Annotating Concept Abstractness by Common-Sense Knowledge

Enrico Mensa, Aureliano Porporato, and Daniele P. Radicioni

Dipartimento di Informatica – Università degli Studi di Torino, Italy
aureliano.porporato@edu.unito.it,
{enrico.mensa,daniele.radicioni}@unito.it

Abstract. Dealing with semantic representations of concepts involves collecting information on many aspects that collectively contribute to (lexical, semantic and ultimately) linguistic competence. In the last few years mounting experimental evidences have been gathered in the fields of Neuroscience and Cognitive Science on conceptual access and retrieval dynamics that posit novel issues, such as the imageability associated to terms and concepts, or abstractness features as a correlate of figurative uses of language. However, this body of research has not yet penetrated Computational Linguistics: specifically, as regards as Lexical Semantics, in the last few years the field has been dominated by distributional models and vectorial representations. We recently proposed COVER, that relies on a partly different approach. Conceptual descriptions herein are aimed at putting together the lexicographic precision of BabelNet and the common-sense available in ConceptNet. We now propose Abs-COVER, that extends the existing lexical resource by associating an abstractness score to the concepts contained therein. We introduce the detailed algorithms and report about an extensive evaluation on the renewed resource, where we obtained correlations with human judgements in line or higher compared to state of the art approaches.

Keywords: Concept Abstractness, Concept Representation, Lexical Resources, Knowledge Representation, Figurative Language, NL Semantics

1 Introduction

Ordinary experience shows that semantic representation and lexical access and processing of concepts can be affected by concepts' concrete/abstract status: concrete meanings, more ingrained in the perceptual experience, are acknowledged to be more quickly and easily delivered in human communication than abstract meanings [1]. Such kind of information grasps a complex combination of experiential (e.g., sensory, motor) and strictly linguistic features, such as verbal associations arising through co-occurrence patterns and syntactic information [20]. These features make conceptual abstractness matter of broad interest for computational linguistics, and the investigation on conceptual abstractness a challenging though only superficially explored field. Information on conceptual abstractness impacts on many diverse NLP tasks, such as the word sense disambiguation task [10], the semantic processing of figurative language [3, 18, 14], the automatic translation and simplification [22], the characterisation of web queries with difficulty scores [21], the processing of social tagging information [2], and many others, as well.

One very first issue is, of course, that it is not straightforward to define abstractness [9]. Provided that more fine grained distinctions on abstract and concrete word meanings can be drawn, the term 'abstract' has two main interpretations: *i)* what is far from perception (as opposed to perceptible directly through the senses), and *ii)* what is more general (as opposed to low-level, specific). To implement the second view, the concreteness or *specificity* —the opposite of abstractness— can be defined as a function of the distance intervening between a concept and a parent of that concept in the top-level of a taxonomy or ontology [5]. In this setting, the second definition can be used to automatically compute abstractness given an ontology-like resource (like WordNet or BabelNet [17]) without any additional information from human beings. On the other side, the first definition seems to better correlate with what is perceived as "abstract" in human judgement [19].

In this work we basically refer to the first aspect —perceptually salient abstractness—, and enrich it with common-sense information; additionally, different from most existing literature (e.g., [4]), we consider abstractness as a feature of word meanings (concepts) rather than a feature of word forms (terms): our work thus consists of annotating with abstractness information the concepts in COVER [15], a lexical resource developed in the frame of a long-standing research aimed at combining ontological and common-sense reasoning [8, 13, 11]. As a result, we propose an extended lexical resource, ABS-COVER,[1] where each and every concept is automatically annotated with an abstractness score ranging in the $[0, 1]$ interval, where the left bound 0.0 features fully concrete concepts, and the right bound 1.0 stands for maximally abstract concept.

The paper is organised as follows. We first propose a review of the related work on this and close issues (Section 2); in Section 3 we then illustrate how the abstractness score featuring COVER concepts is computed. Later on we extensively evaluate the proposed annotation and discuss the obtained results (Section 4). We conclude by pointing out future work, to refine our approach and improve the quality of the abstractness annotation.

