Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Note

Atypical associations to abstract words in Broca's aphasia

Mikael Roll [a,*], *Frida Mårtensson* [a], *Sverker Sikström* [b], *Pia Apt* [c], *Rasmus Arnling-Bååth* [b] and *Merle Horne* [a]

[a] *Department of Linguistics and Phonetics, Lund University, Sweden*
[b] *Department of Cognitive Science, Lund University, Sweden*
[c] *Department of Neurology, Malmö, Skåne University Hospital, Sweden*

ARTICLE INFO

Article history:
Received 19 January 2010
Reviewed 10 May 2010
Revised 10 March 2011
Accepted 15 November 2011
Action editor Rhonda Friedman
Published online 23 November 2011

Keywords:
Aphasia
LSA, Latent semantic analysis
Concreteness
Abstractness

ABSTRACT

Introduction: Left frontal brain lesions are known to give rise to aphasia and impaired word associations. These associations have previously been difficult to analyze. We used a semantic space method to investigate associations to cue words. The degree of abstractness of the generated words and semantic similarity to the cue words were measured.
Method: Three subjects diagnosed with Broca's aphasia and twelve control subjects associated freely to cue words. Results were evaluated with latent semantic analysis (LSA) applied to the Swedish Parole corpus.
Results: The aphasic subjects could be clearly distinguished from controls by a lower degree of abstractness in the words they generated. The aphasic group's associations showed a negative correlation between semantic similarity to cue word and abstractness of cue word.
Conclusions: By developing novel semantic measures, we showed that Broca's aphasic subjects' word production was characterized by a low degree of abstractness and low degree of coherence in associations to abstract cue words. The results support models where meanings of concrete words are represented in neural networks involving perceptual and motor areas, whereas the meaning of abstract words is more dependent on connections to other word forms in the left frontal region. Semantic spaces can be used in future developments of evaluative tools for both diagnosis and research purposes.

1. Introduction

1.1. *The concrete brain*

Both abstract and concrete words can be understood on the basis of their relation to other words derived from the discourse contexts they have been experienced in (Elman, 1991). Concrete words in addition rapidly activate sensory and motor networks engaged in perceiving or executing their referred object or action (Pulvermüller and Fadiga, 2010). This is a possible reason for the 'concreteness effect' (Paivio, 2010), which refers to the fact that concrete words are processed faster and more accurately than abstract words (De Groot, 1989; Kroll and Merves, 1986).[1] Abstract words are less

* *Corresponding author*. Department of Linguistics and Phonetics, Lund University, Box 201, 22100 Lund, Sweden.
E-mail address: mikael.roll@ling.lu.se (M. Roll).

[1] We consider rapid, pre-attentive activation of sensorimotor regions to distinguish concrete word meaning from abstract rather than the imagery of the *dual coding theory* (Paivio, 2010). Automatic semantic activations fade away quickly (Shtyrov, 2010), and are thus likely to influence online production and perception of words, but not necessarily their recall (cf. Schwanenflugel et al., 1992).

0010-9452/$ – see front matter © 2011 Elsevier Srl. All rights reserved.
doi:10.1016/j.cortex.2011.11.009

Table 1 – Results from diagnostic testing using PAPAP (Apt, 1997).

Subtest	Maximum	Cut-off	Subject 1	Subject 2	Subject 3
A. Auditory language comprehension	183	178	162	157	176
B. Repetition	196	178	143	151	145
C. Oral naming and narration	140	128	104	96	124
D. Digits and arithmetic	123	118	95	103	122
E. Writing	276	269	213	179	264
F. Reading	192	191	149	151	187
Total	1110	1062	866	837	1018

straightforwardly associated with sensory and motor representations, but instead increase activation in the frontotemporal language areas of the left hemisphere, indicating a greater dependence on other word forms for their interpretation (Binder et al., 2006; Just et al., 2004; Kounios and Holcomb, 1994; Sabsevitz et al., 2005). If concrete word meaning relies partly on widely distributed sensorimotor representations, lesions restricted to core language areas should affect abstract word meaning more than concrete. In the present study, cued associations of participants diagnosed with Broca's aphasia were measured for abstractness as well as for semantic similarity as a function of the cue words' abstractness.

