

A Note on Densities of the L2 Mental Lexicon Network*

Yusuke Higuchi^{1#}, Rumi Takahashi², Mikihiro Tanaka³

¹Mathematics Laboratories, College of Arts and Sciences, Showa University, Yamanashi, Japan

²Faculty of Arts and Sciences, Showa University, Yamanashi, Japan

³Department of English Language and Culture, Konan Women's University, Hyogo, Japan

Email: [#]higuchi@cas.showa-u.ac.jp

How to cite this paper: Higuchi, Yu., Takahashi, R., & Tanaka, M. (2019). A Note on Densities of the L2 Mental Lexicon Network. *Open Journal of Modern Linguistics*, 9, 476-507.

<https://doi.org/10.4236/ojml.2019.96037>

Received: October 18, 2019

Accepted: December 1, 2019

Published: December 4, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this note, we discuss a supplementary explanation about works of Takahashi and Tanaka, which are a kind of replication study of Wilks & Meara (2002), where they discuss the structure of learners' mental lexicon network from the viewpoint of psycholinguistics. Here we mainly discuss some mathematical structures which support their studies latently. Using some simple random graph techniques, we propose that the role of "densities" is important and useful under some suitable assumptions. Then, we exhibit the effectiveness of our method by applying it to the data of experiments in Takahashi and Tanaka, where some inequalities obtained mathematically help us to investigate some of the aspects of the L2 mental lexicon network. Mathematical proofs and detailed discussion are collected in the final section, in which we give also mathematical interpretation on some models which were shown by computer simulations in previous works.

Keywords

Mental Lexicon, Random Graph, Density, Complex Network, Second Language, Mathematical Modelling

1. Introduction

The basic object to examine "mental lexicon" is to investigate how humans organize huge "words" in their mind. Many researchers in various areas are attracted to the structure of mental lexicon and try to illustrate it. Many types of findings are found in the standard textbooks, for example, (Aitchison, 1987,

*Dedicated to the memory of Professor Yoichiro Takahashi.

2012) and (Boyd, 1993); an overview of various fields on psycholinguistics including mental lexicon is found in (Gaskell, 2009). There exist various methods or ideas due to the difference of the communities where researchers belong to, but no one will imagine each word in the mental lexicon is completely isolated. Thus it is natural to consider words are linked one another and the whole words form “word-web”, a kind of network. In this sense, a model of the mental lexicon becomes a network, that is, graph; a graph is a notion in mathematics and constructed by vertices corresponding to “words” and edges corresponding to relationships among words. Classically graphs are known to be applied effectively to chemistry, physics, social science and so on. Also in the context of linguistics and psychology, there exist many works on graph-theoretical mental lexicon. For instance, refer to (Kiss, 1968; Meara, 1992; Rapaport, Livant, & Boyd, 1966; Scott, 1991; Wilks & Meara, 2002; Wilks & Meara, 2007; Wilks, Meara, & Wolter, 2005). In many works, graph-invariants, which are some stable quantities measuring a given graph, in the structure of mental lexicon are discussed: the distribution of degrees, cluster coefficient, path length, community and so on. See also Marcus (2001) for the study on the mental lexicon with computer sciences. Moreover, after the original idea of complex network, or, scale-free network are brought by Barabási & Albert (1999) and Barabási (2009), one can see this concept describes many phenomena in sociology, biology or economics well. Therefore many researchers try to apply it to the mental lexicon: (Gruenenfelder & Mueller, 2007; Gruenenfelder & Pisoni, 2005, 2009; Schur, 2007; Stella, Beckage, Brede, & De Domenico, 2018; Vitevitch, 2008; Vitevitch, Goldstein, Siew, & Castro, 2014), for instance. In summary, graph-theoretical approach for the mental lexicon has been well developed.

