Testing the cognitive relevance of a geometric model on a word association task: A comparison of humans, ACOM, and LSA

HYNGSUK JI
Sungkyunkwan University, Seoul, Korea

BENOÎT LEMAIRE
Laboratoire TIMC-IMAG, University of Grenoble, Grenoble, France

HYUNSEUNG CHOO
Sungkyunkwan University, Seoul, Korea

AND

SABINE PLOUX
Institut des Sciences Cognitives, L2C2, UMR5230 CNRS, Université Claude Bernard Lyon 1, France

The general aim of this study is to validate the cognitive relevance of the geometric model used in the semantic atlases (SA). With this goal in mind, we compare the results obtained by the automatic contexonym organizing model (ACOM)—an SA-derived model for word sense representation based on contextual links—with human subjects' responses on a word association task. We begin by positioning the geometric paradigm with respect to the hierarchical paradigm (WordNet) and the vector paradigm (latent semantic analysis [LSA] and the hyperspace analogue to language model). Then we compare ACOM's responses with Hirsh and Tree's (2001) word association norms based on the responses of two groups of subjects. The results showed that words associated by 50% or more of the Hirsh and Tree subjects were also proposed by ACOM (e.g., 71% of the words in the norms were also given by ACOM). Finally, we compare ACOM and LSA on the basis of the same association norms. The results indicate better performance for the geometric model.

Over the past 20 years, much research has been devoted to modeling lexical concepts and word semantics. Most models have attempted to satisfy two sorts of criteria. The first is to provide a suitable formal framework for describing semantic and conceptual facts; the second is to furnish a basis for predicting and simulating experimental outcomes. Once validated, a model can, in turn, supply supporting information for theoretical debates about how human knowledge and the mental lexicon are organized. Arguably, a formal model capable of accounting for a broad range of data is also a plausible basis for describing the way humans process concepts and words.

In what follows, we will begin by briefly summarizing prior research, which has focused mostly on two types of models: hierarchical models and vector models. We will highlight the principles behind the construction of each of these models, the specific kinds of data for which they successfully account, and the challenges faced by each. We then will introduce the geometric paradigm underlying our own model (hereafter referred to as ACOM, for automatic contexonym organizing model), which repre-

sents a third approach. This model seeks to satisfy the same constraints as do the other, well-established ones—namely, to provide a formal foundation for representing lexical concepts while also accounting for behavioral data. The rest of the article will be devoted to determining the extent to which ACOM can account for behavioral data in a word association task.

Hierarchical Models

Hierarchical models were initially inspired by a traditional way of organizing concepts that involves tree structures representing relationships between lexical concepts. According to this classical, Aristotelian approach, concepts are defined as a set of necessary and sufficient properties, and each concept inherits properties from the concepts above it in the tree. For example, *robin* inherits from *bird* the fact that birds have beaks, wings, and so forth. Higher up in the hierarchy, concepts such as *plant* and *animal* inherit properties from the concept *organism*. With respect to behavioral data, hierarchical models are insightful in accounting for selective processing deficits affecting certain

---

S. Ploux, sploux@isc.cnrs.fr

semantic categories, such as *animate*/*inanimate objects*, *parts of the body*, or *fruits* and *vegetables* (see Forde & Humphreys, 2002). The sort of categorical deficits experienced by a patient are specific enough to be characterized by simply cutting off a branch of the tree.

The best-known example of an elaborate hierarchical model is arguably the WordNet database (Fellbaum, 1998), which covers a very large array of lexical concepts and, for this reason, is a crucial reference across several domains (e.g., linguistics, psychology, and automatic language processing). In this lexical database, each lexical concept is represented by a set of synonyms (SynSet), and lexical concepts are linked to each other by relations such as hyponymy, which permits property inheritance. Compilation of the WordNet system was achieved mostly by lexicographers (and thus, "by hand"), on the basis of a predefined hierarchical organization.

Each model has challenges that must be tackled. Here, we will highlight three challenges facing hierarchical models. First, it is not obvious how an Aristotelian approach can organize all concepts. For example, the concept *game* resists an analysis based on necessary and sufficient conditions (as was pointed out by Wittgenstein, 1953). Similarly, nouns lend themselves to categorization more readily than do members of other grammatical categories, such as verbs, adjectives, and adverbs. Second, attempts to export the WordNet approach to other languages have encountered numerous obstacles stemming from difficulty in matching concepts across languages. This runs counter to the classic assumption that the conceptual level is language independent and can capture how each language "breaks down" reality into words (see, e.g., Pottier, 1992). Finally, it has been difficult to reconcile structural representations with basic experimental outcomes. For instance, the sentence *A robin is a bird* is processed more quickly than the sentence *A chicken is a bird*, which suggests that a purely logical approach is not entirely predictive of psychological representations. If mental computation consists simply of verifying a list of necessary and sufficient conditions, these two sentences should be processed similarly (see Collins & Loftus, 1975; Collins & Quillian, 1969; Rips, Shoben, & Smith, 1973; Smith, Shoben, & Rips, 1974).

