Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6057–6062
Marseille, 11–16 May 2020

VROAV: Using Iconicity to Visually Represent Abstract Verbs

Simone Scicluna[1], Carlo Strapparava[2]
[1] University of Trento
Trento, Italy
[2] Fondazione Bruno Kessler
Via Sommarive 18, 38123 - Trento, Italy
simonescicluna@gmail.com, strappa@fbk.eu

Abstract

Abstractness is a feature of semantics that limits our ability to visualise every conceivable concept represented by a word. By tapping into the visual representation of words, we explore the common semantic elements that link words to each other. Visual languages like sign languages have been found to reveal enlightening patterns across signs of similar meanings, pointing towards the possibility of identifying clusters of iconic meanings in words. Thanks to this insight, along with an understanding of verb predicates achieved from VerbNet, this study produced VROAV (Visual Representation of Abstract Verbs): a novel verb classification system based on the shape and movement of verbs. The outcome includes 20 classes of abstract verbs and their visual representations, which were tested for validity in an online survey. Considerable agreement between participants, who judged graphic animations based on representativeness, suggests a positive way forward for this proposal, which may be developed as a language learning aid in educational contexts or as a multimodal language comprehension tool for digital text.
Keywords: Abstract verbs, verb classification, sign language, language comprehension, visual shape of words, symbolic modeling, computer-aided learning (CAL).

1. Introduction

The way we see the world can be expressed using a myriad of words offered by the vocabulary of the language we speak in. Language vocabularies leave us with an infinity of choices to make and this is largely due to the fact that we have different words with similar meanings. While enriching our views of the world, this characteristic also makes language an incredibly complex skill. Children start to acquire the ability to label objects around them when young, and gradually, expand their lexicon based on their lived experiences. This lexicon continues to expand as they learn new ways of expressing varying degrees of the same concept - an idea better known as synonymy. This skill allows humans to express themselves in more complex ways, such as in communicating ideas, opinions and arguments. This inevitably introduces a greater use of abstract words, which comes with challenges such as the lack of reliance of mental images for words and concepts.

In the current study, the latter claim is not taken as a limitation, but rather as a motivation to delve into the existing structures humans may be utilising to more efficiently access these concepts from memory, even when they are abstract. So far, brain research has offered insight on the fact that we develop mental images for words learned. However, what this fails to tell us is how they would look like visually. One way of obtaining this is to observe how humans react to external visual stimuli and make inferences accordingly. This has been largely studied for concrete words, but interestingly, less so for abstract words, as these have less obvious visual structures. The current study supports the claim that visual structures are encoded in the mind for abstractness too, and that these might not be entirely subjective. We turn to sign language to discover the common visual elements of these structures and prove the hypothesis that abstractness poses no limitation when it comes to the visual representation of words.

2. Background

Literature on the topic of the current study spans across many areas, including verb classification, semiotics, sign language and learning. First of all, it is worth highlighting the reason we chose to focus on verbs. Relative to nouns, verbs are found to be more difficult to learn and this explains why children grasp them later than nouns (Gentner, 2006). In fact, children's vocabularies contain significantly more nouns than they do verbs in proportion (Fenson et al., 1994; Nelson, 1973; Stern, 2017). Reasons for this include (i) the fact that verbs can be seen as predicates that take noun phrases as arguments, which implies that a child must first learn the names of objects before combining these with relational terms (Sandhofer et al., 2000), and (ii) the fact that verbs are less likely to have concrete associations to the world than nouns have, and this refers to the *Natural Partition Hypothesis* (Gentner, 2006).

Verbs contain a lack of perceptual and conceptual links, which entails that it must be harder to develop mental images that represent them. According to the *dual coding theory* (Clark and Paivio, 1991), the formation of mental images aids in learning, although this advantage may not be available for abstract words as these are less likely to have corresponding word-to-image referential connections. As a result, abstract words can be more challenging to learn and memorise. Given this path of reasoning, this study has taken the challenge of finding ways to create visual representations of abstract verbs to strengthen their link to the world, and hence potentially contribute to better memorisation of them in the mind.

