Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Applied Logic 2 (2004) 457–467
JOURNAL OF APPLIED LOGIC
www.elsevier.com/locate/jal

Semantic computations of truth based on associations already learned

Patrick Suppes *, Jean-Yves Béziau [1]

Stanford University, Center for the Study of Language, Ventura Hall Room 36, Stanford, CA 94305-4115, USA

Available online 12 November 2004

Abstract

This article sets forth a detailed theoretical proposal of how the truth of ordinary empirical statements, often atomic in form, is computed. The method of computation draws on psychological concepts such as those of associative networks and spreading activation, rather that the concepts of philosophical or logical theories of truth. Axioms for a restricted class of cases are given, as well as some detailed examples.

Keywords: Truth; Computaton; Empirical statements; Associative networks; Spreading activation

1. Introduction: the problem of truth computation

In this article we try to give an account of how one determines the truth or falsity of sentences like: *Paris is the capital of France, Paris is not the capital of France, Rome is the capital of France.*

We want to describe the computations underlying the answers given, taking into account, at least in a qualitative way, the time factor—what psychologists call the latency of a response. Our theory should be able to explain the data gathered by experimentation, for example, why it takes more time to give a negative answer than a positive one, be it true or

* Corresponding author.
E-mail address: psuppes@stanford.edu (P. Suppes).
[1] Work supported by a grant of the Swiss National Science Foundation and Member of the LOCIA project (CNPq/Brazil).

1570-8683/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jal.2004.07.006

false. But the important theoretical question is what is the actual method of computation, a problem not ordinarily considered in philosophical theories of truth, but also not subject to direct empirical observation.

2. Background of the theory

2.1. Philosophy and logic

Philosophers discuss at length various theories of truth—coherence theory, correspondence theory, problem of direct reference, sense and denotation, and so on—but, curiously, do not give an account of how we actually perform truth computations, and even less why we are able to perform them so quickly. Philosophers who claim that "Paris is the capital of France" is true because Paris is the capital of France are generally not interested in explaining how we actually compute the answer. But, since such sentences are almost never remembered, or even previously encountered, a computation is necessary.

Logicians also do not solve these problems. If we want to describe how one answers a question like "Is 49 + 13 equal to 61?", it is certainly wrong to look at the logical foundation of arithmetic, whether it is proof-theoretical or model-theoretical. We answer a question like "Is 49 + 13 equal to 61?" by using a series of small computational algorithms and tricks, not by looking for a formal proof from a set of axioms or by finding a model in which the axioms are true and $49 + 13 = 61$ is false. In the case of a question like "Is Rome the capital of France?", it is even more doubtful that we are trying to deduce the truth or falsity of the sentence from a set of axioms, or by using a truth-table.

From our point of view it is misleading to say that we are making a *deduction* to arrive at the conclusion that "Rome is the capital of France" is false, unless we strongly emphasize that deduction does not reduce to the narrow meaning of deduction in formal logic. To avoid misunderstanding, it is better to say that we are here trying to describe how we *compute* the truth and falsity of such a sentence.

Logicians do not deal with this kind of problem. As stressed by Woods, they "have generally stopped short of trying to actually specify the truth conditions of the basic atomic propositions in their systems, dealing mainly with the specification of the meanings of complex expressions in terms of the meanings of elementary ones" [19, p. 220]. According to Woods, researchers in artificial intelligence are trying to find an alternative solution where logicians failed. But do they have a solution?

2.2. Artificial intelligence and computational linguistics

The *artificial intelligence paradox* is described as follows by Hölldobler, commenting on a paper by Shastri and Ajjanagadde, in which they propose a possible solution to this paradox, which is "the gap between the ability of humans to draw a variety of inferences effortlessly, spontaneously, and with remarkable efficiency, on the one hand, and the results about the complexity of reasoning reported by researchers in artificial intelligence, on the other hand" [13, p. 463]. This paradox shows very well that most research in the field of artificial intelligence does not solve our problem. In order to compare our approach with

the various approaches of AI researchers, or those working in computational linguistics, it is important to emphasize that, generally, it is not clear what they are trying to describe. Is it reasoning, processing of language, deduction, or something else?