2 Related Work

An automatic approach has been devised using abstractness information to analyse web image queries, and to characterise them in terms of difficulty [21]. In particular, the authors compute the abstractness associated to nouns by checking the presence of the *physical entity* synset among the hypernyms of senses in the WordNet taxonomy. This approach also involves a disambiguation step, which is

[1] ABS-COVER is available for download at `https://ls.di.unito.it`.

performed through a model trained on the SemCor *corpus* [16]. The technique carried out in [5] also relies on WordNet information, but the abstractness of a concept (also called *specificity*) is here defined as the distance between the corresponding node in WordNet, and the root of the ontology. The more specific (lower in the hierarchy) is a concept, the more concrete, according to the second definition of abstractness reported in the previous Section. Given that in WordNet we have at most 17 levels of depth, conceptual concreteness varies over the interval $[0, 16]$. A closely related strategy [9], based on similar assumptions, proposes the notions of *precision by depth* (P-depth), together with other two abstractness measures: *precision by inclusiveness* (P-inclusiveness), based on the fraction of descendants of a node with respect to the overall number of nodes in WordNet; and *concreteness*, based on the sensory definition of abstractness.

The authors of the work [19] compare the first two methods, one based on the definition of abstractness that checks for the presence of *physical entity* among the hypernyms of a concept, and one based on the second —specificity-based— notion of abstractness. Interestingly enough, the authors report a 0.17 Spearman correlation between scores obtained with the method in [5] and those obtained in [21], in line with the findings about the correlation of values based on the two different definitions [9]. This measure can be considered as an estimation of the overlap of the two notions of abstractness: the poor correlation seems to confirm that they are rather distinct. Furthermore, the two sets of scores have been compared with those in the MRC data set, reporting a 0.60 Spearman correlation between the abstractness scores proposed by [21] and the human judgements, and a 0.29 correlation between the scores by [5] and human ratings. Such experimental evidence suggested us to adopt the first definition of abstractness.

The role of abstractness has also been explored in the context of the Word Sense Disambiguation [10], leading to the finding that words with very high or very low score of abstractness are easier to disambiguate. Along this line, the association of word senses with senses of different words have been examined, finding that concrete concepts tend to be related to concrete concepts and abstract concepts tend to be related to abstract ones. In particular, concrete concepts would be more related to concrete concepts than are abstract concepts to abstract concepts. Similar conclusions have been recently reached in [7].

3 COVER Annotation

In this Section we describe how the common-sense knowledge already present in COVER has been used to enrich it with abstractness information. Before providing the annotation algorithm, for the sake of self containedness, we briefly introduce COVER.

3.1 Introduction to COVER

COVER is a lexical resource aimed at hosting general conceptual representations. Full details on COVER and on the algorithm designed to build it by

```
Algorithm 1: The COVER Annotation Algorithm.
              Data: a set of COVER elements $C$, a set of COVER dimensions $D$
              Result: a set of pairs $(e, a)$, with $e \in C$ and $a$ updated abstractness score for $e$
First Step    $T \longleftarrow \bigcup_{c \in C}(c, \texttt{IsAbstract}(c.\texttt{wnsi},\ c.\texttt{bsi}))$
Second Step   return $\bigcup_{c \in C}(c, \texttt{RefineAbstracness}(c,\ D,\ T))$
```

integrating BabelNet and ConceptNet can be found in [15]. Each concept c in COVER is identified through a BabelNet synset ID and described as a vector representation $\vec{c}$, composed by a set of semantic dimensions $\mathcal{D} = \{d_1, d_2, \dots d_n\}$. Each such dimension encodes a relationship like, e.g., IsA, UsedFor, *etc.* and reports the concepts connected to c along the dimension d_i. The vector space dimensions are based on ConceptNet relationships.[2] The dimensions are filled with BabelNet synset IDs, so that finally each concept c in COVER can be defined as

$$\vec{c} = \bigcup_{d \in \mathcal{D}} \{\langle ID_d, \{c_1, \cdots, c_k\}\rangle\} \tag{1}$$

where ID_d is the identifier of the d-th dimension, and $\{c_1, \cdots, c_k\}$ is the set of values (concepts themselves) filling d.