Vector Models

Vector models emerged as an alternative to the classical approach. As Kintsch (2001) wrote, "[A vector model such as] Latent Semantic Analysis attempts to provide a computational underpinning of Wittgenstein's claim (that word meanings are not to be defined, but can only be characterized by their 'family resemblance')" (p. 175). Unlike WordNet, vector models are not constructed in view of organizing lexical concepts according to a given predefined structure but are designed to calculate lexical meanings on the basis of how words are used in context. In order to capture real usage, vector models attach values to co-occurrences of words across large corpora. The number of co-occurrences of a given word is stored in a matrix and then used to establish a multidimensional space in which each word is represented by a vector. In this paradigm, a word's semantics is not defined but is characterized by similarities among

words, as determined by the proximity of the vectors assigned to the words. Proximity between two words can be calculated as the cosine of their vectors (Landauer, McNamara, Dennis, & Kintsch, 2007; see also Table 2, which lists the closest neighbors of the word *party* on the basis of the cosine measure) or as the distance between the two word vectors (Burgess & Lund, 1997), and so forth.

The two models that initiated this approach are latent semantic analysis (LSA; for a review, see Landauer et al., 2007) and hyperspace analogue to language (HAL; Burgess & Lund, 1997). Several experiments have been carried out to determine the extent to which vector models can predict human behavior. For example, Lund and Burgess (1996) showed how the distance between word vectors as calculated by HAL provides significant correlations with respect to reaction time speeds in semantic priming studies. The vector model LSA and the hierarchical model WordNet have been compared in terms of their ability to account for the results of behavioral data in tasks sensitive to semantic similarity: semantic priming effects in lexical decision tasks, substitution errors in speeded picture naming, and semantic interference effects of distractor words (Vigliocco, Vinson, Lewis, & Garrett, 2004). Using these tasks involving the processing of semantic similarity between a prime word and a target word or between a distractor and the subjects' responses, Vigliocco et al. showed that correlations between model data and behavioral responses are generally higher for LSA than for WordNet.[1] On the basis of these experiments, it seems that the vector paradigm with computed semantic distances is better able to account for human processing of semantic similarities.

Although vector models are designed to capture the semantic similarity of words, LSA and HAL have also been tested to see whether they can account for categorical differences. For example, Laham (1997) used a list of 140 objects and 14 categories and showed that LSA was able to predict membership between objects and categories well above chance. In the same manner, Lund and Burgess (1996) made comparisons between a list of animals, a list of body parts, and a list of geographical places, using HAL. These experiments showed how vector models are able to account for the fact that distances between word vectors within a category can be lower than distances between word vectors belonging to two different categories. Nevertheless, these studies did not capture individual categories in the way that WordNet does. For example, the set of artifacts is not delineated and separated from all other categories, and vector models do not propose a hierarchical, tiered category structure like that found in hierarchical models (*a robin is a bird*, *a bird is a vertebrate*, etc.). Finally, these experiments were conducted with a small number of categories, unlike those with WordNet, which has the advantage of encompassing a very large range of categories.

To summarize, these two types of approaches differ in basic construction (a priori vs. lexical meanings in context), in the nature of the links between words (hierarchical vs. proximity), and in the method used ("manual" vs. automatic and statistical). However, the two approaches do share a feature—namely, both assign every concept or

word an indivisible, atomic, mathematical representation: a node of a graph in hierarchical models, a vector in vector models. Consequently, neither of these formal models offers a representation of the internal structure of concepts (if one exists) but only the extrinsic relations between them.

Geometric models are designed to represent the internal structure of concepts or word senses. In the next section, we will present the geometric paradigm and will describe how the geometric model ACOM was built. Our goal is to test the suitability of ACOM for modeling a word association task. To this end, we used the Hirsh and Tree (2001) norms, which we will briefly describe first. Then we will compare the responses of subjects with those output by the ACOM model. To our knowledge, this type of validation has never been done. Model validations generally compare word association norms with semantic distances calculated by the model (see, e.g., Griffiths & Steyvers, 2003). They do not test the semantic and cognitive quality of the word's neighbors supplied by the model—that is, the model's ability to produce relevant responses.

Finally, to assess the contribution of geometric modeling, relative to vector modeling, we will compare the results of ACOM and LSA on the basis of the Hirsh and Tree (2001) norms. The ACOM model and the vector models were built in the same way; that is, they were compiled by establishing an inventory of co-occurrences in large corpora of texts. The present comparison of a geometric model and a purely vectorial model will allow us to assess the differences between these two approaches: a structured approach in which the meaning of a word is organized, versus an extrinsic approach in which the meaning of a word is defined solely by its degree of proximity to all other words. This comparison between a geometric model and a vector model is also a first in the field of lexical models.

THE GEOMETRIC PARADIGM

We use the term *geometric model* to refer to any model that assigns a domain[2] in a multidimensional space (and no longer a vector) to each concept or word. With this definition, the internal structure of a concept is represented by interpreting the various zones constituting the domain, the concept's boundary corresponds to the outer edge of the domain, and semantic similarity between units is measured in terms of the overlap between the units' respective areas and the distances between them. Proposals of geometric models are scarce, although Gärdenfors (2000) has offered a philosophical argumentation in their favor and Widdows and Higgins (2004) have presented an example of a geometric computational model for organizing concepts. The ACOM model (Ji, Ploux, & Wehrli, 2003; Pcdico.isc.cnrs.fr) is a geometric model derived from the semantic atlases, initially developed using semantic similarity data from synonym dictionaries (Ploux, 1997; Ploux & Victorri, 1998; Pcdico.isc.cnrs.fr). ACOM is founded on two principles. The first, which it shares with the semantic atlases, pertains to the organization of word senses; the second rests on the assumption that the meaning of a word is induced by its

Table 1
Cliques Generated by the Model for the
Words *Suit*, *Party*, and *Organ*

Word	Examples of Cliques Output by ACOM
Suit	blue, brown, coat, hat, shoes, suit, wearing
	company, court, different, law, person, purpose, suit
Party	company, court, english, family, ladies, party
	democratic, leader, member, members, national, party,
	political, president, public, war
Organ	organ, organs, parts, species, structure
	cathedral, choir, church, music, organ

usage—particularly, its usage in written texts. In the latter respect, ACOM is like LSA and HAL.

