Conceptual Abstractness: from Nouns to Verbs

Davide Colla, Enrico Mensa, Aureliano Porporato & Daniele P. Radicioni
Dipartimento di Informatica – Università degli Studi di Torino
`firstname.surname@unito.it`

Abstract

Investigating lexical access, representation and processing involves dealing with conceptual abstractness: abstract concepts are known to be more quickly and easily delivered in human communications than abstract meanings (Binder et al., 2005). Although these aspects have long been left unexplored, they are relevant: abstract terms are widespread in ordinary language, as they contribute to the realisation of various sorts of figurative language (metaphors, metonymies, hyperboles, *etc.*). Abstractness is therefore an issue for computational linguistics, as well. In this paper we illustrate how to characterise verbs with abstractness information. We provide an experimental evaluation of the presented approach on the largest existing *corpus* annotated with abstraction scores: our results exhibit good correlation with human ratings, and point out some open issues that will be addressed in future work.

Italiano. *In questo lavoro presentiamo il tema dell'astrattezza come una caratteristica diffusa del linguaggio, e un nodo cruciale nell'elaborazione automatica del linguaggio. In particolare illustriamo un metodo per la stima dell'astrattezza che caratterizza i verbi a partire dalla composizione dei punteggi di astrattezza degli argomenti dei verbi utilizzando la risorsa* Abs-COVER.

1 Introduction

Surprisingly enough, most of frequently used words (70% of the top 500) seem to be associated to abstract concepts (Recchia and Jones, 2012). Coping with abstractness is thus central to the investigation of lexical access, representation, and processing and, consequently, to build systems dealing with natural language. Information on conceptual abstractness impacts on many diverse NLP areas, such as word sense disambiguation (WSD) (Kwong, 2008), the semantic processing of figurative uses of language (Turney et al., 2011; Neuman et al., 2013), automatic translation and simplification (Zhu et al., 2010), the processing of social tagging information (Benz et al., 2011), and many others, as well. In the WSD task, abstractness has been investigated as a core feature in the fine tuning of WSD algorithms (Kwong, 2007): in particular, experiments have been carried out showing that "words toward the concrete side tend to be better disambiguated that those lying in the mid range, which are in turn better disambiguated than those on the abstract end" (Kwong, 2008).

A recent, inspiring, special issue hosted by the Topics in Cognitive Science journal on 'Abstract Concepts: Structure, Processing, and Modeling' provides various pointers to tackle abstractness, by posing it as a relevant issue for several disciplines such as psychology, neuroscience, philosophy, general AI and, of course, computational linguistics (Bolognesi and Steen, 2018). As pointed out by the Editors of the special issue, the investigation on abstract concepts is central in the multidisciplinary debate between grounded views of cognition *versus* modal (or symbolic) views of cognition. In short, cognition might be *embodied* and grounded in perception and action (Gibbs Jr, 2005): accessing concepts would amount to retrieving and instantiating perceptual and motoric experience. Typically, abstract concepts, that have no direct counterpart in terms of perceptual and motoric experience, are accounted for by such theories with difficulty. On the other side, modal approaches to concepts are mostly in the realm of distributional semantic models: in this view, the meaning of *rose* is "the product of statistical computations from associations between *rose* and concepts like *flower*, *red*, *thorny*, and *love*" (Louwerse, 2011).[1]

While we do not enter this passionate debate, we start by considering that distributional models are of little help in investigating abstractness, with some notable exceptions, such as the interesting links between abstractness and emotional content drawn in (Lenci et al., 2018). In fact, whilst distributional models can be easily used to express similarity and analogy (Turney, 2006), since they are basically built on co-occurrence matrices, they are largely acknowledged to convey vague associations rather than defining a semantically structured space (Lenci, 2018). As illustrated in the following, our approach is different from such mainstream approach, in that the conceptual descriptions used to compute abstractness and contained in the lexical resources COVER (Mensa et al., 2018c) and ABS-COVER (Mensa et al., 2018b)[2] are aimed at putting together the lexicographic precision and richness of BabelNet (Navigli and Ponzetto, 2012) and the common-sense knowledge available in ConceptNet (Havasi et al., 2007).

One preliminary issue is, of course, how to define abstractness, since no general consensus has been reached on what should be measured when considering abstractness or, conversely, concreteness (Iliev and Axelrod, 2017). The term 'abstract' has two main interpretations: *i)* what is *not perceptually salient*, and *ii)* what is *less specific*, and referred to the more general categories contained in the upper levels of a taxonomy/ontology. According to 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 (Changizi, 2008): the closer to the root, the more abstract. In this setting, existing taxonomies and ontology-like resources can be directly employed, such as WordNet (Miller et al., 1990) or BabelNet (Navigli and Ponzetto, 2012).

