Abstract

This paper proposes a method for measuring semantic similarity between words as a new tool for text analysis. The similarity is measured on a semantic network constructed systematically from a subset of the English dictionary, LDOCE (Longman Dictionary of Contemporary English). Spreading activation on the network can directly compute the similarity between any two words in the Longman Defining Vocabulary, and indirectly the similarity of all the other words in LDOCE. The similarity represents the strength of lexical cohesion or semantic relation, and also provides valuable information about similarity and coherence of texts.

1 Introduction

A text is not just a sequence of words, but it also has coherent structure. The meaning of each word in a text depends on the structure of the text. Recognizing the structure of text is an essential task in text understanding. [Grosz and Sidner, 1986]

One of the valuable indicators of the structure of text is lexical cohesion. [Halliday and Hasan, 1976] Lexical cohesion is the relationship between words, classified as follows:

1. Reiteration:
   Molly likes cats. She keeps a cat.
2. Semantic relation:
   a. Desmond saw a cat. It was Molly’s pet.
   b. Molly goes to the north. Not east.
   c. Desmond goes to a theatre. He likes films.

Reiteration of words is easy to capture by morphological analysis. Semantic relation between words, which is the focus of this paper, is hard to recognize by computers.

We consider lexical cohesion as semantic similarity between words. Similarity is computed by spreading activation (or association) [Waltz and Pollack, 1985] on a semantic network constructed systematically from an English dictionary. Whereas it is edited by some lexicographers, a dictionary is a set of associative relation shared by the people in a linguistic community.

The similarity between words is a mapping \( \sigma : L \times L \rightarrow [0, 1] \), where \( L \) is a set of words (or lexicon). The following examples suggest the feature of the similarity:

\[
\sigma(\text{cat}, \text{pet}) = 0.133722 \quad (\text{similar}),
\]
\[
\sigma(\text{cat}, \text{mat}) = 0.002692 \quad (\text{dissimilar}).
\]

The value of \( \sigma(w, w') \) increases with strength of semantic relation between \( w \) and \( w' \).

The following section examines related work in order to clarify the nature of the semantic similarity. Section 3 describes how the semantic network is systematically constructed from the English dictionary. Section 4 explains how to measure the similarity by spreading activation on the semantic network. Section 5 shows applications of the similarity measure — computing similarity between texts, and measuring coherence of a text. Section 6 discusses the theoretical aspects of the similarity.
2 Related Work on Measuring Similarity

Words in a language are organized by two kinds of relationship. One is a syntagmatic relation: how the words are arranged in sequential texts. The other is a paradigmatic relation: how the words are associated with each other. Similarity between words can be defined by either a syntagmatic or a paradigmatic relation.

Syntagmatic similarity is based on co-occurrence data extracted from corpora [Church and Hanks, 1990], definitions in dictionaries [Wilks et al., 1989], and so on. Paradigmatic similarity is based on association data extracted from thesauri [Morris and Hirst, 1991], psychological experiments [Osgood, 1952], and so on.

This paper concentrates on paradigmatic similarity, because a paradigmatic relation can be established both inside a sentence and across sentence boundaries, while syntagmatic relations can be seen mainly inside a sentence — like syntax deals with sentence structure. The rest of this section focuses on two related works on measuring paradigmatic similarity — a psycholinguistic approach and a thesaurus-based approach.

2.1 A Psycholinguistic Approach

Psycholinguists have been proposed methods for measuring similarity. One of the pioneering works is ‘semantic differential’ [Osgood, 1952] which analyses meaning of words into a range of different dimensions with the opposed adjectives at both ends (see Figure 1), and locates the words in the semantic space.

Recent works on knowledge representation are somewhat related to Osgood’s semantic differential. Most of them describe meaning of words using special symbols like microfeatures [Waltz and Pollack, 1985].