Studies of non-fluent aphasics' word associations have found atypical or no responses to cue words that have large numbers of associates in an association norms database (Wyke, 1962). These cue words have been characterized as "abstract" (Lesser, 1973). However, response words have been difficult to analyze since they are often neither among the cue words' database associates, nor do they seem purely idiosyncratic (Gewirth et al., 1984). Measuring semantic similarity in a semantic space model is a possible solution to this methodological problem (Church and Hanks, 1990). Further, semantic models have also proven to be useful for distinguishing between abstract and concrete words (Audet and Burgess, 1999). In this study, we used a semantic space model to assign abstractness values to cue and response words, and to measure how the semantic strength of the associations changed as a function of the cue word abstractness. Our first hypothesis was that aphasic participants would produce response words with lower abstractness than the control subjects. The reason would be better access to concrete words due to undamaged perceptual and motor areas where part of their meaning is represented. We further hypothesized that abstract cue words would be more difficult for the aphasic group to understand, and thus lead to associations with lower semantic strength.

1.2. *Semantic spaces*

Semantic space models are methods for deriving measures of semantic similarity between words based on their occurrence in the same sort of contexts in large text corpora (Lund and Burgess, 1996). A semantic space model that has been successful in modeling human associations is latent semantic analysis (LSA) (Landauer et al., 2007). In this model, a word by context frequency table is first created. The high number of contexts (typically in the order of millions) is then reduced to a limited number of semantic dimensions (usually in the order of hundreds) by applying a data compression algorithm called Singular Value Decomposition (SVD). This dimension reduction generalizes the information from specific contexts to capture more general semantic traits of each word. The result is that each word can be placed in a multi-dimensional semantic space where words with similar meanings are located near each other. Semantic similarity is measured through the geometric distance between two vectors in the space.

2. Method

2.1. *Subjects*

Three Swedish females diagnosed with Broca's aphasia participated. Diagnosis was based on results from a validated aphasia test standardized for Swedish, Pia APts AfasiPrövning 'Pia Apt's aphasia examination' (PAPAP) (Table 1) (Apt, 1997), carried out by an authorized speech and language pathologist specialized in neurological communication disorders. The subjects were chosen due to their stable neurological condition with moderate to mild aphasia, which made participation in the current study possible. They had regular contact with the language pathologist both before and after taking the PAPAP diagnostic test. Continuous observation and testing, using parts of PAPAP and other standardized tests, showed that the initial diagnosis and degree of Broca's aphasia had not changed when the current study was carried out.

Participant 1 (age 66, junior high school education) had a cerebral infarction on Oct. 8, 2004. X-ray computed tomography (CT) on Oct. 27, 2004 showed low attenuation mainly in the left frontal cortex. The diagnosis following behavioral testing on Oct. 6, 2005 was mild to moderate Broca's aphasia. Her speech was non-fluent, telegraphic, and moderately dyspraxic, displaying perseverations, dysgrammatism, and verbal amnesia, affecting verbs more than nouns. She had mild difficulty comprehending long or grammatically complex sentences, and moderate impairment of auditory verbal short-term memory. She had phonological dyslexia. Writing was impaired for longer words and sentences, but improved with shorter words as long as they were orthographically transparent.

Participant 2 (age 41, high school education) had a cerebral infarction on Apr. 25, 1989. CT on Aug. 22, 2000 displayed left frontoparietal low attenuation with largest extension frontally. Behavioral tests on Mar. 27, 1998 resulted in diagnosis of mild to moderate Broca's aphasia. She had slow, non-fluent, telegraphic speech, manifesting particular difficulty finding and inflecting verbs and producing wh-clauses. Function word production and syntax were also impaired. Perseverations occurred with increased cognitive load due to fatigue, stress, or task difficulty. She showed mild difficulty comprehending long or complex spoken sentences. Short, common words were unproblematic to write, but in longer words and sentences, often only the first letters were correct. Reading was slow, involving the same impediments as her speech comprehension.

Participant 3 (age 31, high school education) suffered a cerebral infarction on Feb. 17, 2007. CT on Feb. 18, 2007 revealed low attenuation temporoparietally in the left hemisphere, beginning at the middle cranial fossa and extending cranially dorsally over the Sylvian fissure at the height of the cella media of the lateral ventricles. The low attenuation also included the insula area and the major part of the external capsule. Magnetic Resonance (MR) on Feb. 20, 2007 revealed lesions in large parts of the irrigation areas of the middle cerebral artery, both frontally and temporally. The diagnosis following behavioral testing on May 21, 2007 was mild Broca's aphasia. Her speech was moderately dyspraxic, slow and articulatorily strained, especially as regards words that were long or difficult to pronounce. Her spontaneous speech and descriptions of pictures and series of pictures were characterized by word fluency difficulties, with verbs more affected than nouns. She had mild dysprosodia and dysgrammatism, involving slight impairments in sentence formulation and function word production. Her arithmetic skills were on a normal level. Reading and writing impairments were on a par with spoken language impediments.