One of the initial hypothesis of this study was that abstract language is more strongly associated with higher levels of language production. The BNC (BNC, 2007) proves that

the top 500 most frequent verbs used in academic texts are, by and large, abstract. This revelation, tied with the literature previously mentioned, leads us to understand that abstract verbs dominate in advanced genres of writing such as scientific, argumentative, critical and evaluative texts, in which "argument predominates over description" (Power, 2007). Verbs used in such genres of writing are referred to as *discourse verbs* (Danlos, 2006), and they include verbs like: *express*, *suggest* and *involve*.
The current study uses a combination of graphic symbols, animation and sign language to contribute towards tackling the following question: how can we represent abstract verbs visually? The study of sign language was given priority due to its excellent mapping from meaning to form, using a wide vocabulary of signs that enables the interpretation of any possible universal concept. This mapping is referred to as a *formational link* (Brennan and Brien, 1992) , pointing towards the highly symbolic value that signs hold, also known as iconicity. As sign language demonstrates, this link also exists for abstract concepts that have no real referents in the physical world, suggesting that **even abstractness can be iconic**. This phenomenon makes sign language indispensable to the objectives of the current study, and we therefore used it as our main inspiration for developing graphic representations of abstract verbs classes.
Research shows that gestures have been useful in enabling successful communication in educational contexts as their iconic nature provides more accessibility to the "form" of meaning (Thompson et al., 2012). For instance, Makaton (Grove and Walker, 1990) is a system of manual signs and graphic symbols proven to enhance learning and communication. The use of gestures in education has also been empirically tested, demonstrating that we remember words better with the use of gestures than without, even when they are abstract (Macedonia and Knösche, 2011). This goes in line with the view of *embodied cognition*, which may also enhance the learning of new languages by reinforcing the sensorimotor representation of a concept (Macedonia and von Kriegstein, 2012). Perhaps the use of gestures reveals that a mental image is being employed in the memory retrieval process, which has actually been found to reinforce the links between L1 and L2 in bilingual brains (Repetto et al., 2017). This suggests that strengthening mental images could improve efficiency of learning and retrieval of new words, and this is a potential application of the resources produced from this study.

3. Methodology

The aim of this study was to achieve graphic representations for abstract verbs and confirm their communicative meaning through naïve human judgements. The methodology employed was motivated by Richardson et al. (2001), a study in which four images depicting the four directions (up, down, left and right) were created in association with concrete and abstract words, proving the hypothesis that "language is spatial".

3.1. Data preparation

Two methods were attempted in the data preparation stage of this study. First, a list of 150,114 words of mixed classes rated by abstractness and imageability were extracted from the MRC psycholinguistic database (Coltheart, 1981) and organised by Tsvetkov et al. (2014) with the intention of applying a top-down approach to narrow this down to our target verbs. Elimination was done in order to obtain a list of words that were only abstract verbs, by tagging with WordNet. Then, weighting was given to each verb to prioritise on high synonymy count, high polysemy count, high abstractness rating, inclusion in the Academic Word List (Coxhead, 2000), and inclusion in the top 500 verbs used in academic texts on the British National Corpus. After being weighted and sorted, the top 500 verbs were passed through clustering techniques like Latent Dirichlet Allocation (Blei et al., 2003), Latent Semantic Analysis (Landauer and Dumais, 1997) and Word2Vec (Mikolov et al., 2013). The motive for this was to cluster verbs by visual shape, however this was not successful. The outcomes showed that non-synonymous words were clustered together (e.g. *expand* and *focus*), as well as antonyms (e.g. *separate* and *connect*). This goes against the objectives of this study, since the differences would result in completely unrelated, and at times opposing, visual shapes.

The second method was to apply a bottom-up approach. Rather than applying elimination on a large list of words, we started building our list from scratch. We considered VerbNet's collection of verb predicates (Schuler, 2005) and inferred verbs that behave in the same way, and therefore would be more likely to be clustered together. Distinct predicate groups were identified and abstract verbs were drawn from them. These were searched through sign language dictionaries like SpreadTheSign (`www.spreadthesign.com`) and BSLSignBank (`www.BSLSignBank.ucl.ac.uk`) to look for the similarities of equivalent signs across sign languages and across sign neighbours. Generally, cases of considerable cross-linguistic similarity between signs were preferred, due to the advantage of universality. This strengthened and improved the grouping done through VerbNet, always bearing visual shape in mind.