In this work we single out the first aspect, and focus on perceptually salient abstractness; we start from a recent work where we proposed an algorithm to compute abstractness (Mensa et al., 2018a) for concepts contained in COVER (Mensa et al., 2018c; Lieto et al., 2016),[3] and we extend that approach in order to characterise also verbs, whose abstractness is presently computed by combining the abstractness of their (nominal) dependents. Different from most literature we treat abstractness as a feature of word meanings (senses), rather than a feature of word forms (terms).

2 Related Work

Due to space reasons we cannot provide a full account of the related work from a scientific perspective nor about applications and systems; we limit to adding a mention to the closest and most influential approaches. Abstractness has been used to analyse web image queries, and to characterise them in terms of processing difficulty (Xing et al., 2010). In particular, the abstractness associated to nouns is computed 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 performed through a model trained on the SemCor *corpus* (Miller et al., 1993).

Methods based on both (perceptual *vs.* specificity-based) notions of abstractness are compared in (Theijssen et al., 2011). Specifically, the authors of this work report a 0.17 Spearman correlation between scores obtained with the method by (Changizi, 2008) and those obtained by (Xing et al., 2010), in line with the findings about the correlation of values based on the two definitions. This score can be considered as an estimation of the overlap of the two notions of abstractness: the poor correlation seems to suggest that they are rather distinct.

Finally, the abstractness scores by (Xing et al., 2010) and (Changizi, 2008) have been compared with those in the Medical Research Council Psycholinguistic (MRC) Dataset (Coltheart, 1981) reporting, respectively, a 0.60 and 0.29 Spearman correlation with the human ratings.

[1] *Modal* or *symbolic* views of cognition should not be confused with the symbolic AI, based on high-level representations of problems, as outlined by the pioneering work by Newell and Simon (such as, e.g., in (Newell, 1980)), that was concerned with physical symbol systems

[2] `https://ls.di.unito.it`.

[3] COVER is a lexical resource developed in the frame of a long-standing research aimed at combining ontological and common-sense reasoning (Ghignone et al., 2013; Lieto et al., 2015; Lieto et al., 2017).

3 From Nouns to Verbs Abstractness

In this Section we recall the conceptual representation implemented in COVER; we then describe how the resource has evolved into ABS-COVER, that provides nouns with abstractness scores. We then show how abstractness scores are computed for verbs.

COVER is a lexical resource aimed at hosting general conceptual representations. Each concept c is identified through a BabelNet synset ID and described as a vector representation $\vec{c}$, composed by a set of semantic dimensions $\mathcal{D} = \{d_1, d_2, \ldots d_n\}$. Each such dimension encodes a relationship like, e.g., ISA, USEDFOR, HASPROPERTY, CAPABLEOF, *etc.* and reports the concepts that are connected to c along the dimension d_i. The vector space dimensions are based on ConceptNet relationships. The dimensions are filled with BabelNet synset IDs, so that finally each concept c in COVER can be defined as

$$\vec{c} = \bigcup_{d \in \mathcal{D}} \{\langle \mathit{ID}_d, \{c_1, \cdots, c_k\} \rangle\}$$

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.

3.1 Annotation of Nouns in ABS-COVER

The annotation of COVER concepts is driven by the hypothesis that the abstractness of a concept can be computed by the abstractness of its ancestor(s) (basically, its hypernyms in WordNet), resorting to their top level super class, either abstract or concrete entity, as previously done in (Xing et al., 2010). In ABS-COVER 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 main algorithm consists of two steps, the *base score computation* and the *smoothing phase* (Mensa et al., 2018a).

The **base score computation** is designed to compute a base abstractness score for each element e in COVER. *a*) The algorithm first looks up for the concepts associated to e in BabelNet and retrieves the corresponding set of WordNet hypernyms: if these contain the *physical entity* concept, the base abstractness score of e is set to 0.0; otherwise it is set to 1.0. *b*) In case of failure (i.e., no WordNet synset ID can be found for e), the direct BabelNet hypernyms of e are retrieved and the step *a* is performed for each such hypernyms. Finally, *c*) in case taxonomic information cannot be exploited for e, the BabelNet main gloss for e is retrieved and disambiguated, thus obtaining a set of concepts N. We then perform steps (*a* and *b*) for each noun $n \in N$. The gloss scores are averaged and the result is assigned as score of e. If the function fails in all of these steps, the abstractness score is set to -1, indicating that no suitable score could be computed. For example, the concept *bomb* as "an explosive device fused to explode under specific conditions",[4] is connected to *physical entity* through its hypernyms in WordNet; thus, its base score is set to 0.0.

