **RQ3**. In particular, we present a method that revisits the classic numeric nature of computing semantic association, as originally approached in [32]. In contrast with the work of Vyas and Pantel, our approach does not simply find explanations of similarity by traversing paths connecting concepts in existing lexical resources rather than finding recurring syntactic dependencies and terms in corpora, whereas by modeling a generic semantic association between concepts as a twofold interaction between conceptual and (enabled) functional (or common sense-like) features. To this end, the resulting alignment of the previous section provides an input for the following module. The motivation behind this contribution derives from what stated in [9], i.e., people explain similarity using semantically-related bundles of features rather than referring to, or interconnecting similar concepts. In addition to this, the explanations of [32] are not comparable with those of our system as they were found through a method based on predefined input queries with a random choice of the number of words. Instead, our proposal does not need any input parameters, as it solely relies on a reasoning process over the whole knowledge base.

First, we give an overview of the paradigm and the definition of what we call *explanation* and *Semantic Association Labeling* (SAL). Then, we illustrate our experimentation on ConceptNet and the similarity dataset SimLex (which we already used in the previous section). This allowed to create an additional semantic knowledge base that enriches the similarity dataset with conceptual features that aim at *explaining* the *reasons* behind the similarity numeric values.

5.1 Basic Idea

The proposed approach aims at giving a semantic explanation of the semantic association between two lexical entities using a semantic knowledge as support. Given two words w_1 and w_2, what is usually done by standard methods is to compare their *word semantic profiles*, i.e., contextual lexical item sets. For example, given the word *cat* and *dog*, their profiles can share terms such as

pet, *fur*, *claws* and so forth. These word profiles can be extracted from co-occurrences in large corpora and/or in available resources (using relations such as synonyms, hypernyms, meronyms, etc.). On the contrary, our idea is to replace *blind* lexical overlappings by a meaningful and coherent matching of semantic information, as explained in the methodology Section 4.5.1. Being bound to ConceptNet, we used the following semantic relations as *conceptual* information (also called *properties* in this section):

$$\{partOf, \ madeOf, \ hasA, \ definedAs, \ hasProperty\}$$

Then, we used the following relations as *functional* information (simply called *functions* from now on):

$$\{capableOf, \ usedFor, \ hasPrerequisite, \ motivatedByGoal,$$
$$desires, \ causes\}$$

5.2 Definitions

5.2.1 Word Explanation (WE)

The first step is the extraction of individual *word explanations* for both w_1 and w_2. A word explanation is a relation-based model which correlates conceptual properties to functions. Generally speaking, the idea is that the functions of a concept is directly correlated with its properties. In a sense, our assumption is that:

> there is a strong relationship (or interaction) between a conceptual property and some function of the object.

In our case, given a word w and a semantic relation r, we use a semantic resource KB to extract all the words that show the semantic instance $r - w$. For example, if $r = has$ and $w = fur$, the query would be $has - fur$, and the result will contain the set of words having that semantic information in KB, for example $\{cat, bear, ...\}$. Then, we obtain all semantics related to such retrieved terms, building a matrix which correlates the *conceptual* properties to the functional features (by using Pointwise Mutual Information, as later described). For example, the property of *having claws* usually correlates with the fact of *climbing trees*.

The set of matrices $M = < M_1, M_2, ..., M_{|r|} >$ (one for each semantic relation provided by KB) represent the semantic explanation e of a single word w, i.e., all *conceptual-to-functional interactions* related to the semantics around a single word w (from different perspectives, by considering all types of semantic relations in KB). For instance, the analysis of the words *bear* and *cat* may lead to explanations that associate different *functional* features to the property of having claws (e.g., *climbing trees* for a cat, *killing people* for a bear). Figure 3 shows two examples of such relation-based explanation.