The Organization of Senses in the Semantic Atlases Model

Our approach to word sense distinguishes two semantic levels: the area assigned to a lexical unit and minimal units, or *semantic points*. In the ACOM model, minimal units are instantiated by cliques generated on the basis of contexonymy. Ji (Ji et al., 2003) defined contexonymy as a contextual link between words that is both relevant and robust. Below, we will give a precise mathematical definition of the notion of *clique*. For our purposes here, a clique is a set of terms, each of which has a contexonymy link with every other term in the clique. These shared links constrain the semantic values of each word in the clique—particularly, the initial term. Table 1 gives some examples of word cliques.

We can see in Table 1 that each clique represents a fine-grained sense context. For the word *organ*, the *cathedral, choir, church, music, organ* clique is associated with the notion of *musical instrument*, whereas the *organ, organs, parts, species, structure* clique refers to the notion of *bodily organ*.

In the next section, we will briefly describe how a lexical context database is compiled and will describe the algorithm used to generate the semantic space assigned to a given term. For more detailed information, the reader may wish to refer to a previous article that describes this procedure more explicitly (Ji et al., 2003).

ACOM

Compiling the Lexical Context Database

First, the model creates a lexical database that, for every word it includes, gives the set of all words connected to it through context. This particular lexical database was obtained from two large corpora: an English corpus maintained by Project Gutenberg, which includes literature, essays, and other writings, and the British National Corpus. The lexical databases compiled separately for these two corpora were merged. The total number of tokens in the training corpora was over 300 million. All types of electronic dictionaries and encyclopedias were excluded from the training corpus so as to avoid having predefined lexical links that were expert or lexicographer dependent.

The method used is rather straightforward, in that it simply considers co-occurrences of words. Inside a window, which can be a sentence or a paragraph (a sentence in this study), co-occurrences of words are counted and then stored in a database. Each word in the corpus is selected as a headword (or entry word) and is denoted as W_n ($1 < n < N$), where N is the total number of types in the corpus. W_n has children (c_js), which are arranged in descending order of co-occurrence with W_n. Children with co-occurrences under a 10,000th of the overall frequency of the entry word are removed to reduce noise. The lexical database is thus a set of lines of the following type:

$$W_n: c_1, c_2, \ldots, c_{Kn}.$$

Contexonyms are not symmetrical or transitive; that is, when target word W_n has contexonyms $c_1, c_2, \ldots, c_{Kn}$, W_n is not necessarily a contexonym of c_i ($1 < i < K_n$). This is also true among c_is. We hypothesize that the better the training corpus is, the more relevant and robust the contexonyms obtained from it will be. A "good" corpus is one that is sufficiently large and well balanced.

Assigning a Semantic Area to a Term

Below is a brief description of the steps in the algorithm, which draws up the contexonym list for each corpus word in order to build a well-organized semantic space.

Step 1. For the target word, a word association table with three factors is generated. In order to eliminate children that rarely co-occur with W_n, the first portion (where $0 < \alpha < 1$) of the word is selected. W_n becomes: $W_n: c_1, c_2, \ldots, c_{kn}$, (where $k_n \leq K_n$). In view of selecting the most relevant children for each c_i, a second parameter, β, is used. Like α, the β portion of the child contexonyms of c_i are selected. The second selection narrows down the set of words preselected on the basis of parameter α in such a way that only the most relevant terms are retained. Let's take the example of the word *organ*. *Organ* occurs 4,863 times in the selected corpus, and the corresponding lexical database contains 5,731 terms associated to it, in decreasing order of frequency. Below are two lists showing the first 30 contexonyms retained, with α set at .1 (10%) and β set first at .4 (40%) and then at .1 (10%).

$\alpha = .1, \beta = .4$:
church, music, played, sound, playing, play, barrel, choir, brain, organs, mouth, piano, grinder, deep, voices, singing, notes, instrument, eye, soul, species, street, developed, natural, function, cathedral, pipe, solemn, pipes, parts

$\alpha = .1, \beta = .1$:
church, music, played, sound, playing, play, barrel, choir, brain, organs, mouth, piano, grinder, deep, voices, singing, notes, instrument, eye, species, street, developed, natural, function, cathedral, pipe, solemn, pipes, parts, tones

We can see here that a more restrictive value for parameter β eliminated certain less relevant contexonyms, such as *soul*, which disappeared on the second list.

Step 2. Step 2 consists of computing all the cliques for a given word. Cliques are computed from the associated word lists output in Step 1. Mathematically, a clique is a maximal complete subgraph of a larger graph. The graph used here is defined by the set of all the words in the corpus, which are nodes of the graph, and the set of arcs that link those words via a contexonymy relation. Cliques are calculated using a third parameter γ ($0 < \gamma < 1$) that governs their granularity. The higher the value of γ, the more terms there are in the clique and, thus, the more precise the represented context. For example, for the word *organ* and with the parameters α and β set at .1, the algorithm finds 33 cliques for $\gamma = .05$ and 44 cliques for $\gamma = .1$. The two cliques calculated with $\gamma = .05$ were the following:

developed, function, natural, organ, organs, parts
and

developed, natural, organ, organs, parts, species,

which reflect a shade of meaning of the word *organ* more specifically related to a function (first clique) and a shade of meaning similar to the first but related to differences between the species (second clique), respectively. These two cliques form a single clique when $\gamma = .1$:

developed, function, natural, organ, organs, parts, species.