3.2 COVER Annotation Algorithm

In order to enrich COVER with abstractness information, we took inspiration from the concreteness criterion exposed in [21], and we follow this idea: a concept is concrete if it descends from *physical entity* in WordNet, abstract otherwise. Algorithm 1 shows the main procedure for the abstractness annotation, that consists of two steps:

1. **First Step** (Algorithm 2): this function is designed to compute a base abstractness score for each element e in COVER, where an *element* is a concept (i.e., a BabelNet synset ID) that either has a vector representation or is a value inside a vector. In order to compute this score, we perform the following steps:
   (a) we attempt to retrieve the list of WordNet synset IDs associated to e in BabelNet, and from those we collect the WordNet hypernyms set (Algorithm 2, (1)); if in this set we find the synset of *physical entity*,[3] the abstractness score of e is set to 0.0; otherwise it is set to 1.0 (3);
   (b) if (a) fails (i.e., no WordNet synset ID can be found for e), we collect the direct BabelNet hypernyms of e and the search described in (a) is performed for each such hypernym (4). If at least one of e hypernyms has *physical entity* among its hypernyms, the base abstractness score of e is set to 0.0, and to 1.0 otherwise (6);

[2] The most relevant relationships include: RelatedTo, IsA, AtLocation, UsedFor, CapableOf, PartOf, HasProperty, MadeOf, HasA, InstanceOf.

[3] The synset for *physical entity* has ID wn:00001930n in WordNet 3.0.

Algorithm 2: Auxiliary `IsAbstract` function.

```
    Input: a BabelNet synset $b$
    Output: the base abstractness score of the COVER element corresponding to $b$
    Function IsAbstract($b$):
 1      $S \longleftarrow$ WordNetHypernyms($b$)
 2      if $S \neq \emptyset$ then
 3          if physical entity $\in S$ then
                return 0
            else
                return 1
        else
            $H \longleftarrow$ BabelNetHypernyms($b$)
 4          $W \longleftarrow \bigcup_{h \in H}$WordNetHypernyms($h$)
 5          if $W \neq \emptyset$ then
 6              if physical entity $\in W$ then
                    return 0
                else
                    return 1
            else
 7              $g \longleftarrow$ GetMainBabelNetGloss($b$)
 8              $N \longleftarrow$ Babelfy($g$)
                $G \longleftarrow$ []
                for each $n$ noun concept $\in N$ do
 9                  $q \longleftarrow$ GetGlossConceptAbstractness(n)
10                  if $q \geq 0$ then
                        append $q$ to $G$
11              if $G$ is not empty then
                    return average of scores in $G$
                else
12                  return $-1$
```

(c) if (b) fails (that is, e has no hypernyms in BabelNet or none of them has an associated WordNet synset ID), we retrieve the BabelNet main gloss for e (7), disambiguate it,[4] thus obtaining a set of concepts N. In order to compute the abstractness score for each noun in the gloss, steps (a) and (b) are performed on each $n \in N$. Finally, the valid scores associated to the nouns are averaged and the result is assigned as the abstractness score of e (9–11).

If the function fails in all of these steps, the abstractness score is set to -1, indicating that no suitable score could be computed (12).

2. **Second Step** (Algorithm 3): the first step enriches every concept in COVER with a base score of abstractness. The goal of the second step is to smooth such scores by following human perception accounts; to do so, we employ the common-sense knowledge available in COVER. Given a vector $\vec{c}$ in the

[4] At the present stage the disambiguation is performed by using Babelfy APIs (`http://babelfy.org/`).