Anticipating the details of the next step, in case of a semantic comparison between the words *cat* and *bear*, the claws-property will be not used as an
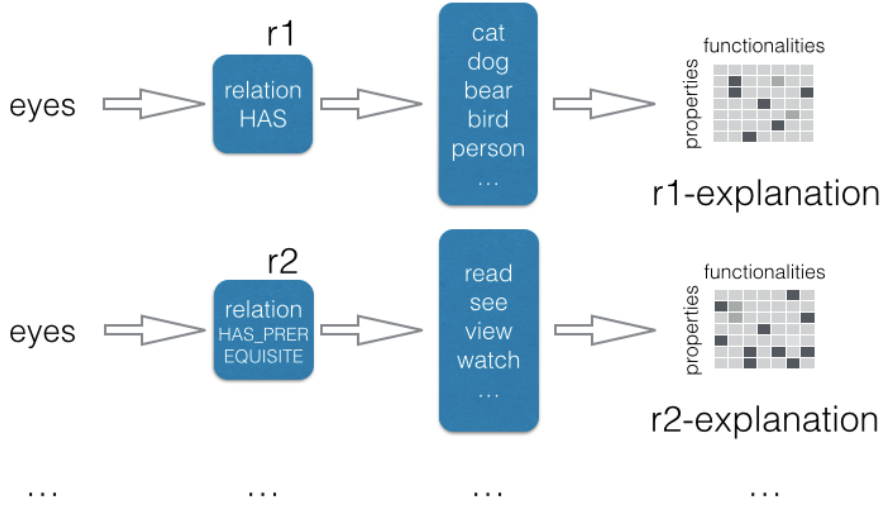
Fig. 3 Example of relation-based explanations for the word *eyes* (*cats*, *dogs*, etc. have eyes; *read*, *see*, etc. have the prerequisite of having eyes). The matrices represent the correlation between properties and functionalities of these resulting words.

element of similarity, because of their different *meanings* in terms of enabled behaviour. This point represents a radical novelty with respect to state-of-the-art approaches which only considers lexical and statistical overlappings without taking into account their actual context-based meaning.

5.2.2 Semantic Association Labeling (SAL)

For each relation r in KB, we then obtain the relative explanations for the words w_1 and w_2, namely e_1 and e_2 respectively. At this point, we make a comparison of e_1 and e_2 by aligning the vectors of the relative sets of matrices M_w1 and M_w2 (i.e., each vector of the matrices of e_1 is aligned to zero or some vector of the matrices of e_2 if they represent the same property). In case of a non-empty alignment (i.e., e_1 and e_2 share some identical property), the two *conceptual* property-vectors are compared in terms of their *functional* features. Again, in case the vectors share identical functions, the numeric product of their weight represents a score (and thus the importance) of a single *semantic association labeling (SAL) instance*. Figure 4 illustrates this process.

In words, the idea is summed up with the following conceptualization:

> a SAL instance that links a row row_1 of one r_x-matrix in e_1 and a row row_2 of one matrix r_y in e_2 would mean that everything that is related with w_1 through the relation r_x has a property p with an overlapping *functional* distribution with the same *conceptual* property p of what is related with word w_2 through the relation r_y.

For example, if $w_1 = cat$, $w_2 = tree$, $r_x = partOf$ and $r_y = usedFor$, one SAL instance contains the *conceptual* property $p = claws$, since claws are parts of a cat and they are used for climbing. In a sense, this instance *explains* an aspect of the semantic association between cats and trees in terms of a *conceptual* property (the claws), saying that cats can climb the trees through this *conceptual* aspect.
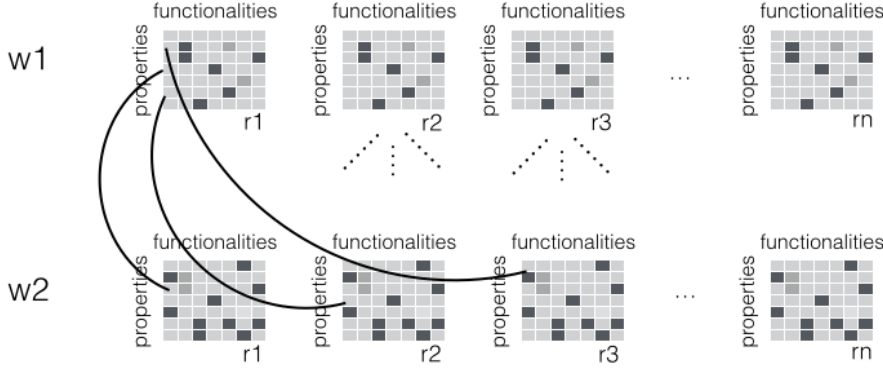

Fig. 4 Scheme of the Semantic Association Labeling (SAL) process, which correlates matrix rows (property vectors) of the explanations of the two words under comparison.