Control subjects ($N = 12$, mean age: 47, $range = 23$ to 79, 10 females) were healthy and matched for educational level. All subjects gave their written informed consent to participate.

2.2. Procedure and materials

Thirty concrete and 30 abstract nouns were chosen from the Medical Research Council psycholinguistic database (Coltheart, 1981) and translated to Swedish. The nouns were chosen based on their high, $M = 605$, $SD = 26$, or low, $M = 277$, $SD = 38$, concreteness rating on a 100–700 scale, $t(58) = 36.7$, $p < .001$. The log frequency of the abstract nouns, $M = 2.51$, $SD = .52$, was higher than that of the concrete, $M = 1.56$, $SD = .66$, $t(58) = 6.17$, $p < .001$, in the Parole corpus (http://spraakbanken.gu.se/parole/).

Participants were instructed to say the first word that came to mind when hearing the cue word. The experiment leader presented the cue words orally one by one. Response times were measured from recordings from the point when the experiment leader had finished uttering the whole cue word to the point when the participant began to utter the response word.[2]

2.3. LSA

Infomap 0.8.6 was used on the Parole corpus to create a semantic space. Parole is a 19.4-million word corpus consisting of 23% novels, 70% newspaper text, 2% popular science magazine texts, and 5% Internet material. Infomap created a co-occurrence matrix that was decomposed using the SVDPACKC implementation of SVD (Berry et al., 1993). The rows in the matrix represented the 400 k most frequent words in the corpus, and the columns represented the contexts of ±15 words around the 1000 most frequent words. The space was decomposed to 100 dimensions through 100 SVD iterations. 15,026 words, consisting of the most frequent words in the space and the cue words, were transferred to LSALAB (http://www.lucs.lu.se/sverker.sikstrom/), where the analyses were performed.

2.4. Abstractness values and data analysis

Abstractness values were generated for both cue and response words based on the semantic space model in the following way: A norm of 1400 words was used to define the concept of abstractness versus concreteness. This norm was based on subjects ranking words on a scale from concrete (1) to abstract (9). Multiple linear regression was used to predict the ranked abstractness of the words. A leave-one-out cross-validation of this regression showed that it was possible to predict concreteness for the word that was left out, $r = .42$, $p < .001$. The dimensions in the semantic space functioned as independent predictors in the regression. The obtained regression coefficients were used to generate predicted abstractness values that could be generalized to all the words in the semantic space. This predicted abstractness constituted the abstractness measure in the current study. Predicted abstractness for response words was compared between groups. Correlation between abstractness of cue words and semantic distance between cue and response words was measured in order to obtain a measure of how abstract words

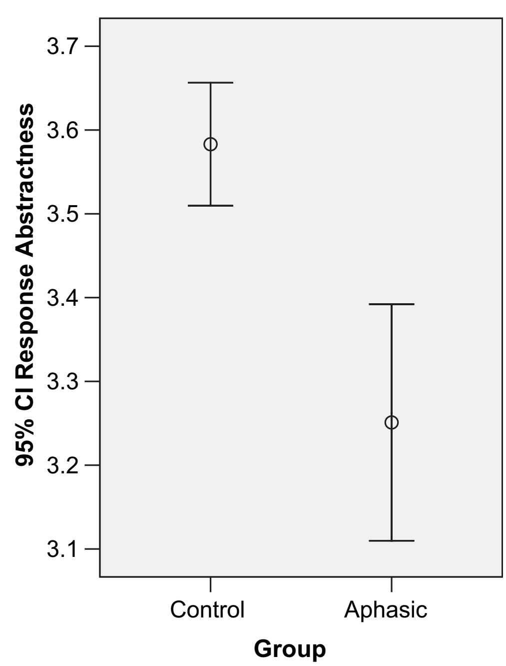

Fig. 1 – Mean abstractness of response words calculated using the semantic space model. There was no overlap between groups as regards the individual means.

[2] Response time, response abstractness, and semantic strength have different Ns because some words were not available for LSA in the corpus.