Algorithm 3: Auxiliary `RefineAbstracness` function.

```
  Input: a COVER element $elem$, a set of COVER dimensions $D$, a set $A$ of
         pairs $(c, a)$, with $c$ COVER element and $a$ base abstractness score of $c$
  Output: the refined abstractness score for $elem$
  Function RefineAbstracness($elem, D, A$):
      $s_{\text{vec-base}} \longleftarrow A(elem)$                // find the score of v in A
      if $elem$ is a COVER vector then
          $L \longleftarrow []$
          for each dimension $dim \in D$ do
              for each value $v \in elem.dim$ do
                  $abstr_v \longleftarrow A(v)$
1                 if $abstr_v \geq 0$ then
                      append $abstr_v$ to $L$
2         if $L$ is not empty then
              $s_{\text{values-avg}} \longleftarrow$ average of scores in $L$
          else
              $s_{\text{values-avg}} \longleftarrow -1$
3         case $s_{vec\text{-}base} \geq 0$ AND $s_{values\text{-}avg} \geq 0$ return $\frac{s_{\text{vec-base}} + s_{\text{values-avg}}}{2}$
4         case $s_{values\text{-}avg} \geq 0$ return $s_{\text{values-avg}}$
5         otherwise return $s_{\text{vec-base}}$
      else
6         return $s_{\text{vec-base}}$
```

resource, we explore a subset of its dimensions:[5] all the base abstractness scores of the concepts that are values for these dimensions are retrieved, and the average score $s_{\text{values-avg}}$ is computed. Concepts having an invalid score are discarded (1, 2). The score $s_{\text{values-avg}}$ is then in turn averaged with $s_{\text{vec-base}}$, that is the base score of $\vec{c}$ (3), thus obtaining the final score for the COVER vector. If either $s_{\text{vec-base}}$ or $s_{\text{values-avg}}$ are invalid scores, the final score of $\vec{c}$ is set to the only valid score available.

It is important to note that the scores computed in the first step are frozen, and they are not dynamically updated during the execution of the second step. This is important to ensure that the order in which the vectors are considered does not impact on the final result of the annotation. Moreover, we did not iterate the second step, since it would potentially drift the scores from the precise information given by WordNet. In the end, any element defined as a *physical entity* in WordNet retains an abstractness score lesser than or equal to 0.5.

4 Evaluation

In order to assess the abstractness scores of Abs-COVER we make use the Medical Research Council Psycholinguistic Dataset (MRC hereafter) [6] and the Brysbaert Dataset (BRYS hereafter) [4]. The MRC *corpus* has been built by merging

[5] We presently consider the following dimensions: RelatedTo, FormOf, IsA, Synonym, DerivedFrom, SimilarTo and AtLocation.

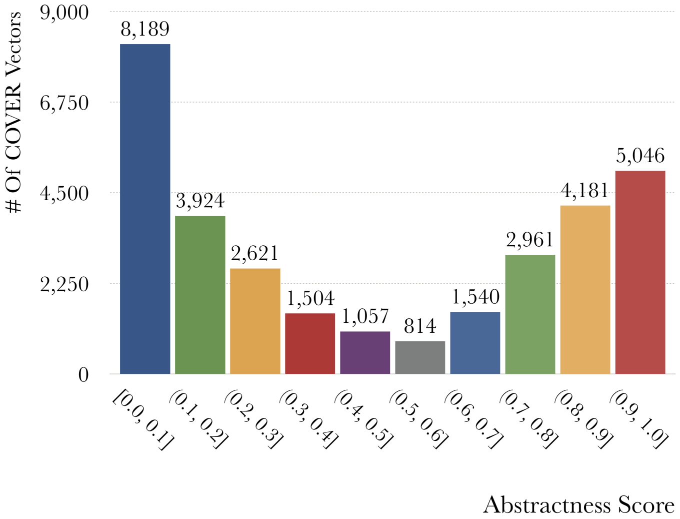

Fig. 1: Distribution of COVER vectors by abstractness score.

three handcrafted corpora containing words abstractness information.The MRC also provides additional information: each word has up to 25 associated features, such as imageability (i.e., how easily a word can evoke mental images) or meaningfulness (i.e., the confidence that a subject has about his understanding of the actual meaning of a word or expression), or its common part of speech. From this data set, containing in total $8,228$ terms with an abstractness value, we extracted $3,977$ nouns. On the other hand, the BRYS data set consists of a large of set of words annotated with abstractness scores through crowdsourcing, for a total of $39,945$ terms. We only use a portion of these terms (that is, the nouns contained in Abs-COVER).