This clique encompasses both of the previously differentiated shades of meaning.

Once the three parameters (α, β, and γ) are set, cliques are computed automatically.

Step 3. Recall that in the model, cliques represent units of meaning finer than the meaning of the word itself and that they are the points in the word's space. To compute this space, we apply a factor correspondence analysis developed by Benzécri (1980) to the matrix M, whose rows contain the cliques and whose columns contain the words. The value of matrix element $M_{i,j}$ is 1 if clique i contains word j, and 0 otherwise. The correspondence analysis applied to this matrix gives the clique's coordinates and the coordinates of the centers of gravity of the areas assigned to the initial word and its contexonyms. In accordance with the geometric paradigm, which assigns an area, the terms are represented by the envelope of the cliques that contain them. The distance χ^2 between two cliques, y_i and y_j, is expressed by the formula

$$\chi^2\left(y_i, y_j\right) = \sum_{1 \leq k \leq n} \left(x_{..} / x_{.k}\right)\left(x_{i,k} / x_{i.} - x_{j,k} - x_{j.}\right)^2, \quad (1)$$

where n is the total number of contexonyms and p is the total number of cliques, $x_{..} = \sum_{1 \leq i \leq p}\sum_{1 \leq j \leq n} x_{i,j}$, $x_{i.} = \sum_{1 \leq j \leq p} x_{i,j}$, $x_{.,k} = \sum_{1 \leq j \leq n} x_{j,k}$, and x_{ij} is equal to 1 if the jth contexonym belongs to the ith clique and is equal to 0 otherwise. Since every clique has its own coordinates, clique distances are proportional to clique relatedness.

Step 4. Cliques are projected onto the principal plane and are classified by hierarchical clustering (see Figure 1 for an example of the area assigned to the word *organ*). ACOM's output can be viewed at the following address: Pcdico.isc.cnrs.fr.
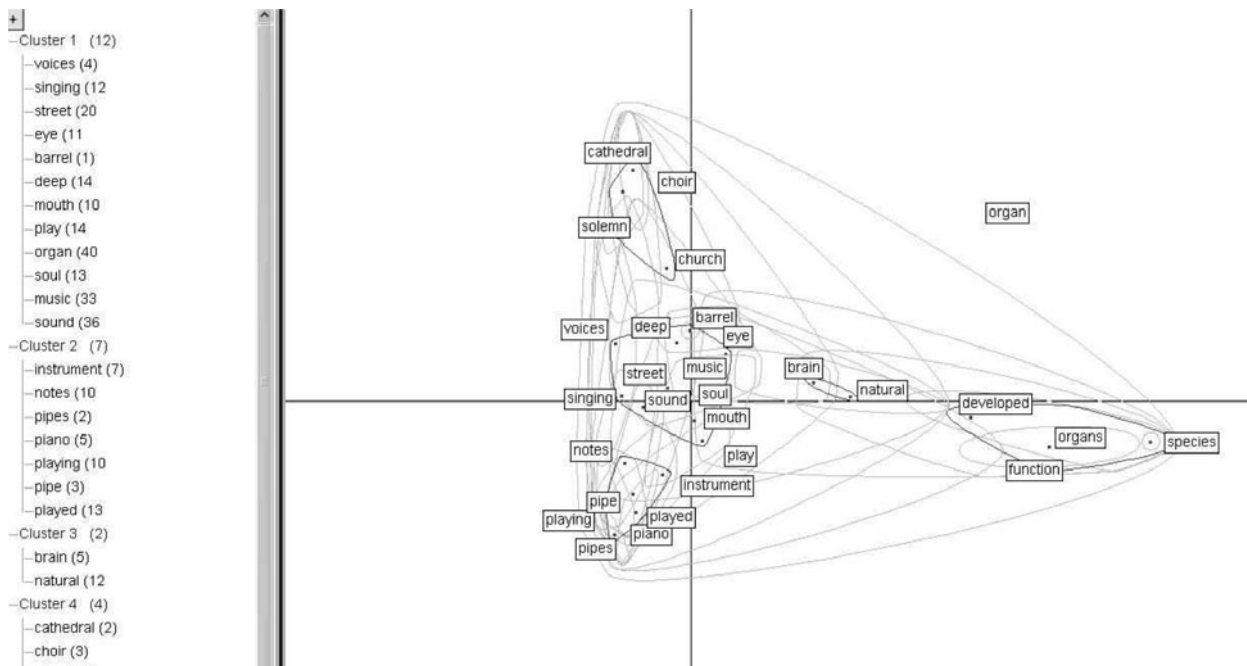
Figure 1. Space assigned to the word *organ*, with five computed clusters. $\alpha = .2$, $\beta = .2$, $\gamma = .05$.

In order to examine the validity of the model—that is, the relevance of the contexonyms that ACOM produces—we chose Hirsh and Tree's (2001) experimental data for comparison with ACOM's output. The next section will briefly sketch Hirsh and Tree's experiment and results.