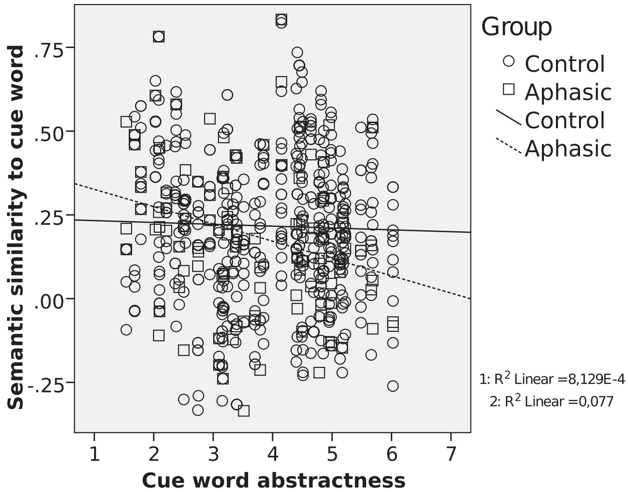

Fig. 2 – Linear regression for associations of aphasic and control subjects. In the aphasic group, semantic similarity decreased as the cue words became more abstract.

affected the semantic strength of word associations. Furthermore, abstractness was correlated with response times, making it possible to detect a concreteness effect. Correlations and comparison of means were evaluated with two-tailed *t* tests.

3. Results

Aphasic participants produced response words with lower abstractness than controls, $t(788) = 3.73$, $p < .001$ (Fig. 1). Average abstractness was 3.25 for the aphasic group, $N = 133$, $SD = .82$, and 3.58 for the control group, $SD = .96$, $N = 657$. There was no overlap in individual means between groups, since the range was 3.15–3.29 for aphasic subjects and 3.42–3.64 for controls.

In the aphasic group, there was a negative correlation between semantic similarity and cue word abstractness, $r = -.278$, $p = .003$, absent in the control group, $r = -.029$, $p = .488$ (Fig. 2). An ANOVA was performed with the factor abstractness (abstract, concrete) based on MRC ratings of the cue words. There were no main effects of abstractness, $p = .452$, or group, $p = .302$. However, an abstractness × group interaction, $F(1, 676) = 7.44$, $p = .007$, and Bonferroni-corrected posthoc comparison revealed lower semantic similarity for associations to abstract cue words in the aphasic group, $M = .149$, $SD = .23$, than in the control group, $M = .238$, $SD = .23$, $p = .006$. There was no significant difference for concrete cue words, $p = .250$.

Correlations between response time, semantic similarity to cue word, and abstractness of cue and response words for the aphasic and control groups are shown in Table 2. In both groups, response time increased together with increasing abstractness in both cue words, $r = .353/.351$ (aphasic/control), $p < .001$, and response words, $r = .296/.244$, $p < .01$. This is the expected concreteness effect. Abstract responses further correlated with abstract cue words, $r = .505/.467$, $p < .001$. Finally, there was a weak tendency for abstract responses to show increased semantic similarity to cue words in the control group, $r = .114$, $p = .007$. The aphasic group had longer response times than controls, $M = 7.85/4.39$ sec, $SD = 10.61/7.98$, $N = 169/715$, $t(882) = 6.95$, $p < .001$ (*t* calculated on logarithmic response time).

Response words had a higher log frequency in the Parole corpus for the control group, $M = 2.48$, $SD = .90$, $N = 690$, than for the aphasics, $M = 2.24$, $SD = .90$, $N = 153$, $t(841) = 2.93$, $p = .004$. Control subjects' responses showed a weak correlation between abstractness and frequency, $r = .14$, $p < .001$, absent in the aphasic subjects, $r = -.003$, $p = .969$.

4. Discussion

Associated words' degree of abstractness and semantic similarity to cue words were measured for subjects diagnosed with Broca's aphasia using a semantic space model. Overall, the aphasic group produced words with a lower degree of abstractness than the control group. In fact, based on individual means, all aphasic participants could be distinguished by the concreteness in their responses without overlap with the control group. Surprisingly, there was no significant main difference between groups in the response words' semantic similarity to the cue words. However, aphasic participants' responses had lower semantic similarity the more abstract the cue words were. Dividing the cue words into an abstract and a concrete group showed lower semantic similarity in aphasic associations only to abstract cue words. There was a weak correlation between frequency and abstractness in the control group's responses, congruent with the tendency for abstract words to be more frequent (Audet and Burgess, 1999). The lack of such a correlation among the aphasic participants and their slightly lower response word frequency are probably consequences of their avoidance even of common abstract words.