![Figure 1. A psycholinguistic measurement (semantic differential [Osgood, 1952]).](image)

2.2 A Thesaurus-based Approach

Morris and Hirst [1991] used Roget’s thesaurus as knowledge base for determining whether or not two words are semantically related. For example, the semantic relation of truck/car and drive/car are captured in the following way:

1. truck ∈ vehicle ⊃ car
   (both are included in the vehicle class),

2. drive ∈ journey → vehicle ⊃ car
   (journey refers to vehicle).

This method can capture almost all types of semantic relations (except emotional and situational relation), such as paraphrasing by superordinate (ex. cat/pet), systematic relation (ex. north/east), and non-systematic relation (ex. theatre/film).

However, thesauri provide neither information about semantic difference between words juxtaposed in a category, nor about strength of the semantic relation between words — both are to be dealt in this paper. The reason is that thesauri are designed to help writers find relevant words, not to provide the meaning of words.

3 Paradigme: A Field for Measuring Similarity

We analyse word meaning in terms of the semantic space defined by a semantic network, called Paradigme. Paradigme is systematically constructed from Glossème, a subset of an English dictionary.

3.1 Glossème — A Closed Subsystem of English

A dictionary is a closed paraphrasing system of natural language. Each of its headwords is defined by a phrase which is composed of the headwords and their derivations. A dictionary, viewed as a whole, looks like a tangled network of words.

We adopted Longman Dictionary of Contemporary English (LDOCE) [1987] as such a closed system of English. LDOCE has a unique feature that each of its 56,000 headwords is defined by using the words in Longman Defining Vocabulary (hereafter, LDV) and their derivations. LDV consists of 2,851 words (as
red\(^1\) /red/ adj -dd- \(^\text{1}\) of the colour of blood or fire: a red rose/dress | We painted the door red. — see also like a red rag to a bull \((\text{rag}^{1}\) \(^\text{2}\) of human hair) of a bright brownish orange or copper colour \(^3\) (of the human skin) pink, usu. for a short time: I turned red with embarrassment/anger. | The child’s eye (= the skin round the eyes) were red from crying. \(^4\) (of wine) of a dark pink to dark purple colour

— ∼ness

\[\text{red adj ; headword, word-class}
\]
\[\text{((of the colour) ; unit 1 -- head-part}
\]
\[\text{((of blood or fire) ) ; det-part}
\]
\[\text{((of a bright brownish orange or copper colour)
}\]
\[\text{((of human hair) )
\]
\[\text{((of a bright brownish orange or copper colour)
}\]
\[\text{((of human hair) )
\]
\[\text{((of a bright brownish orange or copper colour)
}\]
\[\text{((of wine) )
\]

Figure 2. A sample entry of LDOCE and a corresponding entry of Glossème (in S-expression).

\[(\text{red}_1 \text{ (adj) 0.000000 ;; headword, word-class, and activity-value
}\]
\[\text{;; referant
}\]
\[\text{(0.333333 ;; weight of referant 1
}\]
\[\text{(* (0.001594 of_1) (0.001733 the_1) (0.001733 the_2) (0.042108 colour_1)
}\]
\[\text{(0.185058 fire_1) (0.185058 fire_2) ))
}\]
\[\text{;; referant 2
}\]
\[\text{(0.277778
}\]
\[\text{(* (0.00278 of_1) (0.000196 a_1) (0.030997 bright_1) (0.065587 brown_1)
}\]
\[\text{(0.466411 orange_1) (0.000139 of_1) (0.007330 colour_1)
}\]
\[\text{(0.185058 fire_1) (0.185058 fire_2) ))
}\]
\[\text{;; referant 3
}\]
\[\text{(0.163722
}\]
\[\text{(* (0.000328 of_1) (0.000232 a_1) (0.028368 dark_1) (0.028368 dark_2)
}\]
\[\text{(0.123290 to_1) (0.123290 dark_2) (0.00273 to_1) (0.000273 to_2)
}\]
\[\text{(0.000273 to_3) (0.028368 dark_1) (0.028368 dark_2) (0.141273 purple_1)
}\]
\[\text{(0.141273 colour_2) (0.008673 colour_1) (0.008673 colour_2) (0.000164 of_1)
}\]
\[\text{(0.338512 wine_1) ))
}\]
\[\text{;; referent
}\]
\[\text{(0.031058 apple_1) (0.098349 apple_2) (0.028100 purple_2)
}\]
\[\text{(0.028100 purple_2) (0.098349 red_2)
}\]