The **smoothing phase** focuses on the tuning of the base scores previously obtained by following human perception accounts; to do so, we employ the common-sense knowledge available in COVER. Given a vector $\vec{c}$ in the 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. The score $s_{\text{values-avg}}$ is then in turn averaged with $s_{\text{vec-base}}$, that is the base score of $\vec{c}$, thus obtaining the final score for the COVER vector. Continuing our previous example concerning the concept *bomb*, the average abstractness score of its dimension values is mostly low. Specifically, the "bomb" vector in COVER contains, for instance, "bombshell" (with a score of 0.0), "war" (with a score of 1.0) and "explosive material" (with a score of 0.0). The average of *bomb*'s values is 0.2245 and thus the final, smoothed abstractness score for *bomb* is set to 0.112.

3.2 Annotation of Verbs

COVER does not include a conceptual representation for verbs: only nouns are present herein, and this is currently an active line of research aiming at ameliorating the resource. However, in order to build practical applications, we needed to be able to also characterise verb abstractness (Mensa et al., 2018b). In this work we do not aim at extending COVER with verbs representations, but rather to see if the nouns in ABS-COVER can be

[4] Featured by the WordNet synset ID `wn:02866578n`.

[5] We presently consider the following dimensions: RELATEDTO, FORMOF, ISA, SYNONYM, DERIVEDFROM, SIMILARTO and ATLOCATION.

exploited in order to compute verb abstractness.

We start by representing the meaning of verbs in terms of their argument distribution, which is common practice in NLP. We followed this intuition: abstract senses are expected to have more abstract dependents than concrete ones. For example, let us consider the verb *drop*. To drop may be —concretely— intended as “to fall vertically”. In this case, it takes concrete nouns as dependents, such as, e.g., in “the bombs are dropping on enemy targets”. In a more abstract meaning to drop is “to stop pursuing or acting”: in this case its dependents are more abstract nouns, such as, e.g., in “to drop a lawsuit”. Although some counterexamples may also be provided, we found that this assumption holds in most cases.

We retrieved the $1,000$ most common verbs from the Corpus of Contemporary American English, which is a corpus covering different genres, such as spoken language, fiction, magazines, newspaper, academic.[6] In order to collect statistics on the argument structure of the considered verbs, we then sampled $3,000$ occurrences of such verbs in the WaCkypedia_EN *corpus*, a 2009 dump of the English Wikipedia, containing about 800 million tokens, tagged with POS, lemma and full dependency parsing (Baroni et al., 2009).[7] All trees containing the verbs along with their dependencies were collected, and such sentences have been passed to the Babelfy API for disambiguation. We retained all verb senses with at least 5 dependents that are present in COVER. The abstractness score of each sense has been computed by averaging the abstractness scores of all its dependents.

4 Evaluation

In order to assess the computed abstractness scores we make use of the Brysbaert Dataset, which is to date the largest corpus of English terms annotated with abstractness scores. It has been acquired through crowdsourcing, and it contains $39,945$ annotated terms (Brysbaert et al., 2014). One chief issue clearly stems from the fact that the human abstractness ratings are referred to terms rather than to senses, which may bias the results of comparisons between the figures used as a ground truth values and the abstractness scores computed by our system. This issue has been experimentally explored in (Mensa et al., 2018a), where different selectional schemes have been tested to pick up a sense from those associated to a given term. The best results, in terms of both Pearson r correlation and of Spearman ρ correlation with human ratings, have been reached by choosing a ‘best’ sense for the term t based on the distribution of the senses associated to t in the *SemCor* corpus (Miller et al., 1993). Specifically, the correlations between the abstractness scores in ABS-COVER and the human ratings in the Brysbaert Dataset amount to $r = 0.653$ and to $\rho = 0.639$.

	MaxAbs	*MinAbs*	*MaxDep*	*BestSns*
Pearson r	0.4163	0.4581	0.5103	0.4729
Spearman ρ	0.4037	0.4690	0.5117	0.4792

Table 1: Correlation results obtained by comparing our system’s abstractness scores against the human ratings in BRYS.

We presently compare the human ratings contained in the Brysbaert *corpus* and the abstractness score associated to one verb sense (corresponding to each lexical entry in the dataset), as computed by our system. We report the correlation scores obtained by selecting the senses based on four strategies:

1. the sense with highest abstractness (*MaxAbs*);
2. the sense with lowest abstractness (*MinAbs*);
3. the sense with the highest number of dependents (*MaxDep*);
4. the sense returned as the best sense through the BabelNet API (*BestSns*).