WORD ASSOCIATIONS

An association experiment consists of asking subjects to come up with words associated with target words. For example, if the word presented is *snow*, the subject might say *white* or *mountain*. In Hirsh and Tree's (2001) study, the authors defined association norms on the basis of the responses of two cohorts of British subjects. One of the

cohorts contained 21- to 30-year-old subjects, and the other, 66- to 81-year-old subjects. For each of a total of 90 target words, the subjects had to come up with a word associated to the target. Their answers were ordered by frequency and were converted into occurrence percentages for each group. Hirsh and Tree compared the results of the two groups and noted that their responses overlapped only moderately for the most frequently given terms: Only 36 of the 90 most common responses were shared by the two cohorts.

In the next part of this article, which is based on a few sample words found in the Hirsh and Tree (2001) norms, we will point out the specificities of the geometric representation with respect to a vector representation and then will compare the norms with the closest words proposed by ACOM. After these introductory examples, we will report our full analyses, carried out on the entire set of 90 words.

Geometric Versus Vectorial Representation

The 90 words selected referred to concrete notions (e.g., *dragon*, *circus*, *blue*). Some were nouns only (e.g., *puma*, *acrobat*); others could also be adjectives (e.g., *blue*) or verbs (e.g., *cross*). Although the authors did not mention this, some terms were monosemous (e.g., *puma*, *cigarette*), whereas others were polysemous or homonyms (e.g., *party*, *suit*, *organ*). In LSA, associates are listed in order of distance from the target term, without categorization. Table 2 shows LSA's output for the corpus *General Reading Up to 12th Grade*, taken from the Web site lsa .colorado.edu (neighbors whose frequency is below 20 were not included). For the word *party*, for example, the terms *invite* and *birthday* are mixed in with terms per-

Table 2
Words Computed by Latent Semantic Analysis

Word	Closest Neighbors, in Decreasing Order of Proximity
Party	1.00 party, .85 parties, .78 democrats, .75 democratic, .73 candidates, .71 republican, .68 candidate, .67 political, .66 republicans, .66 nomination, .64 politician, .62 supporters, .62 presidential, .61 invite, .61 elections, .61 birthday, .61 invitations, .60 election, .60 precinct, .59 bosses, .58 campaign, .57 nominated, .56 federalist, .56 membership
Suit	1.00 suit, .64 wearing, .64 wore, .62 hat, .62 suits, .59 sleeves, .59 pants, .59 shirt, .58 dressed, .57 dress, .57 trousers, .57 collar, .56 blouse, .55 wear, .54 shoes, .54 hats, .54 cuffs, .53 coat, .53 pockets, .52 velvet, .52 jacket, .52 outfit, .52 clothes, .51 shirts, .51 jeans
Organ	1.00 organ, .87 organs, .74 tissues, .69 tissue, .64 muscular, .61 circulatory, .59 cells, .59 carries, .57 sac, .56 intestines, .56 lining, .55 nerve, .54 enters, .54 specialized, .53 bladder, .53 lymph, .53 vessels, .53 blood, .53 corpuscles, .53 contractions, .51 function, .51 kidneys, .50 liver, .50 digestive, .49 nervous

Table 3
Words Computed by the Automatic Contexonym Organizing Model

Word	Separation Into Clusters
Party	{war, public, led, chief, court, strong, neither, third} {number, met, city, twenty, river, camps, arrived, joined, miles, horses, captain} {states, united, members, leader, political, parties, democratic, republican} {family, ladies, dinner}
Suit	{hat, dress, wore, blue, clothes, silk, dressed, shoes, wearing, coat, shirt, wear, worn, gray, tie} {hair, read, dark, fine, brown, cut, figure, gray} {person, court, different} {needs, follow, business, law}
Organ	{sound, music, play, street, deep, soul, singing, eye, voices, pipes, barrel} {notes, instrument, played, playing, piano, pipe, pipes} {church, solemn, choir, cathedral} {natural, brain} {developed, organs, function, species}

taining to political parties (e.g., the words *democrats*, *republicans*, etc. come before them, and the words *election*, *campaign*, etc. come after).

In ACOM, words are grouped by usage context. Figure 1 shows a geometric representation of the word *organ*. On this map, the outer envelope delineates the area associated to this word. The points represent the cliques computed by the method described in the ACOM section. The clusters classify the different zones on the map. Here, five clusters are computed; they represent semantic values of the word (organ as a musical instrument, organ in a church, organ as a structure that fulfills a function, etc.).

For the three words *party*, *organ*, and *suit*, Table 3 gives the principal clusters distinguished by ACOM and the terms the clusters contain. So for the word *party*, there is a cluster containing the terms *family*, *dinner*, and so forth, which refer to the concept *reception*, and a cluster containing the terms *political*, *democratic*, and so forth, which refer to *political party*. The same is true for *suit*, where the different semantic values (represented by associated words) are separated into a value related to clothing (*clothes*, . . .) a value related to persons (*hair*, . . .), and a value related to the law and business (*law*, . . .).

Norms Versus ACOM

The 90 words used to determine the association norms were input into the ACOM system. Below, we will report some of the comparative results for these words. In Table 4, the first column lists the responses (in decreasing order of frequency) given by the 21- to 30-year-old cohort (45 subjects); the second column lists the responses of the 66- to 81-year-old cohort (45 subjects); and the third column gives ACOM's responses (in decreasing order of the number of cliques to which the term belongs). The greater the number of cliques a term has, the closer its assigned center of gravity is to the origin, which itself is the center of gravity of the initial word. This parameter thus measures the link between the response and the initial word. In the table, the words in boldface are the words output by ACOM that were also suggested by the subjects in Hirsh and Tree's (2001) experiment. For each word, the number in brackets indicates the percentage of subjects who chose that word. For instance "sky [.20]" indicates that 20% of the younger subjects selected the word *sky* as a word associated to *blue*. "Mood, red [.04]" indicates that the words *mood* and *red* were selected as associated to *blue* by 4% of the older subjects.