Figure 3. A sample node of Paradigme (in S-expression).

the headwords in LDOCE) based on the survey of restricted vocabulary \[^{[\text{West, 1953}]}\].

We made a reduced version of LDOCE, called Glossème. Glossème has every entry of LDOCE whose headword is included in LDV. Thus, LDV is defined by Glossème, and Glossème is composed of LDV. Glossème is a closed subsystem of English.

Glossème has 2,851 entries that consist of 101,861 words (35.73 words/entry on the average). An item of Glossème has a headword, a word-class, and one or more units corresponding to numbered definitions in the entry of LDOCE. Each unit has one head-part and several det-parts. The head-part is the first phrase in the definition, which describes the broader meaning of the headword. The det-parts restrict the meaning of the head-part. (See Figure 2.)

3.2 Paradigme — A Semantic Network

We then translated Glossème into a semantic network Paradigme. Each entry in Glossème is mapped onto a node in Paradigme. Paradigme has 2,851 nodes and 295,914 unnamed links between the nodes (103.79 links/node on the average). Figure 3 shows a sample node red\(_1\). Each node consists of a headword, a word-class, an activity-value, and two sets of links: a référent and a réfééré.

A référent of a node consists of several subréférents corresponding to the units of Glossème. As shown in Figure 2 and 3, a morphological analysis maps the word brownish in the second unit onto a link to the node brown\(_1\), and the word colour onto two links to colour\(_1\) (adjective) and colour\(_2\) (noun).
4 Computing Similarity between Words

Similarity between words is computed by spreading activation on \textit{Paradigme}. Each of its nodes can hold activity, and it moves through the links. Each node computes its activity value \(v_i(T+1)\) at time \(T+1\) as follows:

\[
v(T+1) = \phi \left( R_i(T), R'_i(T), e_i(T) \right),
\]

where \(R_i(T)\) and \(R'_i(T)\) are the sum of weighted activity (at time \(T\)) of the nodes referred in the référent and référe respectively. And, \(e_i(T)\) is activity given from outside (at time \(T\)); to 'activate a node' is to let \(e_i(T) > 0\). The output function \(\phi\) sums up three activity values in appropriate proportion and limits the output value to \([0,1]\). Appendix B gives the details of the spreading activation.

4.1 Measuring Similarity

Activating a node for a certain period of time causes the activity to spread over \textit{Paradigme} and produce an activated pattern on it. The activated pattern approximately gets equilibrium after 10 steps, whereas it will never reach the actual equilibrium. The pattern thus produced represents the meaning of the node or of the words related to the node by morphological analysis.

The activated pattern, produced from a word \(w\), suggests similarity between \(w\) and any headword in LDV. The similarity \(\sigma(w, w') \in [0,1]\) is computed in the following way. (See also Figure 4.)

1. Reset activity of all nodes in \textit{Paradigme}.
2. Activate \(w\) with strength \(s(w)\) for 10 steps, where \(s(w)\) is significance of the word \(w\).
3. Then, an activated pattern \(P(w)\) is produced on \textit{Paradigme}.
4. Observe \(a(P(w), w')\) — an activity value of the node \(w'\) in \(P(w)\).
5. Then, \(\sigma(w, w') = s(w') \cdot a(P(w), w')\).