The obtained results are reported in Table 1. The differences in the scores reported in Table 1 provide tangible evidence that the problem of selecting the correct sense for a verb is a crucial one. E.g., if we consider the verb ‘eat’, the sense described as “Cause to deteriorate due to the action of water, air, or an acid (example: The acid *corroded* the metal)” and the sense described as “Worry or cause anxiety in a persistent way (What’s *eating* you?)” exhibit fully different abstractness characterisation. In order to decouple the assessment of the abstractness scores from that of the sense selection, we randomly selected 400 verbs, and manually associated them with an *a priori* reasonable sense,[8] annotated through the cor-

[6] http://corpus.byu.edu/full-text/.

[7] http://wacky.sslmit.unibo.it/doku.php?id=corpora.

[8] Disambiguation proper would require to select a sense in accordance with a given context.

	FULL-400	Pruning ϑ_1
Pearson r	0.6419	0.6848
Spearman ρ	0.6634	0.6854

Table 2: Correlation scores obtained by manually choosing the main sense for 400 verbs (column FULL-400), and correlation scores obtained by removing from the FULL-400 verbs those with abstractness $\leq .1$ (column ϑ_1 pruning).

responding BabelNet Synset Id. This annotation process is definitely an arbitrary one (only one annotator, thus no inter annotator agreement was recorded, *etc.*), and it should be considered as an approximation to the senses underlying the human ratings available in the Brysbaert *corpus*. The correlation scores significantly raise, as illustrated in the first column of Table 2, thus confirming the centrality of the sense selection step.

Furthermore, we observed that most mismatches in the computation of the abstractness scores occur when the verb is featured by very low (lower than 0.1) abstractness score. To corroborate such intuition, we have then pruned from our data set the verbs whose annotated score is lower than a threshold $\vartheta_1 = 0.1$, finally yielding 383 verbs. In this experimental setting we obtained higher correlation scores, thereby confirming that the computation of more concrete entities needs to be improved, as illustrated in the second column of Table 2.

5 Conclusions

In this paper we have introduced a method to compute verbs abstractness based on the ABS-COVER lexical resource. We reported on the experimentation, and discussed the obtained results, pointing out some issues such as the problem of the sense selection, and the difficulty in characterising more concrete concepts.

As regards as future work, the simple averaging scheme on dependents' abstractness scores can be refined in many ways, e.g., by differentiating the contribution of different sorts of dependents, or based on their distribution. Yet, the set of relations that constitute the backbone of ABS-COVER can be further exploited both for computing the abstractness of dependents, and, in the long term, for generating explanations about the obtained abstractness scores, in virtue of the set of relations at the base of the explanatory power of COVER (Colla et al., 2018). Finally, we plan to explore whether and to what extent our lexical resource can be combined with distributional models, in order to pair those strong associative features with the more semantically structured space described by ABS-COVER.

References

[Baroni et al.2009] Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.

[Benz et al.2011] Dominik Benz, Christian Körner, Andreas Hotho, Gerd Stumme, and Markus Strohmaier. 2011. One tag to bind them all: Measuring term abstractness in social metadata. In *Extended Semantic Web Conference*, pages 360–374. Springer.

[Binder et al.2005] Jeffrey R Binder, Chris F Westbury, Kristen A McKiernan, Edward T Possing, and David A Medler. 2005. Distinct brain systems for processing concrete and abstract concepts. *Journal of cognitive neuroscience*, 17(6):905–917.

[Bolognesi and Steen2018] Marianna Bolognesi and Gerard Steen. 2018. Editors' introduction: Abstract concepts: Structure, processing, and modeling. *Topics in cognitive science*.

[Brysbaert et al.2014] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 46(3):904–911.

[Changizi2008] Mark A Changizi. 2008. Economically organized hierarchies in wordnet and the oxford english dictionary. *Cognitive Systems Research*, 9:214–228.

[Colla et al.2018] Davide Colla, Enrico Mensa, Daniele P. Radicioni, and Antonio Lieto. 2018. Tell Me Why: Computational Explanation of Conceptual Similarity Judgments. In *Procs. of the 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Communications in Computer and Information Science (CCIS), Cham. Springer.

[Coltheart1981] Max Coltheart. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505.

[Ghignone et al.2013] Leo Ghignone, Antonio Lieto, and Daniele P. Radicioni. 2013. Typicality-Based Inference by Plugging Conceptual Spaces Into Ontologies. In Antonio Lieto and Marco Cruciani, editors, *Proceedings of the International Workshop on Artificial Intelligence and Cognition*. CEUR.