In computational linguistics people are more interested in syntax: formal grammars, parsing of sentences, and so on. In AI, people are more oriented towards semantics, for example, the development of semantic networks. However, the status of such networks is not clear from the viewpoint of the distinction between syntax and semantics. Woods remarks that "The question of what (semantic) networks have to do with semantics is one which takes some answering" [19, p. 218]. Distinctions between inference, truth and meaning are not clear in semantics networks, which are a mix of many things. Anyway, it seems, from our viewpoint, that the orientation taken by AI researchers is better than the one taken by computational linguists, because with semantic networks they are trying to find a shorter path without going into the syntax and logical representation of natural language.

In a recent book on computational semantics, the authors, Blackburn and Bos, say:

The book is devoted to introducing techniques for taekling the following two questions:

1. How can we automate the process of associating semantic representations with expressions of natural language?
2. How can we use logical representations of natural language expressions to automate the process of drawing inferences? [8, p. iii].

Their idea is to find some algorithms to translate natural language into the language of first-order logic to represent the meaning of natural-language sentences and then to find some additional algorithms to make inferences with these first-order translations. The two steps seem wrong for our purpose. It is doubtful that our brains use first-order logic to compute empirical truths. Both AI researchers and computational linguists have been over-influenced by formal logic. They do not deal directly with the problem of finding the obvious truth or falsity of atomic statements like "Rome is the capital of France".

2.3. *Associations*

We share with AI researchers an emphasis on *associations* (sometimes in AI, "semantics networks" are also called "associative networks"). When answering a question such as "Is Paris the capital of France?", we are using notions which are associated with the input, like Eiffel Tower with Paris, or country with capital. Our purpose here is to try to explain the mechanism of such associations, in connection with the question of truth and falsity (in this point we differ from AI researchers who are concerned with broader problems). But our description should not depend crucially on language, even though we are working with linguistic examples, since we think that this mechanism has a common root in processes involving any language, nonverbal animal behaviour, and stimulus-response phenomena in general.

The viewpoint here is that an associative network is a set of nodes with links between them. One central question is how an associative network is organized. An interesting proposal about the global organization of the lexicon in English, based in psycholinguistic considerations, has been made by the Wordnet project. The organization is based on

three pairs of relationship: antinomy-synonymy, hyperonomy-hyponymy, and meronymy-holonomy (see [10,14]). However, this classification presents several serious limitations, in particular, it does not address the analysis of truth computations.

We will not present any general theory of organization of associative networks, but will focus our attention on how truth computations fit into associative networks. For us 'true' is here no mysterious entity, but a word in the associative network, like 'Paris' or 'capital'. Our main task is to explain, given an input like 'Paris is the capital of France', what happens in the associative network. Our idea is that 'true' or 'false' become linked with 'Paris is the capital of France', on the basis of some already existing associative links. We suppose here that these links are fixed, that they correspond to associations already learned, but, of course, for more general problems the links have to be considered dynamically. For a detailed proof, for example, that grammars can be learned just from associations or conditioning connections, see Suppes [15].

Our approach is similar to earlier work on semantic networks [12]. The most important difference is the detailed consideration of the dynamics of the computation of truth, starting with the dependence on an explicit external cause of activation, i.e., an auditory or visual verbal stimulus being presented to a person. And this activation is followed by spreading activation internally to other nodes not directly activated by the stimulus, as explained in more detail later.

3. The theory

3.1. Intuitive concepts and conventions

A large number of intuitively simple concepts are introduced in the statement of the axioms. We have not introduced a formal mathematical notation for these concepts, because we feel the meaning of the axioms will be much easier to understand if ordinary language is used.

We consider an associative network of nodes and links between the nodes. Of course, not all nodes are linked. A severe restriction of what we analyze in detail is that no learning or forgetting is considered, only performance after associations have been learned. We envisage the networks functioning in the following natural environment. Someone is asked a question, or asked to say whether a sentence about familiar phenomena is true or false—in fact, it is the latter alternative we consider explicitly, although the extension to questions is pretty obvious. So input or stimulation from outside the network, and the brain in which we implicitly assume it is located, comes in the form of sentences expressed orally or presented visually. As in ordinary conversation, everything relevant about the language used, English in our case, is assumed known. There is here no attempt whatsoever to begin from the beginning with the first learning of a first language. We are trying to formalize, at least partially, the processing of simple sentences that are empirically true or false. What we are doing in a general way is making explicit a detailed psychological model of how this processing is done. The model is much simplified to provide an overview of how computations of truth and falsity can be made in such a model that has its origins in ideas that go back to Hume's theory of association, and that have been much studied in psychology

in the last 100 years. What is important is that the fundamental ideas come from empirical scientific efforts rather than logical or philosophical ones.