The word significance \(s(w) \in [0,1]\) is defined as the normalized information of the word \(w\) in the corpus \citep{West1953}. For example, the word \textit{red} appears 2,308 times in the 5,487,056-word corpus, and the word \textit{and} appears 106,064 times. So, \(s(\text{red})\) and \(s(\text{and})\) are computed as follows:

\[
s(\text{red}) = \frac{- \log(2308/5487056)}{- \log(1/5487056)} = 0.500955 ,
\]

\[
s(\text{and}) = \frac{- \log(106064/5487056)}{- \log(1/5487056)} = 0.254294 .
\]

We estimated the significance of the words excluded from the word list \citep{West1953} at the average significance of their word classes. This interpolation virtually enlarged West's 5,000,000-word corpus.

For example, let us consider the similarity between \textit{red} and \textit{orange}. First, we produce an activated pattern \(P(\text{red})\) on \textit{Paradigme}. (See Figure 5.) In this case, both of the nodes \textit{red.1} (adjective) and \textit{red.2} (noun) are activated with strength \(s(\text{red}) = 0.500955\). Next, we compute \(s(\text{orange}) = 0.676253\), and observe \(a(P(\text{red}), \text{orange}) = 0.390774\). Then, the similarity between \textit{red} and \textit{orange} is obtained as follows:

\[
\sigma(\text{red}, \text{orange}) = 0.676253 \cdot 0.390774 = 0.264262 .
\]

\footnote{The morphological analysis maps all words derived by 48 affixes in LDV onto their root forms (i.e., headwords of LDOCE).}
4.2 Examples of Similarity between Words

The procedure described above can compute the similarity $\sigma(w, w')$ between any two words $w$, $w'$ in LDV and their derivations. Computer programs of this procedure — spreading activation (in C), morphological analysis and others (in Common Lisp) — can compute $\sigma(w, w')$ within 2.5 seconds on a workstation (SPARCstation 2).

The similarity $\sigma$ between words works as an indicator of the lexical cohesion. The following examples illustrate that $\sigma$ increases with the strength of semantic relation:

- $\sigma(\text{wine}, \text{alcohol}) = 0.118078$,
- $\sigma(\text{wine}, \text{line}) = 0.002040$,
- $\sigma(\text{big}, \text{large}) = 0.120587$,
- $\sigma(\text{clean}, \text{large}) = 0.004943$,
- $\sigma(\text{buy}, \text{sell}) = 0.135686$,
- $\sigma(\text{buy}, \text{walk}) = 0.007993$.

The similarity $\sigma$ also increases with the co-occurrence tendency of words, for example:

- $\sigma(\text{waiter}, \text{restaurant}) = 0.175699$,
- $\sigma(\text{computer}, \text{restaurant}) = 0.003268$,
- $\sigma(\text{red}, \text{blood}) = 0.111443$,
- $\sigma(\text{green}, \text{blood}) = 0.002268$,
- $\sigma(\text{dig}, \text{spade}) = 0.116200$,
- $\sigma(\text{fly}, \text{spade}) = 0.003431$.

Note that $\sigma(w, w')$ has direction (from $w$ to $w'$), so that $\sigma(w, w')$ may not be equal to $\sigma(w', w)$:

- $\sigma(\text{films}, \text{theatre}) = 0.178988$,
- $\sigma(\text{theatre}, \text{films}) = 0.068927$.

Meaningful words should have higher similarity; meaningless words (especially, function words) should have lower similarity. The similarity $\sigma(w, w')$ increases with the significance $s(w)$ and $s(w')$ that represent meaningfulness of $w$ and $w'$:

- $\sigma(\text{north}, \text{east}) = 0.100482$,
- $\sigma(\text{to}, \text{theatre}) = 0.007259$,
- $\sigma(\text{films}, \text{of}) = 0.005914$,
- $\sigma(\text{to}, \text{the}) = 0.002240$.