[Gibbs Jr2005] Raymond W Gibbs Jr. 2005. *Embodiment and cognitive science*. Cambridge University Press.

[Havasi et al.2007] Catherine Havasi, Robert Speer, and Jason Alonso. 2007. ConceptNet: A Lexical Resource for Common Sense Knowledge. *Recent advances in natural language processing V: selected papers from RANLP*, 309:269.

[Iliev and Axelrod2017] Rumen Iliev and Robert Axelrod. 2017. The paradox of abstraction: Precision versus concreteness. *Journal of psycholinguistic research*, 46(3):715–729.

[Kwong2007] Oi Yee Kwong. 2007. CITYU-HIF: WSD with human-informed feature preference. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 109–112. Association for Computational Linguistics.

[Kwong2008] Oi Yee Kwong. 2008. A preliminary study on the impact of lexical concreteness on word sense disambiguation. In *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation*.

[Lenci et al.2018] Alessandro Lenci, Gianluca E Lebani, and Lucia C Passaro. 2018. The emotions of abstract words: A distributional semantic analysis. *Topics in cognitive science*.

[Lenci2018] Alessandro Lenci. 2018. Distributional models of word meaning. *Annual review of Linguistics*, 4:151–171.

[Lieto et al.2015] Antonio Lieto, Andrea Minieri, Alberto Piana, and Daniele P. Radicioni. 2015. A knowledge-based system for prototypical reasoning. *Connection Science*, 27(2):137–152.

[Lieto et al.2016] Antonio Lieto, Enrico Mensa, and Daniele P. Radicioni. 2016. A Resource-Driven Approach for Anchoring Linguistic Resources to Conceptual Spaces. In *Proceedings of the XVth International Conference of the Italian Association for Artificial Intelligence*, volume 10037 of *Lecture Notes in Artificial Intelligence*, pages 435–449. Springer.

[Lieto et al.2017] Antonio Lieto, Daniele P. Radicioni, and Valentina Rho. 2017. Dual peccs: a cognitive system for conceptual representation and categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(2):433–452.

[Louwerse2011] Max M Louwerse. 2011. Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 3(2):273–302.

[Mensa et al.2018a] Enrico Mensa, Aureliano Porporato, and Daniele P. Radicioni. 2018a. Annotating Concept Abstractness by Common-sense Knowledge. In *Proceedings of the 17th International Conference of the Italian Association for Artificial Intelligence*, LNAI, Cham. Springer.

[Mensa et al.2018b] Enrico Mensa, Aureliano Porporato, and Daniele P. Radicioni. 2018b. Grasping metaphors: Lexical semantics in metaphor analysis. In *The Semantic Web: ESWC 2018 Satellite Events*, pages 192–195, Cham. Springer International.

[Mensa et al.2018c] Enrico Mensa, Daniele P. Radicioni, and Antonio Lieto. 2018c. COVER: a linguistic resource combining common senseand lexicographic information. *Lang Resources & Evaluation*.

[Miller et al.1990] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. 1990. Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244.

[Miller et al.1993] George A Miller, Claudia Leacock, Randee Tengi, and Ross T Bunker. 1993. A semantic concordance. In *Proceedings of the workshop on Human Language Technology*, pages 303–308. Association for Computational Linguistics.

[Navigli and Ponzetto2012] Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

[Neuman et al.2013] Yair Neuman, Dan Assaf, Yohai Cohen, Mark Last, Shlomo Argamon, Newton Howard, and Ophir Frieder. 2013. Metaphor identification in large texts corpora. *PloS one*, 8(4):e62343.

[Newell1980] Allen Newell. 1980. Physical symbol systems. *Cognitive science*, 4(2):135–183.

[Recchia and Jones2012] Gabriel Recchia and Michael Jones. 2012. The semantic richness of abstract concepts. *Frontiers in Human Neuroscience*, 6:315.

[Theijssen et al.2011] DL Theijssen, H van Halteren, LWJ Boves, and NHJ Oostdijk. 2011. On the difficulty of making concreteness concrete. *CLIN Journal*, pages 61–77.

[Turney et al.2011] Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 680–690.

[Turney2006] Peter D Turney. 2006. Similarity of semantic relations. *Computational Linguistics*, 32(3):379–416.

[Xing et al.2010] Xing Xing, Yi Zhang, and Mei Han. 2010. Query difficulty prediction for contextual image retrieval. In *European Conference on Information Retrieval*, pages 581–585.

[Zhu et al.2010] Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd international conference on computational linguistics*, pages 1353–1361. Association for Computational Linguistics.