So let us now turn to the concepts and conventions we use. The first to mention is that of a *brain image*. Reference is continually made to the brain images of words without further explanation. For some detailed attempt to be much more specific about this, see Suppes [16–18]. In this earlier work, brain images of words are taken to be finite temporal segments of superpositions of sine waves of varying frequency, amplitude, and phase, with the frequencies being very much smaller than that of ordinary speech, somewhere in the range from 1 to 30 Hertz. But such details, or even the correctness of this work, are not required here. We need only assume that our brains do have a way of representing words. Our language of images and associations makes no commitment to any particular method of representation.

It is not a new idea that associations should be thought of ultimately as in the brain, not simply in the mind. Here is what William James had to say about this point.

> ...And so far as association stands for a *cause*, it is between *processes in the brain*—it is these which, by being associated in certain ways, determine what successive objects shall be thought [11, p. 554].

Moreover, it is a well-defined problem of current research to conceptualize and test models derived from the physics of electromagnetic fields, and possibly the dynamics of other physical processes, to give an adequate physical grounding to the brain process of association. One other point. Some talk about association seems to claim that the associations we have are between things out in the world. But this view in its pure form cannot give a scientific account of our thinking processes. We go from physical things and processes to their representations (or images) in the brain. The physical associations must be physical phenomena in our brains, at least for those of us who do not hold to some outlandish dogma of dualism.

As already remarked, the brain images of words are permitted only two states here: *quiescent* or *active*. And we emphasize quiescent is not meant to suggest zero energy, but something very small but positive. (Note to logicians: we are serious about the energy remarks, for the ultimate theory of the phenomena we are developing is physical; indeed we would argue that at the most fundamental level the processing of language, on the occurrence of a linguistic stimulus, is primarily in the electromagnetic field generated by the relevant population of neurons, but this rather controversial thesis will also not be defended here. It is only introduced to give some orientation of where the ideas used are coming from.) We also emphasize again how crude the assumption of only two states is; much evidence supports the contrary.

The next idea is that we *activate* quiescent states to produce activated states. The energy for this activation comes, at least partially, from the verbal stimulus input. The speed of this activation is important in detailed studies of such semantic phenomena as listening with clear comprehension to a fast talker, or reading quickly texts of many different kinds. We listen and read quickly enough to keep up a fast pace. Any model of activation, and, therefore, of memory retrieval of the words heard or read, must satisfy such speed constraints. However, we have nothing to say here about the necessary process of identifying

whether an incoming stimulus is verbal in nature, and, if it is, what word is it, and how fast can it be retrieved from long-term memory.

The links between nodes, i.e., the links between the brain images of words, are just Hume's and later psychology's *associations*. All we assume here is that associations are something physical of a definite kind—we resist stating our own favored hypothesis about what this "definite kind" is. Links, like brain images of words, also can have just the same two states, quiescent or active.

In the three simple explicit examples of truth computation we give at the end, we introduce only a few words, which, besides the function words *is, of* and *the,* and the logical constant *not,* are just names of two cities and two countries along with the noun *capital.* Of course, the brain image of the word *Paris,* for example, has many associated brain images of properties, features or relations of the city of Paris, but we do not introduce them here, for they are not activated, and thus not needed, in our examples. But we stress they surely occur in bewildering variety, and any serious empirical model would need to try to survey the most significant ones. Here, we introduce only one property in our examples, the 1–1 *property* of the relation of being a capital in the ordinary political sense, expressed by our use of the word *capital.* The association is between the brain image of the word and the brain image of the property.

Our notation for associations, introduced later, suggests the relation is symmetric, but this seems contrary to experience. For example, in talking about, say, Korean democracy, we associate to the practices of democracy in the United States or the European Union, not vice versa. On the other hand, experience, and also psychological experiments, show that running the relation in reverse, so to speak, against the dominant direction of association, can also occur in a quite natural way. The distinction we want to have is one of relative intensity, not absolute presence or absence. This important conceptual distinction, needed for detailed empirical work, we ignore here. We simply avoid, in the axioms stated later, any commitment about symmetry.