Table 4
Comparison of the Responses Given by Two Cohorts of Subjects and the
Automatic Contexonym Organizing Model (ACOM) for the Words *Blue* and *Dragon*

Word	21- to 30-Year-Old Subjects	66- to 81-Year-Old Subjects	ACOM
Blue	Sky [0.20], red [0.11]	Sky [0.40]	**White**, **green**, sun, **red**, dark, **hair**, brown, **color**, bright, **sky**, trees, rose, soft, silver, grey, gray, **sea**, gold, deep, pale, yellow, wore, pink, flowers, big, across, coat, clear, dress
	Green [0.07]	Color [0.29]	
	Cold, cross, moon, sad, sea, white [0.04]	Sea [0.09]	
	Bird, bottle, boy, depressed, grass, hair, ice, kitchen, mood, Nile, police, sailor, shirt, train, true, water [0.02]	Mood, red [0.04] Depressed, green, lagoon, lamp, miserable, upset [0.02]	
Dragon	Fire [0.29]	Fire [0.24]	**Red**, heads, **green**, mouth, gold, sword, lion, golden, forth, ship, blue, terrible, **fly**, teeth, wings, fiery, fight, **knight**, serpent, **monster**, tail, princess, killed, dragons, **George**, throne, flies, **cave**, Chinese
	Wales [0.11]	Wales [0.18]	
	Slayer, fly [0.07]	Fly, Welsh [0.13]	
	Cave, knight, Welsh [0.04]	George [0.07]	
	Breath, damsel, dinosaur, emblem, exotic, flames, George, green, lance, monster, mother, myth, red, slay, witch [0.02]	Fierce, red [0.04] Animal, blocked, China, fictitious, legend, roar, wife [0.02]	

Note—Words in boldface are the words output by ACOM that were also suggested by the subjects in Hirsh and Tree's (2001) experiment.

Table 5
Agreement Between the Automatic Contexonym Organizing Model (ACOM)
and Subjects: Number of Associated Words Produced

Cutoff Percentage (Lower Bound)	Younger Cohort		Older Cohort		Overall Agreement	
	Given by Subjects	Given by ACOM	Given by Subjects	Given by ACOM	Given by Subjects	Given by ACOM
0% (all words)	1,597	367 (23%)	1,320	329 (25%)	2,324	476 (20%)
10%	189	98 (52%)	217	111 (51%)	305	149 (49%)
20%	87	54 (62%)	110	69 (63%)	149	91 (61%)
30%	39	29 (74%)	52	34 (65%)	70	46 (66%)
40%	23	18 (78%)	36	25 (69%)	44	31 (70%)
50%	13	11 (84%)	19	12 (63%)	24	17 (71%)

Note—Matching rates are shown in parentheses in the ACOM columns.

COMPARATIVE RESULTS

As a final step, we computed the matching rate between the contexonyms produced by the model and the associated words produced by the Hirsh and Tree (2001) subjects. For each word, the parameters α, β, and γ were set at .1 and then incremented by .1 until 30 contexonyms were found.

Table 5 lists the ACOM results alongside those obtained for the subjects in the Hirsh and Tree (2001) experiment. The first column gives the cutoff percentage for the number of subjects who suggested the words. Each of the next three columns has two subcolumns: one showing the number of words produced by the subjects, and the other the number output by ACOM. Take, for example, the 20% row for the younger cohort: The 87 in the "Subjects" column means that for the 90 terms proposed in the experiment, 87 different words were suggested by at least 20% of the younger subjects; the 54 in the "ACOM" column means that 54 of those 87 words (62% matching rate) were also given by ACOM. Note that the matching rate rises with the number of subjects who chose the word. For example, for all the associated words given by more than 10% of the subjects, the matching rate between the subjects' responses and the model's contexonyms is 49%, whereas for words given by 40% of the subjects, the matching rate is 70%.

It seems from Table 5 that for the set of all associated words, ACOM's performance was not as good for the younger subject cohort (the overall agreement between this cohort and ACOM is 23%) as for the older subject cohort (25%). However, the difference in performances was reversed as the frequency of the associated words increased. For example, the agreement between the younger cohort and ACOM for words given by at least 30% of the sub-

jects is 74%, as compared with 65% for the older subject cohort. In order to verify these differences across cohorts, we compared the number of associations given by both ACOM and the younger subject cohort [AYS_{word_i} ($1 \leq i \leq$ 90)] and the number of associations given by both ACOM and the older subject cohort [AOS_{word_i} ($1 \leq i \leq$ 90)], for the entire set of 90 words. In this comparison, the variable number of responses given for each of the 90 words was taken into account by dividing AYS_{word_i} and AOS_{word_i} by the total number of associations given by each cohort for that word. A sign test applied to these values showed that the performance difference across cohorts was significant ($Z = 2.543995$, $p = .010959$) for the whole set of associated words and that, for this set, ACOM's results were better for the older subjects' responses. On the other hand, the same test applied to the words most often proposed by the subjects (by at least 30% or at least 40%) did not yield a significant cross-cohort difference. Note, however, that the inferiority of ACOM's responses for the younger subject cohort and its superiority for the most frequent words showed up again when the initial corpus was modified (see Table 6). Moreover, because the initial set of target words that generated associated words whose frequency was above 30% was relatively small, it would be worthwhile to conduct a new experiment on a larger number of initial target words in order to test the hypothesis that, for highly frequent associated words, ACOM's behavior is more like that of young adults than it is like that of old adults.