We preliminarily observe that vectors associated to abstract concepts are well separated from those associated to concrete ones, as illustrated in Figure 1, showing how abstractness scores are distributed. In particular, the average score of concrete vectors (that is, the score of concepts featured by abstractness score lower than or equal to 0.5) is 0.153 and the average score of abstract vectors (whose abstractness score is greater than 0.5) is 0.837. Concrete vectors are slightly more frequent than abstract vectors: out of the $31,837$ vectors in COVER, we count overall $17,295$ and $14,542$ vectors, respectively.

A first evaluation of the annotation is based on *internal coherence*. We recorded the average abstractness of the values along all dimensions: also based on literature (see, e.g., [10, 7]) we expect that concrete vectors are, on average, more connected to other concrete concepts (e.g., through the Synonym relation). In Table 1 we report, over 15 of the most relevant (populated) relationships in COVER, the average scores that have been obtained by collecting all values along a given dimension. For example, we observe that the average abstract-

Dimension	Average Abstractness	
	Concrete Concepts	Abstract Concepts
RelatedTo*	0.293	0.694
IsA*	0.215	0.787
Synonym*	0.254	0.772
HasContext	0.632	0.805
FormOf*	0.127	0.777
DerivedFrom*	0.227	0.736
Antonym	0.312	0.750
AtLocation*	0.261	0.537
HasA	0.150	0.682
PartOf	0.181	0.681
SimilarTo*	0.241	0.751
UsedFor	0.464	0.719
HasProperty	0.385	0.727
Cause	0.450	0.811
CapableOf	0.473	0.687
HasPrerequisite	0.339	0.723

Table 1: Average abstractness score in COVER vectors' dimensions. Starred dimensions indicate the relations actually used in the Second Step. Concrete concepts are featured by abstractness score ($abs \leq 0.5$), while abstract concepts are those with $abs > 0.5$.

ness score for the values in the IsA relation is 0.215 when the vector involved is concrete, and 0.787 when the vector is abstract. The relations marked with '*' are those actually employed by the auxiliary function `RefineAbstractness`, (Algorithm 3). The only notable exception to this basic homogeneity principle is represented by vectors connected through the HasContext relation, where also concrete concepts are related to abstract concepts.

Although such figures are in accord with cited literature and seem to be reasonable on purely introspective accounts (thus qualitatively corroborating the proposed approach), an extensive experimentation has been devised to fully assess the annotated abstractness scores.

4.1 Correlation with Human Judgement

The second evaluation has been carried out by comparing the computed abstractness scores against human judgements. In particular, we considered the MRC and the BRYS data sets, whose scores were scaled into the range $[0, 1]$.

Before analysing the result, however, it should be noted that these data sets and ours are not directly comparable: while our scores are based on *concept* abstractness, such data sets are based on *word* abstractness. To overcome this

	MRC [6]	BRYS [4]	ABS-COVER
MRC	$r = 1$ $\rho = 1$	$r = 0.941$ $\rho = 0.871$	$r = 0.795$ $\rho = 0.663$
BRYS		$r = 1$ $\rho = 1$	$r = 0.766$ $\rho = 0.649$
ABS-COVER			$r = 1$ $\rho = 1$

Table 2: Pearson's r and Spearman's ρ correlation scores between the computed abstractness and human judgements, on a set of 150 manually disambiguated terms.

problem, we exploited the information available in the label of each COVER vector, a set of words associated to the concept represented by the vector itself. For example, given the word "mother", 5 conceptual descriptions can be found in COVER:

- *mother*, `bn:00029439n`, "The earth conceived of as the female principle of fertility";
- *mother/psi*, `bn:14001852n`, "Role-playing video game series by Shigesato Itoi";
- *mother*, `bn:03824293n`, "Energy drink from Coca-Cola";
- *mother/mommy/mom/motherhood*, `bn:00034027n`, "A woman who has given birth to a child (also used as a term of address to your mother)";
- *mother*, `bn:03322691n`, "A Broadway musical".