A concept essential here, but really never seen as far as we know, in any system of formal logic, is that of *spreading activation,* a concept that applies to the activation of brain images of words not activated directly by the word actually occurring as a stimulus. Axioms S1, S2 and S3 formulate the qualitative properties we need for our limited purposes. For good examples of the use of this concept in psychological theories and models of memory retrieval and related phenomena, some good referenees from the not too distant past are [1–7,9].

Logicians used to carefully formalized concepts, shed of all intuitive meaning, will find especially deviant our use of the concept of *familiar properties.* But here the usage is innocent, referring only to some simplified results assumed either in the initial state or implicitly understood as part of prior experience. Making it explicit for general purposes is something not possible in any practical sense, and a mistaken way to talk about experience in detail. How to deal with familiarity in more complicated contexts we leave to some other occasion, but Bayesian ideas would be at least of some formal use, even if not very substantive. A prime instance of being familiar is the example of the 1–1 property of the binary relation of being a capital, mentioned earlier.

Finally, we come to a non-familiar concept that has, all the same, a simple intuitive explanation. We need to be able to refer to the *associative core* of a sentenee S, in our

notation, $c(S)$. When we are faced with rapidly doing almost any kind of task, our brains strip it down to essentials and focus but little on the parts that do not matter. In the brain experiments referred to earlier (Suppes et al.), we presented persons simple auditory or visual geography sentences about every four seconds and asked them to judge each as true or false. Given this kind of task—and there are many like it in repetitious tasks of ordinary life—persons quickly learn to consider only the key reference words which vary in an otherwise fixed sentential context, or occur in a small number of such contexts. So, for example, the associative core of the sentence *Paris is the capital of France* is the string of brain images of the three words *Paris, capital* and *France,* for which we use the notation PARIS/CAPITAL/FRANCE. This last example takes us to our few special conventions of notation. We use capital letters to denote the brain images of words, e.g., PARIS is the brain image of *Paris*. We use as variables w_i for words and the corresponding capital letters W_i for brain images. The notation for the quiescent state of links is $\sim$, as in PARIS $\sim$ FRANCE, and the active state is $\approx$, as in ROME $\approx$ ITALY. For simplicity we do not use notation that differentiates the quiescent or active state of brain images serving as nodes in our associative networks. In more elaborate work this would be necessary.

3.2. *Statement of axioms*

3.2.1. *Axioms for the initial state*

Axiom I1. No brain images of words are activated; all those present are in the quiescent state.

Axiom I2. All the links between brain images of words are quiescent.

Axiom I3. There are no links between brain images of alethic words ('*true*', '*false*') and proper words.

Axiom I4. There are quiescent links between the brain images of '*false*' and '*not*', and between '*true*' and '*not*'.

3.2.2. *Axioms for activation*

Axiom A1. Given an input sentence $S(w_1, \ldots, w_n)$, the probability of the brain image of a word w_i $(1 \leqslant i \leqslant n)$ of S in a quiescent state being activated is equal to or greater than $1 - \epsilon_i$.

Axiom A2. Given an input sentence $S(w_1, \ldots, w_n)$, the events of the brain images of words w_i being activated are probabilistically independent of each other.

Axiom A3. If the brain images of two words are activated and there is a quiescent link between them, then this link becomes activated.

3.2.3. Axioms for true

Axiom T1. If the associative core $c(S)$ of a sentence S is activated, then a link is activated between the brain images of '*true*' and $c(S)$, and the brain image of '*true*' is activated.

3.2.4. Axioms for false

Axiom F1. If the brain images of '*true*' and '*not*' are activated, then the brain image of '*false*' is activated.

Axiom F2. If there is an activated link between the brain image of '*true*' and the associative core $c(S)$ of a sentence S, and the brain image of '*not*' is activated, then a link is activated between the brain image of '*false*' and the core $c(\textit{not}\, S)$ of the sentence *not S*.

3.2.5. Axioms for spreading activation

Axiom S1. If W is activated, then the most familiar properties associated with W are also activated.