Norm-Based Comparison Between LSA and ACOM

It seemed interesting to compare ACOM's output (presented in Table 5) with that of a recognized model such

Table 6
Agreement Between the Automatic Contexonym Organizing Model (ACOM),
Latent Semantic Analysis (LSA), and Subjects: Number of Associated Words Produced

Cutoff Percentage (Lower Bound)	Younger Cohort			Older Cohort			Overall Agreement		
	Given by Subjects	Given by LSA	Given by ACOM	Given by Subjects	Given by LSA	Given by ACOM	Given by Subjects	Given by LSA	Given by ACOM
0%	1,597	151 (9.5%)	255 (16%)	1,320	155 (12%)	209 (15.8%)	2,324	217 (9.5%)	316 (13.6%)
10%	189	46 (24%)	66 (35%)	217	52 (24%)	72 (33%)	305	71 (23%)	95 (31%)
20%	87	24 (28%)	37 (42.5%)	110	32 (29%)	49 (44.5%)	149	42 (28%)	61 (31%)
30%	39	12 (31%)	20 (51%)	52	19 (36.5%)	26 (50%)	70	23 (33%)	33 (47%)
40%	23	7 (30.5%)	14 (61%)	36	14 (39%)	18 (50%)	44	17 (39%)	22 (50%)
50%	13	7 (54%)	10 (77%)	19	9 (47%)	11 (58%)	24	12 (50%)	14 (58%)

Note—Matching rates are shown in parentheses.

as LSA. To do so, we worked together with LSA experts. However, it was not possible to base the comparison on the corpus used initially to test ACOM, because the steps taken by the two models are quite different. ACOM creates a lexical database first, whereas LSA operates directly on a co-occurrence matrix that cross-tabulates words and paragraphs or texts. Creating a lexical database allows ACOM to process very large corpora, but the corpus used to obtain the results above generated a matrix that could not be processed by our LSA implementations. So, in order to compare the two models on the same corpus, we proceeded by randomly selecting a subset of the initial corpus, while making sure that the distribution of the text types remained balanced. The new corpus represented 12.3% of the initial corpus (31.6 million words). The results obtained are, therefore, different from the ones given in Table 5. To make the comparison, we again chose to retain the first 30 associated terms calculated by each model. For LSA, the 30 closest terms were computed. For ACOM, in order to get the first 30 contexonyms, the program increments the parameters α, β, and γ, set at an initial value of .1, until 30 associated words are obtained.

The results are given in Table 6. They indicate better performance by ACOM than by LSA for each cohort individually and also for both cohorts pooled. These performance differences were checked by applying a sign test to the set of 90 words. ACOM's results were better than LSA's for the younger subject cohort ($Z = 4.387862$, $p = .000011$), for the older subject cohort ($Z = 2.687728$, $p = .007194$), and for the two cohorts combined ($Z = 3.836172$, $p = .000125$). The cross-cohort performance difference between ACOM and LSA (i.e., the difference, for each word and each cohort, between the number of correct answers given by ACOM and the number of correct answers given by LSA, divided by the total number of associated words produced for that word by the subjects in that cohort) was also significant ($Z = 2.250176$, $p = .024438$). This test showed that ACOM agreed a little better with the younger subject cohort than LSA did. This slight superiority should, nevertheless, be validated on a larger set of words.

DISCUSSION AND CONCLUSION

One of the main features of the ACOM model consists of its ability to produce associated words (contexonyms) for a given word. In order to evaluate the reliability of ACOM's associated word output, a comparison study was carried out. The results showed that the ACOM model produces relevant associated words comparable to the ones given by human subjects.

It is noteworthy that ACOM produced such a reliable set of contexonyms. Knowing a set of associated words could be helpful in various research fields dealing with cognitive processes, such as cognitive neuroscience and cognitive psychology. However, few such sets are available, since producing them in a nonautomated way is very time consuming. Large sets of associated words obtained from ACOM could contribute significantly to research in these fields. Direct use of ACOM-generated word sets as

normative associated word sets could be applied to studies such as one in which the degree of neurological damage is determined on the basis of and investigation of a patient-produced set of associated words.

Associated words can be obtained using other computational models, such as LSA or HAL. Although the ACOM model has a number of features in common with vector models, it differs from them in both its design and its architecture. It is thus reasonable to contend that this is what caused the observed differences in performance. Below, we will recall the similarities and dissimilarities between ACOM and LSA. Then we will go into greater detail about the resulting performance differences and relate them to the features specific to each model.

Regarding the similarities between LSA and ACOM, they use analogous methods (a factor analysis) for representing semantic distances between words, and both are constructed from co-occurrences in text corpora. Their principal difference lies in their approaches: LSA takes a global approach, whereas ACOM relies on a local approach organized around each word. LSA, as a high-dimensional space model, considers the relations between a word and all other words in the lexicon, whereas ACOM starts with candidate words only, instead of looking at all chosen entries. The selection of candidate words does not involve any extensive computations but only a cutoff process. A consequence of this difference is that the overly costly processing load prevented LSA from producing associated words for the whole corpus, so we had to take a subset of the corpus for comparing ACOM and LSA. This difference has cognitive implications. One can reasonably contend that the way in which ACOM processes words is more intuitive than LSA's, since humans probably do not get associated words after scrutinizing all of the words in the lexicon.