Considering the entire set of relations r in KB, the SAL of two input words w_1 and w_2 is the r-based sets of SAL instances which represent the direct matching between identical properties and their functional aspects.

5.3 Data

In this paper, we used the dataset named *SimLex-999* [16], which contains one thousand word pairs that have been manually annotated with similarity scores. The inter-annotation agreement is 0.67 (Spearman correlation), highlighting the complexity of the task (and somehow underlining the motivations of this proposal). *SimLex-999* also includes word pairs of another dataset, *WordSim-353* [11], that contains a mix of relatedness- and similarity-based items.

5.4 Algorithm

In this section, for the purpose of reproducibility, we present the details of the algorithm for the extraction of *explanations* and the final *SAL* between the input words. The pseudo-code of the entire approach is shown in Algorithm 1, 2, and 3.

Given two input words w_1 and w_2, and a semantic resource KB, the system creates a set of SAL_r instances, one for each type of relation r among the whole

set R of relations in KB. Each SAL_r instance is the result of a comparison between the two explanations of the two words according to a specific target semantic relation r. In detail, we query the semantic resource KB with the relation-word pair $r - w$. At this point, we query KB with each of these words, and collect the co-occurrence values between conceptual and functional information. In particular, we build a matrix M of $nr = |P|$ rows and $nc = |F|$ columns, where P and F are respectively the set of property features and the set of functional features, and where each value $M_{i,j}$ contains the co-occurrence of the property p_i with the functionality f_j (with $0 < i < |P|$, and $0 < j < |F|$).

Once the matrix of co-occurrences M is calculated, it is then transformed in a PMI-based matrix where each value $M_{i,j}$ is replaced with:

$$M'_{i,j} = \frac{P_{i,j}}{P_i * P_j}$$

where $P_{i,j}$ is the probability of having a non-zero co-occurrence value for the property p_i and the functionality f_j (that is, $M_{i,j} > 0$) in the semantics of the input word, while P_i and P_j are the individual probability to find the property p_i and the functionality f_j respectively. The utility of M' is to capture the strength of the associations between properties and functionalities also considering their individual frequency. Each horizontal vector of a matrix M_r represents a word explanation, i.e., how a property is related to some functionalities with respect to the considered semantic information related to $r - w$. Finally, we align the explanations of w_1 with the ones of w_2. Given a property vector of M'_{w1}, if the property is also contained in M'_{w2}, we calculate their matching functionalities. If the matching is not empty, we add the SAL_r instance in the final result. At the end of the process, the whole SAL output of the two input words w_1 and w_2 will be the set of SAL_r instances obtained for all the relations r in R.

Algorithm 1: Main method to return the final SAL output.

Data: word w_1, word w_2, semantic resource KB
Result: A set of SAL instances, given two input words and a
        semantic resource (ConceptNet in our case).
R = set of relations in KB;
result = empty set of relation-based SAL instances;
for *each relation r in R* **do**
    Explanation e_1 = getExplanation(w_1, KB, r);
    Explanation e_2 = getExplanation(w_2, KB, r);
    SemanticAssociationLabeling SAL_r = getSAL(e_1, e_2);
    result.add(r, SAL_r);
end
return result;

Algorithm 2: Method *getExplanation* for extracting explanations from an input word and a semantic resource.