Axiom S2. If $W_1 \approx W_2$ or $W_2 \approx W_3$, and $W_1 \sim W_3$, then the associative core $W_1W_2W_3$ is activated.

Axiom S3. If the associative core $W_1W_2W_3$ is activated, W_1' or W_3' is activated, $W_1' \neq W_1$ and $W_3' \neq W_3$, and $W_2 \approx$ 1–1, then the brain image of '*false*' is activated and the link is activated between the brain image of '*false*' and the activated associative core $W_1'W_2W_3$ or $W_1W_2W_3'$, as the case may be.

3.3. Examples

For all our examples we have the same initial state of the network, with all the links quiescent, e.g., PARIS $\sim$ CAPITAL, and after activation we use the notation PARIS $\approx$ CAPITAL. In the examples themselves we show only the activated links. And, in our notation, as remarked earlier, we do not distinguish between the quiescent and activated state of brain images. So in the first time point t1 of the first example, PARIS, it is assumed it is activated in accordance with Axiom A1.

Initial State:

PARIS $\sim$ CAPITAL, ROME $\sim$ CAPITAL

FRANCE $\sim$ CAPITAL, ITALY $\sim$ CAPITAL

PARIS $\sim$ FRANCE, ROME $\sim$ ITALY

CAPITAL $\sim$ 1–1

TRUE $\sim$ NOT, FALSE $\sim$ NOT

Example 1. Paris is the capital of France.

$t1$. PARIS	Ax A1
$t2$. CAPITAL, CAPITAL $\approx$ 1–1	Ax A1, SI
$t3$. FRANCE, PARIS $\approx$ CAPITAL	Ax A1, A3
$t4$. CAPITAL $\approx$ FRANCE, PARIS $\approx$ FRANCE	Ax A3
$t5$. PARIS/CAPITAL/FRANCE	Ax S2
$t6$. TRUE $\approx$ PARIS/CAPITAL/FRANCE	Ax T1

Example 2. Paris is not the capital of France.

$t1$. PARIS	Ax AI
$t2$. NOT, CAPITAL, CAPITAL $\approx$ 1–1	Ax A1, S1
$t3$. FRANCE, PARIS $\approx$ CAPITAL	Ax AI, A3
$t4$. CAPITAL $\approx$ FRANCE, PARIS $\approx$ FRANCE	Ax A3
$t5$. PARIS/CAPITAL/FRANCE	Ax S2
$t6$. TRUE $\approx$ PARIS/CAPITAL/FRANCE	Ax T1
$t7$. FALSE	Ax F1
$t8$. FALSE $\approx$ PARIS/NOT/CAPITAL/FRANCE	Ax F2

Example 3. Rome is the capital of France.

$t1$. ROME	Ax AI
$t1$. CAPITAL, CAPITAL $\approx$ 1–1	Ax A1, S1
$t3$. FRANCE, ROME $\approx$ CAPITAL	Ax A1, A3
$t4$. CAPITAL $\approx$ FRANCE	Ax A3
$t5$. PARIS/CAPITAL/FRANCE ROME/CAPITAL/ITALY	Ax S2
$t6$. TRUE $\approx$ PARIS/CAPITAL/FRANCE TRUE $\approx$ ROME/CAPITAL/ITALY	Ax T1
$t7$. FALSE $\approx$ ROME/CAPITAL/FRANCE	Ax S3

4. Some philosophical consequences of the theory

From our point of view, for a normal adult speaker of English there are no basic differences between the truth of "15 + 29 = 44" and "Paris is the capital of France": both are the result of performance computations on already learned associative networks. It seems therefore difficult to say that the truth of "15 + 29 = 44" is analytic or a priori in opposition to the truth of "Paris is the capital of France" being synthetic or a posteriori.

We can, however, draw some distinctions between the computation of truth of different sentences. An important difference about the computation of different sentences is the time

necessary to perform the computation. But can we say that the truth of sentences whose truth computation requires less time is analytic or a priori rather than synthetic or a fortiori? Then what is the time constant which will mark the difference? From this point of view a negative or false sentence would often be less analytic than a true sentence.