To exemplify the process, let us take the class FUSE (see Figure 1). This consisted of verbs like *meet, unite, match, collect, integrate, associate, combine,* to mention a few, most of which belong to the same VerbNet classes such as "amalgamate-22.2-2", "harmonize-22.6" and "mix-22.1", as well as the same VerbNet predicates like "harmonize", "together" and "mingled". Similarly, many of these words have common visual patterns in the respective signs across sign languages - namely, the element of both hands coming together in the centre. The basis of the FUSE class, as with any other class, was formed by manually grouping together abstract verbs that appeared in the same or similar contexts with regards to VerbNet's predicates and sign neighbours, as both kinds of mappings suggest semantic closeness. Synonyms of members were added in order to populate the classes. These did not necessarily have sign equivalents or VerbNet relations, but they communicated the same meaning and therefore fitted with the respective class adequately. It's important to note that not all words belonging to the same class reflected in identical signs in sign language; at times these were similar or even

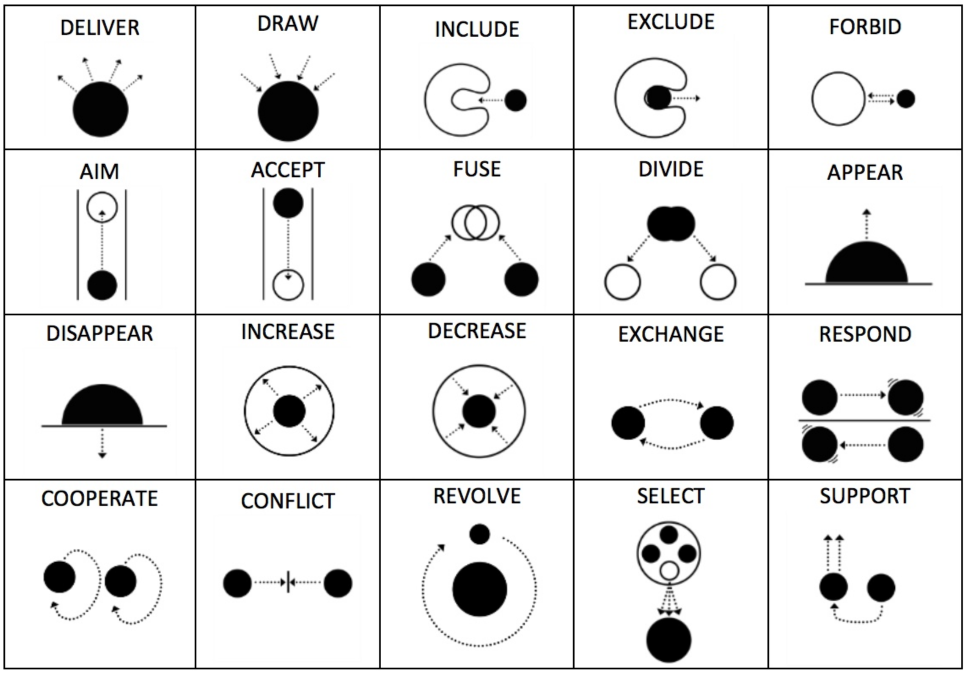

Figure 1: The 20 classes of abstract verbs created

completely dissimilar, but VerbNet predicates or synonymy provided us with an alternative way to link those verbs together. The creation of the graphic visuals was motivated by the most common sign language elements observed for verbs in the same class, which in the case of FUSE translates to two objects coming together and becoming one in the centre. Class names were given by choosing the most generic word member that suits the common meaning, as well as the visual.

VerbNet together with sign language provided for an excellent combination of lexical resources that made it possible to produce a collection of roughly 500 abstract verbs, classified into 20 different verb classes, each represented by distinct visual shapes. Figure 1 displays all abstract verb categories with their corresponding graphic visual. The graphic visuals were made up of basic colours and shapes: black and white circles and lines expressing an interaction. Initially, the designs were static and used arrows to denote movement. After running a pilot study on these designs, they were given animation as well as other visual improvements by a graphic artist, substituting the use of arrows with motion. The motivation for using animation was to maximise the resemblance with sign language.

The process of interpreting signs into iconic graphic visuals is illustrated in Figure 2, which displays the BSL signs for groups of verbs that acted as bases for classes APPEAR and INCLUDE. Sign elements were discussed and visualised together with the graphic artist, ensuring that the interpretation the signs was as clear-cut as possible. Interesting observations were made during this stage, such as the fact that even non-synonymous words such as *happen* and *create* shared sign elements, suggesting underlying common semantics.