Data: word w, relation r, semantic resource KB
Result: A set of explanations for a given word.
$context_{words}$ = query(KB, r, w);
for *each word cw in $context_{words}$* **do**

    $cw_{semantics}$ = query(KB, cw);
    cw_{CE} = selectCESemantics($cw_{semantics}$);
    cw_{BF} = selectBFSemantics($cw_{semantics}$);
    updateCoOccurrence(cw_{CE}, cw_{BF});

end
return the PMI values calculated over the co-occurrences;

Algorithm 3: Method *getSAL* for comparing explanations of two words, returning a SAL instance.

Data: Explanation e_1, Explanation e_2
Result: A SAL instance.
$p - vectors_1$ = getPMIValues(e_1);
$p - vectors_2$ = getPMIValues(e_2);
for *each CE property vector pw_1 (of property p) in $p - vectors_1$* **do**

    if *$p - vectors_2$ has CE property vector pw_2 of the same CE property p* **then**

        $matching_{BF}$ = match(pw_1, pw_2);
        if *$matching_{BF}$ is not empty* **then**
            instance.add(p, $matching_{BF}$)
        end

    end

end
return the instance;

5.5 Results

In this section, we present a running example to show the richness of a SAL process compared with a standard model for labeling semantic association based on co-occurrences or knowledge-base intersections. In particular, we selected a set of word pairs from a manually-annotated dataset with a various degree of similarity from the used similarity dataset. The goal was twofold to evaluate the ability to identify the key semantic points (SAL instances) of annotated word pairs. Figure 5 shows some examples of SAL instances, while the complete set of SAL-enriched word pairs for the SimLex dataset will be released, creating a new benchmark for further computational studies on these topics.

In order to evaluate the ability of the proposed system to also identify plausible semantic association explanations even in case of dissimilar word

Table 8 The 8 word pairs that have been resulted to be the most difficult to explain (in terms of their similarity) by the 8 participants in the test.

Word pair		Relation A	Relation B	Similarity	SAL Manual Validat. Accuracy
loop	belt	RelatedTo	RelatedTo	3.1	5 out of 9 items
car	elevator	RelatedTo	RelatedTo	1.03	18 out of 22 items
diet	apple	RelatedTo	RelatedTo	1.18	6 out of 7 items
water	salt	HasA	HasA	1.3	27 out of 30 items
foot	head	PartOf	PartOf	2.3	5 out of 6 items
room	bed	HasA	HasA	2.35	16 out of 30 items
motor	boat	RelatedTo	RelatedTo	2.57	17 out of 18 items
friend	teacher	IsA	IsA	2.62	8 out of 9 items
				Avg Acc.	**77.33%**

pairs (i.e., the 239 noun word pairs scored with a value lower than 3.3 in the SimLex-999 dataset), we conducted a two-phases experiment with 8 participants, asking

1. to mark the word pairs with a degree of complexity (*low, medium, high*) in terms of explanation of similarity. Specifically, they were asked to think how difficult was to think at labels explaining the similarity between the two terms in the pair;
2. we selected those (8) cases that received the highest number of *high*-marks and asked again the same participants to manually evaluate the automatically extracted SALs for such difficult cases.

Table 8 shows the 8 word pairs which seemed to be the most difficult to explain by the participants. The automatically extracted labels have been manually marked as simply *correct* or *incorrect*. The results show that more than 3 out of 4 times the presented labels have been recognized as right semantic explanations (77.33% of *correct* marks). By having additional lexical knowledge at disposal, even in case of unrelated words, it is possible to imagine novel ways to put forward automatic reasoning approaches in tasks such as WSD, IR, etc. where finding semantic links or lexical overlap is fundamental (see Section 5.6).