Note that the reflective similarity $\sigma(w, w)$ also depends on the significance $s(w)$, so that $\sigma(w, w) \leq 1$:

\[
\sigma(\text{wine}, \text{wine}) = 0.003431 ,
\]

\[
\sigma(\text{wine}, \text{wine}) = 0.003431.
\]

4.3 Similarity of Extra Words

The similarity of words in LDV and their derivations is measured directly on Paradigme; the similarity of extra words is measured indirectly on Paradigme by treating an extra word as a word list $W = \{w_1, \ldots, w_n\}$ of its definition in LDOCE. (Note that each $w_i \in W$ is included in LDV or their derivations.)

The similarity between the word lists $W, W'$ is defined as follows. (See also Figure 6.)

\[
\sigma(W, W') = \psi \left( \sum_{w \in W} s(w') \cdot \alpha(P(W), w') \right),
\]

where $P(W)$ is the activated pattern produced from $W$ by activating each $w_i \in W$ with strength $s(w_i)^2 / \sum s(w_k)$ for 10 steps. And, $\psi$ is an output function which limits the value to $[0, 1]$.

As shown in Figure 7, bottle-l and wine-l have high activity in the pattern produced from the phrase “red alcoholic drink”. So, we may say that the overlapped pattern implies “a bottle of wine”.

For example, the similarity between linguistics and stylistics, both are the extra words, is computed as follows:

\[
\sigma(\text{linguistics, stylistics}) = \sigma(\{\text{the, study, of, language, in, general, and, of, particular, languages, and, their, structure, and, grammar, and, history}\} ),
\]

\[
\sigma(\text{linguistics, stylistics}) = \sigma(\{\text{the, study, of, language, in, general, and, of, particular, languages, and, their, structure, and, grammar, and, history}\} ).
\]
5 Applications of the Similarity

This section shows the application of the similarity between words — measuring similarity between texts, and measuring text coherence.

5.1 Measuring Similarity between Texts

Suppose a text is a word list without syntactic structure. Then, the similarity \( \sigma(X, X') \) between two texts \( X, X' \) can be computed as the similarity of extra words described above.

The following examples suggest that the similarity between texts indicates the strength of coherence relation between them:

\[
\sigma(\text{"I have a hammer."}, \\
\text{"Take some nails."}) = 0.100611, \\
\sigma(\text{"I have a hammer."}, \\
\text{"Take some apples."}) = 0.005295, \\
\sigma(\text{"I have a pen."}, \\
\text{"Where is ink?"}) = 0.113140, \\
\sigma(\text{"I have a pen."}, \\
\text{"Where do you live?"}) = 0.007676.
\]

It is worth noting that meaningless iteration of words (especially, of function words) has less influence on the text similarity:

\[
\sigma(\text{"It is a dog."}, \\
\text{"That must be your dog."}) = 0.252536, \\
\sigma(\text{"It is a dog."}, \\
\text{"It is a log."}) = 0.053261.
\]

The text similarity provides a semantic space for text retrieval — to recall the most similar text in \( \{X_1', \ldots, X_n'\} \) to the given text \( X \). Once the activated pattern \( P(X) \) of the text \( X \) is produced on Paradigme, we can compute and compare the similarity \( \sigma(X, X_1'), \ldots, \sigma(X, X_n') \) immediately. (See Figure 8.)

5.2 Measuring Text Coherence

Let us consider the reflective similarity \( \sigma(X, X) \) of a text \( X \), and use the notation \( c(X) \) for \( \sigma(X, X) \).

Then, \( c(X) \) can be computed as follows:

\[
c(X) = \psi(\sum_{w \in X} s(w)a(P(X), w)).
\]

The activated pattern \( P(X) \), as shown in Figure 7, represents the average meaning of \( w_i \in X \). So, \( c(X) \) represents cohesiveness of \( X \) — or semantic closeness of \( w \in X \), or semantic compactness of \( X \). (It is also closely related to distortion in clustering.)