The hypothesis is that, although minimal, the graphic visuals would be informative enough to distinguish verb classes from each other and communicate the meaning of a word when given a choice or the appropriate context. To see the animated visuals along with the complete collection of abstract verbs visit `http://vroav.online`, which contains all the information linked to the current project.

3.2. Data collection

All 20 animated visuals developed in this study were used in an online survey (`vroav.online/survey`). The survey consisted of only 20 questions, each of which consisting of one animated visual along with four verb choices. Participants were asked to select the one that best describes the visual. The questions were designed in such a way that each one included one valid verb picked at random from the valid class and three invalid verbs picked at random from the invalid classes. The order of the images displayed throughout the survey was randomised and there was always one valid verb displayed in every multiple choice question. This feature made the survey unique for each participant, keeping bias at a minimum. The survey was completed by approximately 100 participants. Nationality as well as level of English (intermediate, fluent and native) was also recorded from each participant. No time constraints were used. In total, the number of individual image judgments collected was 2,010.

4. Results

Results show an average correctness percentage of 61.4% across trials and 62.5% across verb categories (sd 12.3). The DECREASE class ranked highest with 82.7% correctness, followed by CONFLICT and FORBID; and the DIVIDE class ranked lowest with 41.6% correctness, preceded by SELECT and DRAW. A set of 16 classes that

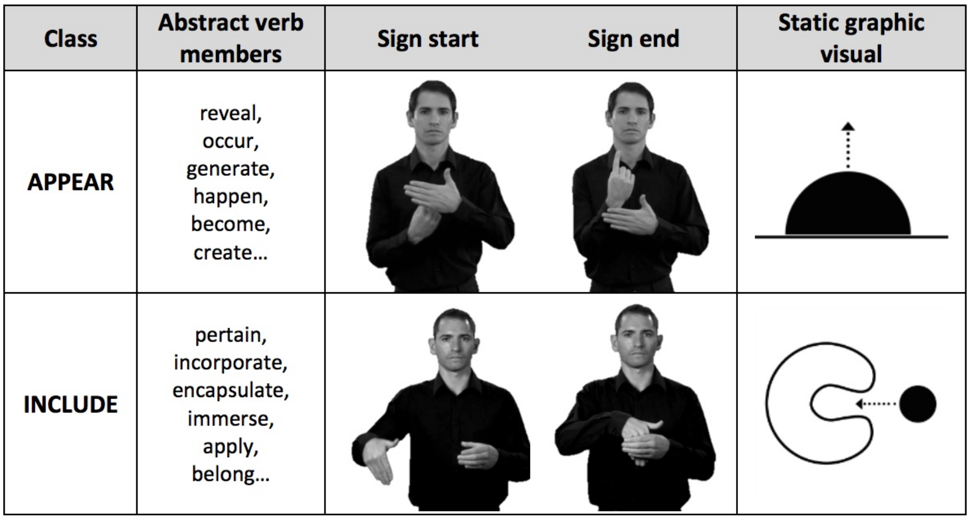

Figure 2: The graphic visual for the class APPEAR as inspired by several sign languages that all display the same hand movements (sign images from `bslsignbank.ucl.ac.uk`)

got an average correctness score of 50% and above makes up 80% of total number of classes. Only 4 classes (20% of total classes) scored below 50%, and these were DIVIDE, SELECT, DRAW and RESPOND. This indicates that there was significant agreement in the way subjects associated abstract verbs to their assigned graphic representations. Considering that we were dealing with abstract verbs, which are normally linked to advanced genres of writing, and with images that the participants had never seen, the results are rather compelling, indicating that abstract words still exhibit agreement in the judgment of potential graphic representations. A more generic reorganisation of the verb classes later resulted in an increase of 8.2% in correctness (70.7%), with the highest average correctness across classes being 78%, raising more interesting points for discussion.

5. Discussion

Further speculation on the recorded responses was made in order to explore other aspects of the results or potential factors that led to them. We tested whether nationality and level of English had an effect on the overall survey responses. Correctness across nationalities resulted in 62% (range 55-75%), and correctness across levels of English resulted in 60.7% (range 58-63%). This implies that none of these variables seem to have influenced the result for correctness.