5.6 Further Considerations

We presented the SAL process as a way to build real-time and context-based semantic knowledge that can enhance any possible consequent natural language processing task. In the case of semantic similarity, we were able to create additional semantic information regarding the comparison of two lexical items. Under a more general point of view, the proposed method can be applied on several other scenarios. We list here some of possible extension of our proposal, to pave the way for further research directions:
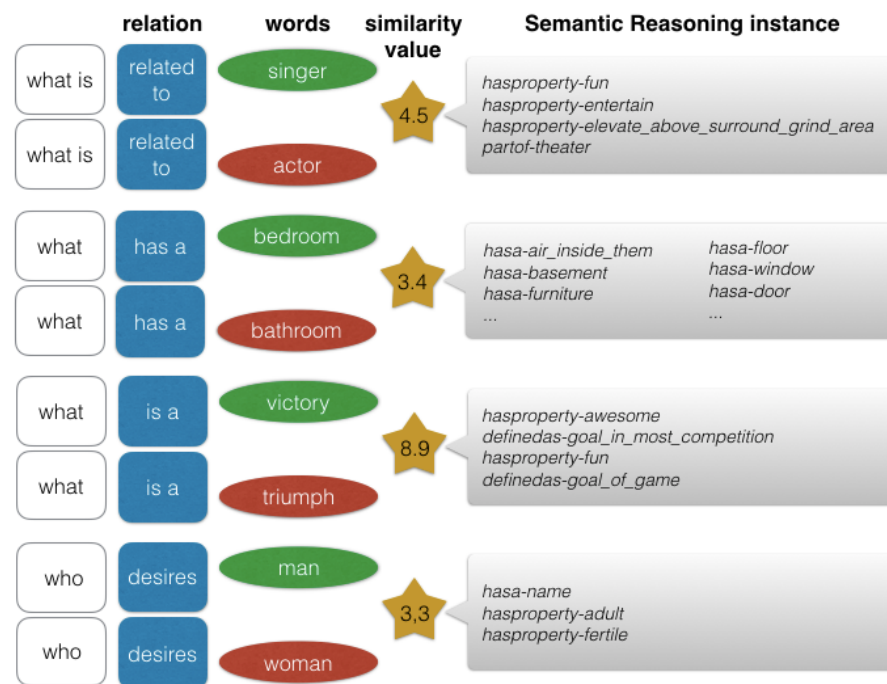
Fig. 5 Examples of SAL instances with different types of semantic relations. The first white blocks on the left helps read the instances, e.g., "what is related to both *singer* and *actor* has the properties *fun*, *entertain*, etc. and it is a part of the concept *theater*".

– In the context of Word Sense Disambiguation, an explanation-based similarity may constitute a novel kind of approach where words in a specific context could match with word senses through the use of correlation between semantic explanations rather than overlapping of word profiles (or vectors).

– In the Information Retrieval field (IR), complex queries may be seen as sets of shared explanations among the keywords in the query, possibly improving both Precision and Recall. In other words, instead of treating a query as a bag of words, it can be transformed into the explanations obtained by the proposed semantic similarity reasoning. For example, let us consider the 3-keywords query "*wolf dog behaviors*". The word "*dog*" should not be considered in the role of a pet, so results concerning pets (and so related to cats and parrots, for example) are out of the scope of the query. In a sense, the aim of the proposed method would be the removal of unnecessary senses related to the used words by shifting the analysis from a lexical to a semantic resource-based basis of correlations.

– Syntactic parsing is a procedure that often requires semantic information. A semantic reasoning approach could alleviate ambiguity problems at syn-

tactic level by using explanations. For example, major problems for syntactic parsing are prepositional-phrase and verbal attachments.

– Finally, this model could also help improve the state of the art on Natural Language Generation (NLG), and Summarization, since similarity reasoning can output lexical items which can be also not correlated to the used words in general, but that can play a requested role in a specific linguistic construction.

– The proposed approach could put the basis of a novel research methodology concerning Textual Entailment (TE). Actually, we think that the complexity of the task of understanding whether a lexical entity entails another lexical entity can be only solved at a semantic level rather than at vector-based level.

6 Related Work

This section is intended to introduce the necessary background knowledge and the state of the art with respect to the theories and the ideas discussed in this work: 1) the notion of *meaning* and *semantic similarity* with the relative connections with the proposed approach; 2) the difference between encyclopedic and common-sense semantic resources; and 3) ontology learning and integration of semantic resources.

6.1 Semantic Similarity and Meaning

As already mentioned in the Introduction, research on Computational Linguistics is often focused on the calculation of similarity scores between texts at different granularity (e.g., word, sentence, discourse, etc.) [20].