Clearly, in order to select one given sense, some sort of disambiguation is needed. Before evaluating the ABS-COVER scores in a pipeline that also includes this disambiguation step, we performed a preliminary exploration of the abstractness scores on a randomly chosen set of 150 terms present in both the MRC data set and in the BRYS data set. We computed the correlation between the abstractness values contained in each data sets, and those in ABS-COVER. Not to mix disambiguation errors with the evaluation of the abstractness scores, we manually performed the word sense disambiguation by selecting the sense of the word that seemed more relevant (taken in isolation, with no disambiguation context). Table 2 reports Pearson's r and Spearman's ρ correlations between the abstractness scores in these data sets and those in ABS-COVER. As it can be seen, the correlation between the two human-annotated sets is very high (although it shows that even human ratings are far from full agreement), and the correlation with ABS-COVER scores is high. We consider this sample test as an upper bound to the correlation that can be reached by undertaking some algorithmic approach to select a given sense in ABS-COVER. In the following, we consider the full problem of calculating the abstractness score for a word by using one or more vectors (concepts) from ABS-COVER.
Baseline. As the simplest strategy to retrieve ABS-COVER scores for comparison with the terms in the MRC and BRYS *corpora*, we took the average of the

		MRC	BRYS
Baseline – simple average	r	0.614	0.588
Baseline – simple average	ρ	0.612	0.590
No common-sense KW	r	0.547	0.524
No common-sense KW	ρ	0.544	0.518
Weighted average	r	0.655	0.608
Weighted average	ρ	0.648	0.612
Most salient sense	r	0.627	0.594
Most salient sense	ρ	0.646	0.615
SemCor-frequency	r	0.732	0.653
SemCor-frequency	ρ	0.704	0.639

Table 3: Pearson (r) and Spearman (ρ) correlation scores obtained by computing the abstractness scores from Abs-COVER senses according to several strategies.

abstractness scores featuring all senses available in Abs-COVER for a given input term. Table 3 reports Pearson and Spearman's correlation values obtained through simple average. From the MRC, we were able to retrieve overall $3,977$ nouns, 431 of which do not appear in the label (i.e., the set of lexicalizations) of any Abs-COVER vector. About the remaining $3,546$ nouns, $1,158$ are associated to 1 concept, and on average a word occurs in the label of 2.56 conceptual descriptions. The BRYS data set lacks of the part of speech annotation, so out of the total $39,954$ words (or compound expressions), $15,779$ occur in some vector's label ($8,722$ only once, 1.88 times on average).

No common-sense KW. In order to investigate how relevant is common-sense information, we devised a subtractive experiment. We created a version of Abs-COVER annotated without executing the second step of the annotation algorithm (Algorithms 1 and 3). We obtained a version of COVER with 14 unlabelled vectors (while no vector remains without abstractness annotation after executing the whole algorithm), $15,649$ zero-valued vectors (+547.99%) and $12,832$ one-valued vectors (+799.86%). As expected, the first step of the algorithm alone produced less smoothed scores, which were compared with human ratings. The correlation figures are reported in Table 3.

Weighted average. Some senses underlying any term (e.g., the term “mother”) are by far more common than other ones: in order to individuate one chief sense —which is ideally that considered by human annotators—, we examined the cardinality (that is, the sum of all concepts) of all dimensions featuring each Abs-COVER vector. The assumption is that the more broadly a sense is used in language and available to annotators, the larger the set of its connections to other senses. More on the possible criteria to perform selection, clustering and/or filtering of the sense inventory can be found in [12]. For example, the sense `bn:03824293n` associated to the term *mother* (an energy drink) has, over all dimensions, 6 values in COVER, while for the sense `bn:00034027n` (a woman who has given birth to a child) we find 119 values. In this experiment, the abstractness scores have been averaged over senses based on the overall amount of information available for each sense. The correlation values are reported in Table 3.

Most salient sense. Based on the same rationale, we designed another experimental condition. Instead of using all the senses extracted from ABS-COVER, weighted by the number of the concepts related with them, in this case we used only the most salient concept. As anticipated, the most salient sense is chosen as the vector featured by the highest number of concepts filling its dimensions. The results obtained by individuating abstractness scores through this strategy are reported in Table 3.