The following examples suggest that \( c(X) \) indicates the strength of coherence of \( X \):

\[
c(\text{"She opened the world with her typewriter. Her work was typing. But She did not type quickly."}) = 0.502510 \quad \text{(coherent)}, \\
c(\text{"Put on your clothes at once. I can not walk ten miles. There is no one here but me."}) = 0.250840 \quad \text{(incoherent)}.
\]

However, a cohesive text can be incoherent; the following example shows cohesiveness of the incoherent text — three sentences randomly selected from LDOCE:
"I saw a lion.
A lion belongs to the cat family.
My family keeps a pet."

= 0.560172 (incoherent, but cohesive).

Thus, \( c(X) \) can not capture all the aspects of text coherence. This is because \( c(X) \) is based only on the lexical cohesion of the words in \( X \).

6 Discussion

The structure of Paradigme represents the knowledge system of English, and an activated state produced on it represents word meaning. This section discusses the nature of the structure and states of Paradigme, and also the nature of the similarity computed on it.

6.1 Paradigme and Semantic Space

The set of all the possible activated patterns produced on Paradigme can be considered as a semantic space where each state is represented as a point. The semantic space is a 2,851-dimensional hypercube; each of its edges corresponds to a word in LDV.

LDV is selected according to the following information: the word frequency in written English, and the range of contexts in which each word appears. So, LDV has a potential for covering all the concepts commonly found in the world. This implies the completeness of LDV as dimensions of the semantic space. Osgood’s semantic differential procedure [1952] used 50 adjective dimensions; our semantic measurement uses 2,851 dimensions with completeness and objectivity.

Our method can be applied to construct a semantic network from an ordinary dictionary whose defining vocabulary is not restricted. Such a network, however, is too large to spread activity over it. Paradigme is the small and complete network for measuring the similarity.

6.2 Connotation and Extension of Words

The proposed similarity is based only on the denotational and intensional definitions in the dictionary LDOCE. Lack of the connotational and extensional knowledge causes some unexpected results of measuring the similarity. For example, consider the following similarity:

\[
\sigma(\text{tree}, \text{leaf}) = 0.008693 .
\]

This is due to the nature of the dictionary definitions — they only indicate sufficient conditions of the headword. For example, the definition of tree in LDOCE tells nothing about leaves:

\[
\text{tree } n \ 1 \ a \ tall \ plant \ with \ a \ wooden \ trunk \ and \ branches, \ that \ lives \ for \ many \ years \ 2 \ a \ bush \ or \ other \ plant \ with \ a \ treelike \ form \ 3 \ a \ drawing \ with \ a \ branching \ form, \ esp. \ as \ used \ for \ showing \ family \ relationships
\]

However, the definition is followed by pictures of leafy trees providing readers with connotational and extensional stereotypes of trees.

6.3 Paradigmatic and Syntagmatic Similarity

In the proposed method, the definitions in LDOCE are treated as word lists, though they are phrases with syntactic structures. Let us consider the following definition of lift:

\[
\text{lift } v \ 1 \ to \ bring \ from \ a \ lower \ to \ a \ higher \ level; \raise \ 2 \ (of \ movable \ parts) \ to \ be \ able \ to \ be \ lifted \ 3 \ \ldots
\]

Anyone can imagine that something is moving upward. But, such a movement can not be represented in the activated pattern produced from the phrase. The meaning of a phrase, sentence, or text should be represented as pattern changing in time, though what we need is static and paradigmatic relation.

This paradox also arises in measuring the similarity between texts and the text coherence. As we have seen in Section 5, there is a difference between the similarity of texts and the similarity of word lists, and also between the coherence of a text and cohesiveness of a word list.

However, so far as the similarity between words is concerned, we assume that activated patterns on Paradigme will approximate the meaning of words, like a still picture can express a story.

7 Conclusion

We described measurement of semantic similarity between words. The similarity between words is computed by spreading activation on the semantic network Paradigme which is systematically constructed from a subset of the English dictionary LDOCE. Paradigme can directly compute the similarity between any two words in LDV, and indirectly the similarity of all the other words in LDOCE.