Although many measures have been proposed in the literature, this work is related to some cognitive theories such as the one of the *affordances* by James Gibson [13]. According to this theory, which refers to the perception of physical objects but that can be revisited in virtual or abstract situations, the element of a domain (e.g., objects in the physical world rather than lexical items in natural language) give clues about specific actions or meanings. Still, actions change the type of perception of an object, which models itself to fit with the context of use. The Gestalt theory [18] contains different notions about the perception of meaning according to interaction and context. In particular, the core of the model is the complementarity between the *figure* and the *ground*. In the linguistic case, a word can be seen as the figure while the context as the ground making light on its specific sense.

As a matter of fact, words are organized in a lexicon as a complex network of semantic relations which are basically subsumed within the Saussure's

paradigmatic (the axis of combination) and syntagmatic (the axis of choice) axes [28]. Some authors [5] have already suggested theoretical and empirical taxonomies of semantic relations consisting of some main families of relation (such as contrast, similars, class inclusion, part-whole, etc.). As Murphy pointed out [23], lexicon has become more central in linguistic theories and, even if there is no widely accepted theory on its internal semantic structure and how lexical information are represented in it, the semantic relations among words are considered in scholarly literature as relevant to the structure of both lexical and conceptual information and it is generally believed that relations among words determine meaning.

Distributional Analysis of natural language, such as Distributional (or vector-based) semantics, exploits Harris's distributional hypothesis (later summarized with Firth's sentence *"you shall know a word by the company it keeps"*) and sees a word meaning as a vector of numeric occurrences (i.e., frequencies) in a set of linguistic contexts (documents, syntactic dependencies, etc.). This approach provides a semantics of similarity which relies on a geometrical representation of the word meanings, and so in terms of vector space models (VSMs, [27]). This view has been recently gaining a lot of interest and success, also due to the growing availability of large corpora from where to obtain statistically-significant lexical correlations. Data Mining (DM) techniques fully leveraging on VSMs and Latent Semantic Analysis (LSA) [8] have been successfully applied on text since many decades for information indexing and extraction tasks, using matrix decompositions such as Singular Value Decomposition (SVD) to reconstruct the latent structure behind the abovementioned distributional hypothesis, often producing concept-like entities in the form of words clusters sharing similar contexts. However, distributional approaches are usually good in finding lexical relatedness rather than similarity.

6.2 Semantic Resources: Computational Lexicons and Common-Sense

In the last 20 years, the Artificial Intelligence (AI) community working on Computational Linguistics (CL) has been using one knowledge base among all, i.e., WordNet [22]. In few words, WordNet was a first answer to the most important question in this area, which is the treatment of language ambiguity. Generally speaking, a word is a symbolic expression that may refer to multiple meanings (polysemy), while distinct words may share the same meaning (synonymy). Syntax reflects grammar rules which add complexity to the overall communication medium, making CL one of the most challenging research area in the AI field.

From a more detailed perspective, WordNet organizes words in *synsets*, i.e., sets of words sharing a unique meaning in specific contexts (synonyms), further described by descriptions (*glosses*) and examples. Synsets are then structured in a taxonomy which incorporates the semantics of generality/specificity of the referenced concepts. Although extensively adopted, the limits of this resource

are sometimes critical: 1) the top-down and general-purpose nature at the
basis of its construction lets asking about the actual need of some underused
meanings, and 2) most Word Sense Disambiguation approaches use WordNet
glosses to understand the link between an input word (and its context) and
the candidate synsets.

As a matter of fact, these tasks require an incredibly rich semantic knowl-
edge containing facts related to behavioural rather than conceptual informa-
tion, such as what an object may or may not do or what may happen with
it after such actions. In the light of this, an interesting source of additional
gloss-like information is represented by common-sense knowledge (CSK), that
may be described as a set of shared and possibly general facts or views of the
world. Being often crowd-sourced, CSK represents a promising (although often
complex and uncoherent) type of information which can serve complex tasks
such as the ones mentioned above. ConceptNet is one of the largest sources of
CSK, collecting and integrating data from many years since the beginning of
the MIT Open Mind Common Sense project[7]. However, terms in ConceptNet
are not disambiguated, so it is difficult to use due to its large amount of lexical
ambiguities. Table 9 shows some of the semantic relations in ConceptNet.