SemCor-based frequency. Another strategy to define suitable weights for word senses is to take into account their frequency in some *corpus*, assuming it reflects the distribution of senses in the considered language. So we collected sense frequency information based on the SemCor corpus [16], where words are annotated with information on part-of-speech and WordNet synset ID. Exploiting such information, we collected a set of word senses that are present in ABS-COVER and in the considered *corpora*. Namely, we took $2,417$ words that are present both in SemCor and in the MRC data set that have at least one associated WordNet Synset ID in ABS-COVER (including $1,433$ terms with only one sense, 1.76 senses per word on average); and $6,383$ words both in SemCor and in the BRYS data set (including $4,785$ terms with only one sense, 1.43 senses per word on average). Note that we not only considered only words present in both SemCor and the selected data set, but we restricted to the WordNet senses for the terms present in both the SemCor and COVER, thus finally reducing the number of both words and concepts: for example, in this setting the word "mother" is linked to a single sense (with ID `wn:10332385n`), "A woman who has given birth to a child". Table 3 reports the correlation obtained with the abstractness scores in ABS-COVER weighted according to the sense frequencies in the SemCor *corpus*.

4.2 Discussion

The reported figures are either in line or directly improve on state of the art approaches, such as [21] and [19]. We first computed the correlation between the human ratings contained in the MRC and BRYS data sets ($r = 0.94$, and $\rho = 0.87$, as shown in Table 2). This makes these figures a solid experimental base for comparison with the scores computed to annotate COVER. This datum is not new; in essence, it replicates previous experiments made by the authors of the BRYS data set [4]. Nevertheless, it was relevant to preliminarily verify the agreement between the two. As regards as the evaluation of our abstractness scores, as expected, we obtained the highest correlation in the first experimental condition where a small sample of terms was disambiguated by hand (Table 2).

This datum shows that WSD is still an open issue (though of secondary relevance, in the present setting), as the task of choosing one best sense for a given term.

The results obtained in the *baseline* experimental condition by computing a simple average among all senses' abstractness show a significant drop in the Pearson correlation with respect to results in Table 2: around 15% and 10% on the MRC on the BRYS data sets, respectively, as reported in Table 3. Similar drop is observed on the Spearman coefficients, with a more limited reduction, in the order of 4% in both data sets. The results obtained in the *No common-sense KW* condition seem to confirm the central role of common-sense knowledge in determining how abstract/concrete concepts are: in fact, these figures are the worst ones, with a reduction in the correlation with human judgement in the order of 20% for the r metrics, and 10% for the ρ metrics on both data sets. The *weighted average* condition obtained correlation above the baseline on both data sets and correlation metrics. It may be interpreted as a cue that the coverage of our resource is sufficient to provide information on the relevance of senses associated to a given term. Also, additional evidence should be collected to investigate on the aptitude of human annotators: when assessing the abstractness of a word, did they consider only one (most salient) concept taken in isolation, or a pool of concepts, where the most salient is the most prominent one surrounded by a set of satellite senses? Systems for Natural Interaction would benefit from this sort of information. The last experimental condition, based on the *SemCor frequency*, is that maximally approaching the first condition results (Table 2). Here we observe good correlation with human ratings (in the specific case of MRC data set and ρ metrics resulting in an improved correlation with scores obtained); also, it confirms the quality of the contribution provided by the SemCor *corpus* to the tasks ingrained in senses disambiguation.

5 Conclusions

In this paper we have proposed Abs-COVER, the novel version of COVER annotated with abstractness scores. At first, we have introduced the research question underlying the present work, and shown that this sort of information may be relevant to different NLP tasks. We have then introduced related work, and illustrated existing research efforts and approaches. Then the algorithms devised to compute abstractness have been introduced and illustrated in full detail. We have finally reported about an experimentation where we obtained valuable agreement with human ratings on concepts abstractness.

Finally, we realised that although the BRYS data set provides a great deal of information, some further elements would be beneficial for NLP experiments. Terms should be sense-annotated, hopefully by adopting the *de-facto* standard naming convention of BabelNet; additionally, some measure of inter-annotator agreement should be reported, so to distinguish cases that are straightforward (at least for human rating) from those more complicated, to further refine systems' accuracy on the latter ones.