As was stated above, if the differing abilities of ACOM and LSA to simulate a word association task are rooted in the difference between the intrinsic local representation of word meaning by the former and an extrinsic global representation of words by the latter, one can wonder whether words with a complex internal semantic organization—for example, polysemous words—are processed better by the geometric model than by the vector model. To test this hypothesis, we separated the 90 words in the norms into two groups. The first group contained monosemous words (like *hamster*), and the second group contained polysemous words (like *cross*). To define these two groups in an objective way, we noted the number of different senses of the words given by an outside source, *The New Shorter Oxford English Dictionary* (*SOED*; 1999). A word was assigned to Group 1 if all the associated terms given by the subjects pertained to only one sense in the *SOED*; a word was assigned to Group 2 if the associated words given by the subjects corresponded to at least two different senses in the dictionary. For example, the word *cross* had *angry* and *road* as associated words, which refer to two different meanings given by the *SOED*. With this method, Group 1 (monosemous words) ended up with 40 of the 90 target words, and Group 2 (polysemous words) with the other 50. This approach did not change ACOM's comparative performance on either group: ACOM still obtained bet-

ter results than did LSA (Group 1, $Z = 2.599735$, $p = .009330$; Group 2, $Z = 2.683282$, $p = .007290$), and the comparison between the two groups on the ACOM/LSA performance difference was nonsignificant (ANOVA, $F = 0.85639$, $p = .357281$).

This result, which needs to be replicated on a larger set of target words, nevertheless seems to suggest that the way the internal structure of words is modeled has an impact on the representation of meaning, not only for polysemous words, but also for monosemous ones. This advantageous feature may be due to the fact that the neighbors of monosemous words can themselves be polysemous (e.g., the word *acrobat*, which belongs to the first group, has the polysemous word *exercise* as an associated word). The structure of these associated polysemous words, along with the links interrelating the various associated words to each other, could have an indirect effect on the quality of the results obtained for the initial words, even ones that are monosemous themselves.

In conclusion, this experiment allowed us both to validate ACOM from the cognitive standpoint and to evaluate the contribution of a geometric organization of word meaning—as compared with a vector-based representation—on an association task. In the future, we plan to extend this study in several directions. The first will be to include other types of tasks involving a measure of semantic proximity, as in the Vigliocco et al. (2004) study, in order to determine whether the geometric organization is useful for these different tasks. The second direction will consist of evaluating the model on categorization tasks, where current hierarchical models seem so far to do the best job.

AUTHOR NOTE

REFERENCES

Benzécri, J.-P. (1980). *L'analyse des données: L'analyse des correspondances*. Paris: Bordas.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language & Cognitive Processes*, **12**, 177-210.

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic memory. *Psychological Review*, **82**, 407-428.

Collins, A., & Quillian, M. (1969). Does category size affect categorization time? *Journal of Verbal Learning & Verbal Behavior*, **8**, 240-248.

Fellbaum, C. (Ed.) (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Forde, E. M. E., & Humphreys, G. W. (Eds.) (2002). *Category specificity in brain and mind*. Hove, U.K.: Psychology Press.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.

Griffiths, T. L., & Steyvers, M. (2003). Prediction and semantic association. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 11-18). Cambridge, MA: MIT Press.

Hirsh, C., & Tree, J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, **14**, 1-44.

Ji, H., Ploux, S., & Wehrli, E. (2003). Lexical knowledge representation with contexonyms. In *Proceedings of the 9th Machine Translation Summit* (pp. 194-201).

Kintsch, W. (2001). Predication. *Cognitive Science*, **25**, 173-202.

Laham, D. (1997). Latent semantic analysis approaches to categorization. In M.-G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

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

The New Shorter Oxford English Dictionary (1999). Oxford: Oxford University Press.

Ploux, S. (1997). Modélisation et traitement informatique de la synonymie. *Linguisticae Investigationes*, **21**, 1-28.

Ploux, S., & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires informatisés des synonymes. *Traitement Automatique des Langues*, **39**, 161-182.

Pottier, B. (1992). *Sémantique générale*. Paris: PUF.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, **81**, 1-20.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214-241.

Vigliocco, G., Vinson, D., Lewis, W., & Garrett, M. (2004). Representing the meanings of object and action words: The featural and unitary semantic system hypothesis. *Cognitive Psychology*, **48**, 422-488.

Widdows, D., & Higgins, M. (2004). Geometric ordering of concepts, logical disjunction, and learning by induction. In S. D. Levy & R. Gayler (Eds.), *Compositional connectionism in cognitive science* (AAAI Fall Symposium Series). Menlo Park, CA: AAAI Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

NOTES

1. In Vigliocco et al.'s (2004) article, the authors compared LSA and WordNet, but also a model called FUSS, which they developed themselves. It is beyond the scope of the present article to describe the particularities of FUSS, as compared with LSA and WordNet. However, note that FUSS was set up on the basis of subjects' descriptions of concepts, using semantic features. With such a compilation mode, it has not been possible so far to obtain broad lexical coverage, unlike with LSA and WordNet. Note also that FUSS uses a self-organizing maps type of representation that also assigns a vector to each word. The superiority of FUSS described in that study for comparisons with behavioral data obtained from tasks sensitive to semantic similarity is another element in favor of vector modeling for tasks such as these.

2. A domain is primarily a topological space, but it is assumed, in addition, that each of its points has a neighborhood homeomorphic to an open set of R^n. For example, the areas delineated by a circle, a square, a rectangle, or an ellipse are domains.