Table 9 Some of the existing semantic relations in ConceptNet, with example sentences in
English.

Relation	Example sentence
MadeOf	NP is made of NP.
DefinedAs	NP is defined as NP.
HasA	NP has NP.
HasProperty	NP is AP.
UsedFor	NP is used for VP.
CapableOf	NP can VP.
HasPrerequisite	NP—VP requires NP—VP.
MotivatedByGoal	You would VP because you want VP.

Among the more unusual types of relationships (28 in total), it contains
information like "*ObstructedBy*" (i.e., referring to what would prevent it from
happening), "and *CausesDesire*" (i.e., what does it make you want to do). In
addition, it also has classic relationships like "*is_a*" and "*part_of*" as in most
linguistic resources (see Table 9 for examples of property-based and function-
based semantic relations in ConceptNet).

This paper presents a method for the automatic enrichment of WordNet
with disambiguated semantics of ConceptNet 5. In particular, the proposed
technique is able to disambiguate common-sense instances by linking them to
WordNet synsets.

---

[7] http://media.mit.edu/research/groups/5994/open-mind-common-sense

6.3 Ontology Learning and Integration of Semantic Resources

Ontology Learning is one of the cornerstone of recent research on Computational Linguistics, as it stands for any automatic (or semi-automatic) techniques able to unravel structured lexical-based semantic knowledge from raw or semi-structured texts. The main inspirational work in the literature is the one by Hearst [15] for automatically extracting hypernyms by using fixed lexico-syntactic patterns.

The most well-known paradigm for knowledge representation at the lexical level is the concept of *word sense*. Word senses are the basis of computational lexicons such as WordNet [22] and its counterparts in other languages [2]. Wordnets usually contain human-readable concept definitions and recognize meanings in terms of paradigmatic relations such as hyperonymy and meronymy that hold between them.

There exists a large literature on integrating different computational lexicons with other types of structured knowledge. Among all, BabelNet [24] represents maybe the widest effort of integration from different sources (e.g., Wikipedia, WordNet, OmegaWiki, etc.) and in a multilingual context. However, an unsolved issue remains the fine granularity of the word sense inventory and its sparse coverage and actual usage.

This contribution is strictly related to the concept of word sense and on the ideas of Ontology Learning and resources integration. Specifically, it constructs novel combinations of lexical semantic relations my aligning different computational lexicons at word sense-level. However, it avoids numeric computation of weights (as with the above-mentioned DS) for building new semantic information. On the contrary, it applies the general idea of collaborative filtering for assessing the correctness of candidate alignments between two different semantic resources.

7 Conclusions and Future Works

The aim of this paper was threefold: 1) we studied how common-Sense knowledge enables the perception of semantic similarity in a cognitive experiment (Research Question n.1); then, we proposed an integration between one of the largest CSK resource (i.e., ConceptNet) into the well-known WordNet, demonstrating how the resulting resource can improve the recognition of word similarity using two different similarity datasets (Research Question n.2).; finally, we highlighted the power of common-sense for labeling semantically-associated words, in order to produce *explanations* which can be further exploited to contextualize lexical comparisons (Research Question n.3).

We evaluated all three proposed modules and we publicly released the generated data (the semantic integration of ConceptNet into WordNet, and the output of the Semantic Association Labeling process on the word pairs of the SimLex similarity dataset).

In future work, we aim at extending our idea to lexical entities of higher granularity (such as n-grams, sentences, etc), through other recently-published annotated data such as the Blue Norwegian Parrot dataset [19].