However, there is considerable disagreement over how the mental lexicon is constructed in human mind. In particular, it is not clear how different the mental lexicons between first language (L1) and second language (L2) are. Refer to, for instance, (Harley, 2010) for the mental lexicon for bi- or multi-lingual. Theoretically it is discussed that the link occurs between two words when they have some semantic relationship, and that co-ordination or collocation is stronger than others, superordination or synonymy, for instance, (Aitchison, 1987, 2012); this seems to be agreed for L1. However, no one may know whether this holds or not for L2; the co-ordination or collocation may be stronger in the semantic relation, or, the syntactical or phonological relation may be stronger than the semantic one, or, the relation in the wide meaning of “co-occurrence” may be stronger than others. Then, to examine how humans link among words actually and the mental lexicon for L1 and/or L2, we believe we should pursue psycholinguistic experiments. In the word association task of Wilks and Meara (Wilks & Meara, 2002), participants were 30 learners of French whose L1 was English and 30 native speakers of French and were asked to identify associations from a given set of randomly chosen five French words. These experiments bring the number of “hits”, that is, pairs of associated words, for each participant, then we

know, as a result, native speakers than nonnative ones perceive significantly more associations. This simple experiment gives to us slight information on how the L2 mental lexicon is organized; that is, “densities” in word associations are different between L1 and L2 since the network of word associations is considered to be denser as the number of word associations increases. Thus main conclusion in (Wilks & Meara, 2002) is that L1 networks are denser than L2 networks of the learners in French. We should remark the linkage “association” in (Wilks & Meara, 2002) depends on the individual impression of each participant, that is, we cannot know what type of reasons generates their links. In other words, it is natural to investigate the L2 mental lexicon without any restriction in a way to associate.

Following the work stated above, Takahashi and Tanaka (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) try to investigate how the L2 lexical network is organized in English, where some mathematical methods essentially help them to state conclusions but explicit expressions and details are hidden since their discussion should have been done mainly in the context of (psycho)linguistics. Therefore one of the main objects of this paper is to give a supplementary explanation on mathematical ideas used in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) explicitly. To investigate the structure of the mental lexicon as a network, we need many psychological experiments for a large number of persons and many tools guaranteed theoretically for analysing properties of the model. As a first step, our methods with mathematically rigorous discussion in the sense of random graph, which are quite basic and useful, will be more developed in the future research.

This paper is organized as follows: In Section 2, we briefly review the study of Wilks & Meara (2002) (Section 2.1) and explain their methods in terms of graph theory and random graphs (Subsection 2.2). We discuss the main study in this paper in Section 3: firstly the brief review of the studies of Takahashi and Tanaka (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) is given in Subsection 3.1. Then we explain some latent mathematical ideas for (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) in Subsections 3.3. After that, we propose and apply our methods to the data of experiments in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) in Subsection 3.4. Finally we discuss concluding remarks about our methods and mental lexicon networks of L2 learners’ in Section 4. Some supplements in the text (some mathematical proofs, some sample of experiments in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019), mathematical interpretations and so on) are exhibited in the **Appendix**.

2. Background

2.1. The Study of Wilks and Meara

In this paper, our methods are almost the same as in (Wilks & Meara, 2002). In a given set of words, one construct a kind of network of words individually; such

a network is so-called a graph. Here the set of vertices (in the terms of Graph Theory) corresponds to that of words and the set of edges does to that of “associations”, where association implies the existence of strongly link of two words. The density of a given graph is defined as the ratio of the number of all the edges in the number of all possible edges which equals to $n(n-1)/2$, where n is the number of all the words. Wilks & Meara (2002) investigated a difference between the densities of L1 and L2 in French. Their experiment was designed as 40 questions (items) in one questionnaire and each question is: for 5 words chosen randomly from the first 1000 most frequency words,

1) if any pair of 5 words are thought to be no association, nothing should be written;

2) if one or more pairs of 5 words are thought to be associated, the only one pair of strongest association should be circled.

We say the “hit rate” for the ratio of the number of answers to be associated over that of all questions.

Through this “five-word task” experiment, we should remark that we cannot obtain the structure of associations of the individual mental lexicon network since, under the instruction (2), we can only see the strongest link pair of two words; so to say, we asked the participant to break the “detail structure” of associations. However, it is important to see the ratio of items with no-association. Once the network of mental lexicon was constructed, the density p behaves like the existence probability of association between any pair of words in this network. Under the assumption that such a probability p is identically and independent distributed, the probability of an item with no-association occurring is estimated as $(1-p)^{10}$ since the number of all possible pairs is 10. Therefore, by observing the ratio of items with no-association, we can get the density p . We are very careful to treat items with association, which was forced to be indicated one-pair only.