Two other variables that were suspected to potentially have had an effect on the results were word frequency and degree of abstractness. The distribution of word correctness against frequency was analysed for each category. This showed that there was a higher tendency for lower frequency words to achieve lower correctness than higher frequency words did, however this effect was not significant enough since some low frequency words also achieved high correctness. Distribution of word correctness against degree of abstractness was also analysed. This displayed a low correlation, suggesting that verbs of high abstractness performed just as well as verbs of lower abstractness. This also applies to incorrectness. Therefore, we cannot assume a positive correlation between concreteness and correctness.

Another possible variable could be the variance between categories and their members. Could the results have improved if certain words were added or excluded, or shifted across categories? Can we give linguistic or semantic explanations to justify why certain words achieved higher correctness than others? Or perhaps, could this study enlighten us about the words that are more visually congruent and incongruent with each other? The confusion matrix in Figure 3 shows that there is a dominance of correctness in every class. The incorrectness displayed by each class is shared with 10 to 19 other classes and makes up 70% of the matrix. Only 10% of this proportion had a convergence score of 5 or higher, which means that most cases actually displayed very low convergence. Class pairs with no convergence at all make up 25% of the matrix.

From our analysis, it has been confirmed that (i) participant context (nationality, level of English) and (ii) lexical context (word frequency, degree of abstractness) had no observed effect on the results. The next obvious context to scrutinize is, of course, the visual context. When analysing the results, it was observed that a subset of the verb classes tended to be confused with certain others, thus forming clusters of confused classes. Once a number of "high confusion clusters" were identified and studied, an interesting pattern was found. The graphic visuals of frequently confused clusters had similarities in visual shape, so much so that three common visual elements could be established: *direction* (upward or downward), *movement* (“coming together" or “coming apart") and *contact* (physical impact). With these common form elements identified, we could justify why some classes tended to behave similarly in terms of responses. This similarity is also observed in the signs used as inspiration for the visuals. The findings go to show that form elements in the visual representations did reflect in

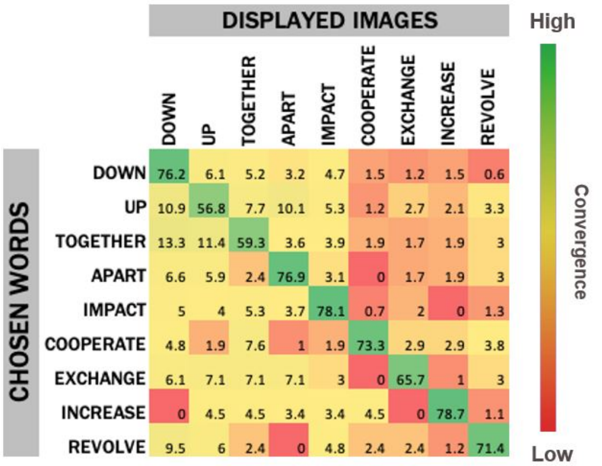

Figure 3: Confusion matrix of the results, showing convergence between classes. Classes displayed here are a result of a reorganisation of the 20 verb classes in Figure 1, where classes sharing common visual elements were clustered into parent classes, as seen in Figure 4.

participants' judgment, but fine-grained visual details were disregarded. Grouping the words from these high confusion clusters together leaves us with 6 verb categories instead of 20, and the correctness average resulting from this reorganisation is 70.7% (8.2% higher than the original), suggesting that 20.8% of incorrect responses could have been justified by visual similarity. Figure 4 displays the 6 new clusters produced after the reorganisation of the classes based on their confusion with one another.

DIRECTION		MOVEMENT			CONTACT
UP	DOWN	TOGETHER	APART	CIRCULAR	IMPACT
APPEAR	DISAPPEAR	FUSE	DIVIDE	COOPERATE*	CONFLICT
DELIVER	DRAW	INCLUDE	EXCLUDE	EXCHANGE*	RESPOND
AIM	ACCEPT	SUPPORT	SELECT	REVOLVE*	FORBID
INCREASE*	DECREASE				

Figure 4: The new reorganisation of the classes resulting from clustering high confusion classes together based on their co-occurrence of visual shape. (* The asterisk marks low confusion classes. Most of these fall into the MOVEMENT-CIRCULAR class. The rest of the classes are high confusion classes.)

The results from the study reveal positive implications towards our hypotheses, which were: (i) that sign language is a successful basis for the classification of verbs by visual shape, and (ii) that we can represent abstract words in graphic visuals that are informative enough to communicate abstract word meanings. Furthermore, we have identified three generic form elements common in 80% of our classes, which are direction, movement and contact, and which contributed to confusion between our least distinctive verb classes. The latter is not seen as a confounding variable, but rather, an indication of the spatial elements involved in the semantics of words, and essentially, the spatial elements that make up our representation of concepts.