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
2. Bond, F., Foster, R.: Linking and extending an open multilingual wordnet. In: ACL (1), pp. 1352–1362 (2013)
3. Bruni, E., Boleda, G., Baroni, M., Tran, N.K.: Distributional semantics in technicolor. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1, pp. 136–145. Association for Computational Linguistics (2012)
4. Caro, L.D., Boella, G.: Automatic enrichment of wordnet with common-sense knowledge. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016. (2016). URL http://www.lrec-conf.org/proceedings/lrec2016/summaries/38.html
5. Chaffin, R., Herrmann, D.J.: The similarity and diversity of semantic relations. Memory & Cognition **12**(2), 134–141 (1984)
6. Chen, J., Liu, J.: Combining conceptnet and wordnet for word sense disambiguation. In: IJCNLP, pp. 686–694 (2011)
7. Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S.: Language-specific tuning of visual cortex? functional properties of the visual word form area. Brain **125**(5), 1054–1069 (2002)
8. Dumais, S.T.: Latent semantic analysis. Annual review of information science and technology **38**(1), 188–230 (2004)
9. Fillenbaum, S.: Words as feature complexes: False recognition of antonyms and synonyms. Journal of Experimental Psychology **82**(2), 400 (1969)
10. Fillmore, C.J.: Scenes-and-frames semantics. Linguistic structures processing **59**, 55–88 (1977)
11. Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E.: Placing search in context: The concept revisited. In: Proceedings of the 10th international conference on World Wide Web, pp. 406–414. ACM (2001)
12. Gärdenfors, P.: Conceptual spaces: The geometry of thought. MIT press (2004)
13. Gibson, J.J.: The Theory of Affordances. Lawrence Erlbaum (1977)
14. Harris, Z.S.: Distributional structure. Word (1954)
15. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th conference on Computational linguistics-Volume 2, pp. 539–545. Association for Computational Linguistics (1992)
16. Hill, F., Reichart, R., Korhonen, A.: Simlex-999: Evaluating semantic models with (genuine) similarity estimation. arXiv preprint arXiv:1408.3456 (2014)
17. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 50–57. ACM (1999)
18. Köhler, W.: Gestalt psychology. (1929)
19. Kruszewski, G., Baroni, M.: Dead parrots make bad pets: Exploring modifier effects in noun phrases. Lexical and Computational Semantics (* SEM 2014) p. 171 (2014)
20. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT press (1999)
21. McRae, K., Cree, G.S., Seidenberg, M.S., McNorgan, C.: Semantic feature production norms for a large set of living and nonliving things. Behavior research methods **37**(4), 547–559 (2005)
22. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
23. Murphy, M.L.: Semantic relations and the lexicon. Cambridge University Press (2003)

24. Navigli, R., Ponzetto, S.P.: Babelnet: Building a very large multilingual semantic network. In: Proceedings of the 48th annual meeting of the association for computational linguistics, pp. 216–225. Association for Computational Linguistics (2010)
25. Navigli, R., Ponzetto, S.P.: Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence **193**, 217–250 (2012)
26. Ruggeri, A., Cupi, L., Di Caro, L.: Word similarity perception: an explorative analysis. In: Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science (2015)
27. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11), 613–620 (1975). DOI 10.1145/361219.361220. URL http://doi.acm.org/10.1145/361219.361220
28. Saussure, F.d.: Course in general linguistics, trans. R. Harris, London: Duckworth (1983)
29. Shrout, P.E., Fleiss, J.L.: Intraclass correlations: uses in assessing rater reliability. Psychological bulletin **86**(2), 420 (1979)
30. Speer, R., Havasi, C.: Representing general relational knowledge in conceptnet 5. In: LREC, pp. 3679–3686 (2012)
31. Speer, R., Havasi, C.: Conceptnet 5: A large semantic network for relational knowledge. In: The People's Web Meets NLP, pp. 161–176. Springer (2013)
32. Vyas, V., Pantel, P.: Explaining similarity of terms. Coling 2008: Companion volume: Posters pp. 131–134 (2008)
33. Wilson, M.: Mrc psycholinguistic database: Machine-usable dictionary, version 2.00. Behavior Research Methods, Instruments, & Computers **20**(1), 6–10 (1988)