The results show impaired comprehension and production of abstract words in Broca's aphasia. This could possibly be

Table 2 – Correlations between different measures for the aphasic (upper right triangle) and control (lower left triangle) groups.

	Response time	Semantic similarity	Cue abstractness	Response abstractness
Response time	–	−.141 (aphasic)	.353** (aphasic)	.296* (aphasic)
Semantic similarity	.082 (control)	–	−.278* (aphasic)	.084 (aphasic)
Cue abstractness	.351** (control)	−.029 (control)	–	.505** (aphasic)
Response abstractness	.244* (control)	.114* (control)	.467** (control)	–

*$p < .01$, **$p < .001$.

due to abstract word meaning relying more on word-to-word relations that are damaged in the aphasics. Concrete meaning would more likely be spared if based on activation of sensory and motor areas. However, it cannot be excluded that the deteriorated abstract word processing could be associated with some more general cognitive decline, which might also be present in other patient groups. Examining subjects with lesions in different locations could lead to stronger conclusions about the relationship between lesion location and behavior. The fact that aphasics' associations to concrete words had a normal degree of semantic similarity to cue word in the LSA space does not necessarily mean that the aphasic participants understood the words in the same way as the model represents them, i.e., through word-to-word relations. The sensory experiences activated by concrete words are likely to be implicitly contained in the semantic space, since concrete discourse contexts should contain words denoting sensory experiences of the referred situations. The results show that semantic space models can be helpful in judging the semantic similarity of associations, since they provide the possibility of evaluating almost any association, and can be used for languages where no association norms exist. This could have practical implications in clinical settings, where diagnostic tools based on semantic spaces could be developed.

Acknowledgments

This work was supported by grant 2009-1773 from the Swedish Research Council. We thank Gerd Waldhauser for providing the abstractness norm.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cortex.2011.11.009.

REFERENCES

Audet C and Burgess C. Using a high-dimensional memory model to evaluate the properties of abstract and concrete words. In Hahn M and Stoness SC (Eds), *Proceedings of the Cognitive Science Society*. Mahaw, NJ: Lawrence Erlbaum, 1999.

Apt P. *Manual till Pia Apts afasiprövning (PAPAP)*. Stockholm: Psykologiförlaget, 1997.

Berry M, Do T, O'Brien G, Krishna V, and Sowmini V. *SVDPACKC (version 1.0) User's Guide. Technical Report CS-93-194, Department of Computer Science*. University of Tennessee, 1993.

Binder JR, Westbury CF, McKiernan KA, Possing ET, and Medler DA. Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17(6): 905–917, 2006.

Church KW and Hanks P. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1): 22–29, 1990.

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

De Groot AMB. Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15: 824–845, 1989.

Elman JL. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7: 195–225, 1991.

Gewirth LR, Shindler AG, and Hier DB. Altered patterns of word associations in dementia and aphasia. *Brain and Language*, 21(2): 307–317, 1984.

Just MA, Newman SD, Keller TA, McEleney A, and Carpenter PA. Imagery in sentence comprehension: An fMRI study. *NeuroImage*, 21: 112–124, 2004.

Kounios J and Holcomb PJ. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4): 804–823, 1994.

Kroll JF and Merves JS. Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12: 92–107, 1986.

Landauer K, McNamara DS, Dennis S, and Kintsch W (Eds), *Handbook of Latent Semantic Analysis*. Mahaw, NJ: Lawrence Erlbaum, 2007.

Lesser R. Word association and availability of response in an aphasic subject. *Journal of Psycholinguistic Research*, 2(4): 355–367, 1973.

Lund K and Burgess C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2): 203–208, 1996.

Paivio A. Dual coding and the mental lexicon. *The Mental Lexicon*, 5(2): 205–230, 2010.

Pulvermüller F and Fadiga L. Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11: 351–360, 2010.

Sabsevitz DS, Medler DA, Seidenberg M, and Binder JR. Modulation of the semantic system by word imageability. *NeuroImage*, 27(1): 188–200, 2005.

Schwanenflugel PJ, Akin C, and Luh W-M. Context availability and the recall of abstract and concrete words. *Memory & Cognition*, 20(1): 96–104, 1992.

Shtyrov Y. Automaticity and attentional control in spoken language processing: Neurophysiological evidence. *The Mental Lexicon*, 5(2): 255–276, 2010.

Wyke M. An experimental study of verbal association in dysphasic subjects. *Brain*, 85: 679–686, 1962.