6. Conclusion

As literature states, verbs are harder to conceptualise relative to other word classes, and abstract verbs even harder. Through this study, we have attempted to come closer towards finding out how abstract verbs map to visual shape, paving the way towards a system that can represent them visually. Our 20 verb classes and visual representations cover only 500 abstract verbs, and the possibility of having more is not excluded. However, it may be difficult to do this for all abstract verbs in the English vocabulary. We have only picked verbs that are interpreted as iconic signs in sign language, and the current study by no means presents a conclusive, fully-encompassing system. With further research, the number of verb classes based on visual shape could increase, and potentially, even the number of verbs they cover.

A more neuroscience-related possibility for future research is to test what effect these visual representations have on memory. This could be investigated in a behavioural experiment to find out whether words accompanied by the images produced by this study aids the retention of words, which could confirm the *multimedia learning theory* (Mayer, 2005) as well as the *dual coding theory* (Clark and Paivio, 1991), applied to abstract words. Furthermore, we could explore whether the method of using multimedia with new words could facilitate first or second language learning amongst adults. The creation of this system may also inspire methods employed in language comprehension when reading digital text, with which pictorial representations could potentially be integrated as an aid for understanding. This may also be seen as additional supralinguistic value added to digital text, which generally lacks it.

Overall, the current study offers a positive way forward for the use of multimedia associated to words. A strong point has been made with regards to the correctness of human judgment on the word-image association for abstract verbs, which was not only found to be high across verb classes, nationalities, levels of English, word frequencies and abstractness, but even higher when clusters of verb classes sharing common visual elements were formed. With these findings being proven, we can be more confident in saying that language is indeed spatial, abstract words can be successfully represented visually, and sign language serves as a bridge to make this possible.

7. References

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(3):993–1022.

(2007). The British National Corpus. University of Oxford.

Brennan, M. and Brien, D. (1992). *Dictionary of British Sign Language/English*. Faber and Faber.

Clark, J. and Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3):149—210. Retrieved from http://link.springer.com/article/10.1007/BF01320076.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33(4):497–505.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2):213–238.

Danlos, L. (2006). Discourse verbs and discourse periphrastic links. In *Proceedings of Constraintsin Discourse'06*, Maynooth, Ireland.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5):1–185.

Gentner, D., (2006). *Action Meets Word: How Children Learn Verbs*, chapter Why Verbs Are Hard to Learn, pages 544–564. Oxford University Press.

Grove, N. and Walker, M. (1990). The makaton vocabulary: Using manual signs and graphic symbols to develop interpersonal communication. *Augmentative and Alternative Communication*, 6(1):15–28.

Landauer, T. K. and Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.

Macedonia, M. and Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind Brain and Education*, 5(4):196–211.

Macedonia, M. and von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, 6(3-4):393–416.

R. E. Mayer, editor. (2005). *Cognitive Theory of Multimedia Learning*. Cambridge University Press, New York, USA.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, et al., editors, *Advances in Neural Information Processing Systems*, volume 26, pages 3111–3119. Curran Associates, Inc.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, 38(1-2):1–135.

Power, R. (2007). Abstract verbs. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG '07)*, pages 93–96, Schloss Dagstuhl, Germany, June.

Repetto, C., Pedroli, E., and Macedonia, M. (2017). Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in Psychology*, December.

Richardson, D. C., Spivey, M. J., Edelman, S., and Naples, A. J. (2001). "language is spatial": Experimental evidence for image schemas of concrete and abstract verbs. In *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, Edinburgh, Scotland.

Sandhofer, C. M., Smith, L. B., and Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language*, 27(3):561–585, October.

Schuler, K. K. (2005). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Stern, W. (2017). *Psychology of early childhood: up to the sixth year of age*. Routledge.

Thompson, R. L., Vinson, D. P., and Vigliocco, G. (2012). The road to language learning is iconic: evidence from british sign language. *Psychological Science*, 23(12):1443–1448, December.

Tsvetkov, Y., Boytsov, L., Gershman, A., Nyberg, E., and Dyer, C. (2014). Metaphor detection with cross-lingual model transfer. In *Proceedings of ACL 2014*, pages 248–258. Association for Computational Linguistics.