References

1. Bambini, V., Resta, D., Grimaldi, M.: A dataset of metaphors from the Italian literature: exploring psycholinguistic variables and the role of context. PloS one 9(9), 1–13 (2014)
2. Benz, D., Körner, C., Hotho, A., Stumme, G., Strohmaier, M.: One tag to bind them all: Measuring term abstractness in social metadata. In: Procs. of ESWC. pp. 360–374. Springer (2011)
3. Birke, J., Sarkar, A.: A clustering approach for nearly unsupervised recognition of nonliteral language. In: Procs. of the 11th conference of EACL (2006)
4. Brysbaert, M., Warriner, A.B., Kuperman, V.: Concreteness ratings for 40,000 generally known english word lemmas. BEHAV RES METH 46(3), 904–911 (2014)
5. Changizi, M.A.: Economically organized hierarchies in wordnet and the oxford english dictionary. Cognitive Systems Research 9, 214–228 (2008)
6. Coltheart, M.: The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A 33(4), 497–505 (1981)
7. Frassinelli, D., Naumann, D., Utt, J., Walde, I., Schulte, S.: Contextual characteristics of concrete and abstract words. In: IWCS 2017 (2017)
8. Ghignone, L., Lieto, A., Radicioni, D.P.: Typicality-Based Inference by Plugging Conceptual Spaces Into Ontologies. In: Procs. of AIC. CEUR (2013)
9. Iliev, R., Axelrod, R.: The paradox of abstraction: Precision versus concreteness. Journal of psycholinguistic research 46(3), 715–729 (2017)
10. Kwong, O.Y.: Sense abstractness, semantic activation, and word sense disambiguation. International Journal of Speech Technology 11(3-4), 135 (2008)
11. Lieto, A., Mensa, E., Radicioni, D.P.: A Resource-Driven Approach for Anchoring Linguistic Resources to Conceptual Spaces. In: Procs. of the 15th AIxIA. LNAI (10037), Springer (2016)
12. Lieto, A., Mensa, E., Radicioni, D.P.: Taming sense sparsity: a common-sense approach. In: Procs. of CLiC-it 2016 (2016)
13. Lieto, A., Minieri, A., Piana, A., Radicioni, D.P.: A knowledge-based system for prototypical reasoning. Connection Science 27(2), 137–152 (2015)
14. Mensa, E., Porporato, A., Radicioni, D.P.: Grasping metaphors: Lexical semantics in metaphor analysis. In: The Semantic Web: ESWC 2018 Satellite Events. pp. 192–195. Springer, Cham (2018)
15. Mensa, E., Radicioni, D.P., Lieto, A.: COVER: a linguistic resource combining common sense and lexicographic information. LANG RESOUR EVAL (Jun 2018)
16. Miller, G.A., Leacock, C., Tengi, R., Bunker, R.T.: A semantic concordance. In: Procs. of the workshop on Human Language Technology. pp. 303–308. ACL (1993)
17. Navigli, R., Ponzetto, S.P.: BabelNet: Building a very large multilingual semantic network. In: Procs. of the 48th ACL. pp. 216–225. ACL (2010)
18. Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., Frieder, O.: Metaphor identification in large texts corpora. PloS one 8(4), e62343 (2013)
19. Theijssen, D., van Halteren, H., Boves, L., Oostdijk, N.: On the difficulty of making concreteness concrete. CLIN Journal pp. 61–77 (2011)
20. Vigliocco, G., Meteyard, L., Andrews, M., Kousta, S.: Toward a theory of semantic representation. Language and Cognition 1(2), 219–247 (2009)
21. Xing, X., Zhang, Y., Han, M.: Query difficulty prediction for contextual image retrieval. In: European Conference on Information Retrieval. pp. 581–585 (2010)
22. Zhu, Z., Bernhard, D., Gurevych, I.: A monolingual tree-based translation model for sentence simplification. In: Procs. of the 23rd international conference on computational linguistics. pp. 1353–1361. ACL (2010)