The similarity between words provides a new method for analysing the structure of text. It can be applied to computing the similarity between texts, and measuring the cohesiveness of a text which suggests coherence of the text, as we have seen in Section 5. And, we are now applying it to text segmentation [Grosz and Sidner, 1986, Youmans, 1991], i.e. to capture the shifts of coherent scenes in a story.
In future research, we intend to deal with syntagmatic relations between words. Meaning of a text lies in the texture of paradigmatic and syntagmatic relations between words [Hjelmlev, 1943]. Paradigme provides the former dimension — an associative system of words — as a screen onto which the meaning of a word is projected like a still picture. The latter dimension — syntactic process — will be treated as a film projected dynamically onto Paradigme. This enables us to measure the similarity between texts as a syntactic process, not as word lists.

We regard Paradigme as a field for the interaction between text and episodes in memory — the interaction between what one is hearing or reading and what one knows [Schank, 1990]. The meaning of words, sentences, or even texts can be projected in a uniform way on Paradigme, as we have seen in Section 4 and 5. Similarly, we can project text and episodes, and recall the most relevant episode for interpretation of the text.

Appendix A. Structure of Paradigme — Mapping Glossème onto Paradigme

The semantic network Paradigme is systematically constructed from the small and closed English dictionary Glossème. Each entry of Glossème is mapped onto a node of Paradigme in the following way. (See also Figure 2 and 3.)

**Step 1.** For each entry \( G_i \) in Glossème, map each unit \( u_{ij} \) in \( G_i \) onto a subréférent \( s_{ij} \) of the corresponding node \( P_i \) in Paradigme. Each word \( w_{ijn} \in u_{ij} \) is mapped onto a link or links in \( s_{ij} \), in the following way:

1. Let \( t_n \) be the reciprocal of the number of appearance of \( w_{ijn} \) (as its root form) in Glossème.
2. If \( w_{ijn} \) is in a head-part, let \( t_n \) be doubled.
3. Find nodes \( \{p_{n1}, p_{n2}, \ldots\} \) corresponds to \( w_{ijn} \) (ex. red \( \rightarrow \) \{red.1, red.2\}). Then, divide \( t_n \) into \( \{t_{n1}, t_{n2}, \ldots\} \) in proportion to their frequency.
4. Add links \( l_{n1}, l_{n2}, \ldots \) to \( s_{ij} \), where \( l_{nm} \) is a link to the node \( p_{nm} \) with thickness \( t_{nm} \).

Thus, \( s_{ij} \) becomes a set of links: \( \{l_{ij1}, l_{ij2}, \ldots\} \), where \( l_{ijk} \) is a link with thickness \( t_{ijk} \). Then, normalize thickness of the links as \( \sum_k t_{ijk} = 1 \), in each \( s_{ij} \).

**Step 2.** For each node \( P_i \), compute thickness \( h_{ij} \) of each subréférent \( s_{ij} \) in the following way:

1. Let \( m_i \) be the number of subréférents of \( P_i \).
2. Let \( h_{ij} \) be \( 2m_i - 1 - j \).
   (Note that \( h_{11} : h_{im} = 2 : 1 \).)
3. Normalize thickness \( h_{ij} \) as \( \sum_j h_{ij} = 1 \), in each \( P_i \).

**Step 3.** Generate référé of each node in Paradigme, in the following way:

1. For each node \( P_i \) in Paradigme, let its référé \( r_i \) be an empty set.
2. For each \( P_i \), for each subréférent \( s_{ij} \) of \( P_i \), for each link \( l_{ijk} \) in \( s_{ij} \):
   a. Let \( p_{ijk} \) be the node referred by \( l_{ijk} \), and let \( t_{ijk} \) be thickness of \( l_{ijk} \).
   b. Add a new link \( l' \) to référé of \( p_{ijk} \), where \( l' \) is a link to \( P_i \) with thickness \( t' = h_{ij} \cdot t_{ijk} \).
3. Thus, each \( r_i \) becomes a set of links: \( \{l'_{i1}, l'_{i2}, \ldots\} \), where \( l'_{ij} \) is a link with thickness \( t_{ij}^* \). Then, normalize thickness of the links as \( \sum_j t_{ij}^* = 1 \), in each \( r_i \).