It is better to focus on the level of activation necessary to compute the truth of a sentence. If no brain images of proper words other than those of the words of the sentence were necessary to compute the truth of it, then we could say that the truth or falsity of the sentence is analytic, and that the truth or falsity would be synthetic if it required the activation of other words not in the sentence. But it is not the purpose of this article to defend in any detail this reformulation.

We could distinguish the pair analytic/synthetic from the pair a priori/a posteriori by saying that the truth of a sentence would be a priori if its computation did not require something else outside of the network. Using this terminology, we could say that in our examples we have considered only a priori truth. In contrast, a posteriori truth depends on the use of external perceptions, such as looking up some fact in an atlas. This empirical construct of the analytic and a priori would fit into a philosophical tradition that goes back at least to John Stuart Mill.

In our context of ordinary usage, we do not see analytical truths as tautological truths or trivial identities of the type 'The capital of France is the capital of France'. In fact, if we ask the question 'Is the capital of France the capital of France?' to an ordinary man, he may have some difficulty in understanding what we are asking, and he takes more time to give a positive answer to this question than to a question like 'Is Paris the capital of France?', because the trivial identity question sounds like nonsense. Much the same can be said about less tautological questions such as 'Is the capital of France in France?'.

References

[1] J.R. Anderson, Retrieval of prepositional information from long-term memory, Cognitive Psychol. 5 (1974) 451–474.

[2] J.R. Anderson, Item-specific and relation specific interference in sentence memory, J. Exp.: Human Learning and Memory 104 (1975) 249–260.

[3] J.R. Anderson, A spreading activation theory of memory, J. Verbal Learning and Verbal Behavior 22 (1983) 261–295.

[4] J.R. Anderson, G.H. Bower, Recognition and retrieval processes in free recall, Psychological Rev. 79 (1972) 97–123.

[5] J.R. Anderson, G.H. Bower, Human Associative Memory, Winston, Washington, 1973.

[6] R.C. Atkinson, J.F. Juola, Factors influencing speed and accuracy in word recognition, in: S. Kornblum (Ed.), Attention and Performance, vol. IV, Academic Press, New York, 1973.

[7] R.C. Atkinson, D.J. Herrmann, K.T. Wescourt, Search processes in recognition memory, in: R.L. Solso (Ed.), Theories in Cognitive Psychology, Lawrence Erlbaum, Hillsdale, NJ, 1974.

[8] P. Blackburn, J. Bos, Representation and inference for natural language (A first course in computational semantics), http://www.comsem.org, 1999.

[9] W.K. Estes, Some targets for mathematical psychology, J. Math. Psychol. 12 (1975) 263–282.

[10] C. Fellbaum, Wordnet, an Electronic Lexical Database for English, MIT Press, Cambridge, MA, 1998.

[11] W. James, The Principles of Psychology, vol. I, Henry Holt and Company, New York, 1890.

[12] F. Lehman, Semantic networks, in: F. Lehman (Ed.), Semantics Networks in Artificial Intelligence, Pergamon Press, 1992, pp. 1–50.

[13] L. Shastri, V. Ajjanagadde, From simple associations to systematic reasoning: a connectionist representation of rules, variables and dynamic bindings using temporal synchrony, Behavioral and Brain Sciences 16 (1993) 417–494.
[14] M. Sigman, G.A. Cecchi, Global organization of the wordnet lexicon, Proc. Natl. Acad. Sci. 99 (2002) 1742–1747.
[15] P. Suppes, Stimulus-response theory of finite automata, J. Math. Psychol. 6 (1969) 327–355.
[16] P. Suppes, Z.-L. Lu, B. Han, Brain wave recognition of words, Proc. Natl. Acad. Sci. 94 (1997) 14965–14969.
[17] P. Suppes, B. Han, J. Epelboim, Z.-L. Lu, Invariance between subjects of brain wave representations of language, Proc. Natl. Acad. Sci. 96 (1999) 12953–12958.
[18] P. Suppes, B. Han, J. Epelboim, Z.-L. Lu, Invariance of brainwave representations of simple visual images and their names, Proc. Natl. Acad. Sci. 96 (1999) 14658–14663.
[19] W.A. Woods, What's in a link? Foundations for semantic networks, in: R.J. Brachman, H.J. Levesque (Eds.), Readings in Knowledge Representation, Morgan Kaufmann, Los Altos, CA, 1985, pp. 217–241.