Appendix B. Function of Paradigme — Spreading Activation Rules

Each node \( P_i \) of the semantic network Paradigme computes its activity value \( v_i(T+1) \) at time \( T+1 \) as follows:

\[
v_i(T+1) = \phi \left( \frac{R_i(T) + R_i'(T)}{2} + e_i(T) \right),
\]

where \( R_i(T) \) and \( R_i'(T) \) are activity (at time \( T \)) collected from the nodes referred in the référent and référé respectively; \( e_i(T) \in [0, 1] \) is activity given from outside (at time \( T \)); the output function \( \phi \) limits the value to \([0,1]\).

\( R_i(T) \) is activity of the most plausible subréférent in \( P_i \), defined as follows:

\[
R_i(T) = S_{im}(T),
\]

\[
m = \arg\max_i \{h_{ij} \cdot S_{ij}(T)\},
\]

where \( h_{ij} \) is thickness of the \( j \)-th subréférent of \( P_i \).

\( S_{ij}(T) \) is the sum of weighted activity of the nodes referred in the \( j \)-th subréférent of \( P_i \), defined as follows:

\[
S_{ij}(T) = \sum_k t_{ijk} \cdot a_{ijk}(T),
\]

where \( t_{ijk} \) is thickness of the \( k \)-th link of \( s_{ij} \), and \( a_{ijk}(T) \) is activity (at time \( T \)) of the node referred by the \( k \)-th link of \( s_{ij} \).

\( R_i'(T) \) is weighted activity of the nodes referred in the référé \( r_i \) of \( P_i \):

\[
R_i'(T) = \sum_k t'_{ik} \cdot a'_{ik}(T),
\]

where \( t'_{ik} \) is thickness of the \( k \)-th link of \( r_i \), and \( a'_{ik} \) is activity (at time \( T \)) of the node referred by the \( k \)-th link of \( r_i \).
References

[Church and Hanks, 1990] K. Church and P. Hanks. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16:22–29, 1990.

[Grosz and Sidner, 1986] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

[Halliday and Hasan, 1976] M. A. K. Halliday and R. Hasan. *Cohesion in English*. Longman, Harlow, Essex, 1976.

[Hendler, 1989] J. A. Hendler. Marker-passing over microfeatures: Towards a hybrid symbolic / connectionist model. *Cognitive Science*, 13:79–106, 1989.

[Hjelmslev, 1943] L. Hjelmslev. *Omkring Sprogteoriens Grundlæggelse*. Akademisk Forlag, København, 1943.

[LDO, 1987] *Longman Dictionary of Contemporary English*. Longman, Harlow, Essex, new edition, 1987.

[Morris and Hirst, 1991] J. Morris and G. Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17:21–48, 1991.

[Osgood, 1952] C. E. Osgood. The nature and measurement of meaning. *Psychological Bulletin*, 49:197–237, 1952.

[Schank, 1990] R. C. Schank. *Tell Me a Story: A New Look at Real and Artificial Memory*. Scribner, New York, 1990.

[Waltz and Pollack, 1985] D. L. Waltz and J. B. Pollack. Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, 9:51–74, 1985.

[West, 1953] M. West. *A General Service List of English Words*. Longman, Harlow, Essex, 1953.

[Wilks et al., 1989] Y. Wilks, D. Fass, C. M. Guo, J. McDonald, T. Plate, and B. Slator. A tractable machine dictionary as a resource for computational semantics. In B. Boguraev and E. J. Briscoe, editors, *Computational Lexicography for Natural Language Processing*. Longman, Harlow, Essex, 1989.

[Youmans, 1991] G. Youmans. A new tool for discourse analysis: The vocabulary-management profile. *Language*, 67:763–789, 1991.