Of course, we know such instruction (2) may be suitable for actual experiment with psychological reasons for participants. If such a experiment was done for subjects not having human emotions, for example, in computer simulations, the restriction in (2) would not be needed. Namely, instead of (2), if the condition that all pairs of association might be indicated was asked, it would be somewhat difficult for participants to choose “no-association” in the item just after indicating “many associations”.

Repeatedly we should say it is important to obtain the number of items with no-association as accurately as possible.

2.2. Random Graphs

2.2.1. Interpretation of the Setting of Wilks and Meara

Among huge words in one’s mind, it is natural to consider a kind of network of words, mental lexicon network, is constructed, and some associations appear and other ones disappear every time one sees or hears words. Namely, such a network will be drastically changing its structure depending on the input of words. However, once the context generating words is confirmed, it is consi-

dered to be almost stable with small fluctuation. On the other hand, how does the network of “mental lexicon” behave every time words are given randomly without any context? We believe the network appearing in such a situation may be a “basic” or “core in (Wilks & Meara, 2002)” mental lexicon; although it will change its structure at every input of words, its density is considered to be stable in the same way as in (Wilks & Meara, 2002), where they actually examined to compare the relative densities of first language (L1) in French and second language (L2) in such “basic” lexical networks. Here the word “density” can be treated as the proportion of the number of the associated word pairs over the number of all possible pairs in the whole given words. In addition, its “density” can correspond to a probability p ; if a pair of words is chosen randomly, they will associate each other with probability p . The model of network with probability is usually said to be a random graph in mathematics. In particular, the Erdős-Rényi model is well-known and much research on it have been actively studied; the first remarkable paper about this model is appeared in (Erdős & Rényi, 1959). Once the density, or probability, is given, we can generate a sample network of mental lexicon in the context of random graphs. The set of words and that of associations are the set of vertices and edges in the term of graph theory, respectively. Then $G(n, p)$ denotes a random “undirected” graph such that every possible edge occurs independently with probability $0 < p < 1$ for n vertices. See also **Figure 1**.

In Wilks and Meara (Wilks & Meara, 2002), they derived the density by many computer simulations to generate regular random graphs and observe the existence of edges in randomly selected subgraph of 5 vertices. However, we emphasize the relationship between a hit rate and its density can be obtained analytically, whose concrete form will be seen in Proposition 2.1 in Section 2.2.2. Namely, once we obtained the hit rate, we can derive its density immediately. In this sense, we calculated the densities in the paper of Wilks and Meara, then their evaluations were about half of our ones. We imagine there was a little bit of mistake or ambiguity in generating random “regular” graphs in their paper. They introduced all notions in graph theory in terms of “undirected” graph: for instance, symmetric adjacency matrices representing the structure of undirected graphs, the density represented by the ratio of the number of all undirected edges in the complete graph and so on. So it is natural to recognize their results in the setting “undirected” graphs. In fact, we re-examined their strategy in using a statistical programming language R (R Core Team, 2019) and a graph theoretical package iGraph (Csardi & Nepusz, 2006), then our estimation seems to be right. However the essence and importance of their paper cannot suffer from tiny errors;

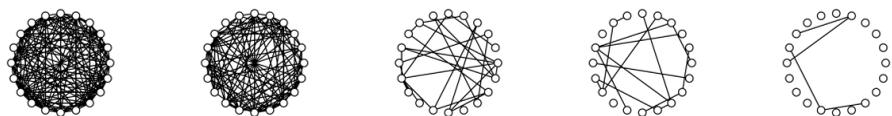


Figure 1. Samples of random graphs of 20 vertices, $G(20, p)$: Edge probabilities p are set as 0.7, 0.5, 0.15, 0.08 and 0.03, respectively.

as will be seen in Remark 2.4, if we apply “directed” random graphs to the association task for 5 words chosen randomly in (Wilks & Meara, 2002), we can find the estimations in our way coincide with the results in their simulation.

2.2.2. Basic Relations around Densities

As stated in the above, we first give a simple relation between the hit rate in the five-word task and the density.

Proposition 2.1. *We can easily find, for the five-word task,*

$$\text{“hit rate”} = 1 - (1 - \text{“density”})^{10}. \quad (1)$$

Thus, once the hit rate is given from actual experiments, we soon obtain the density as

$$\text{“density”} = 1 - (1 - \text{“hit rate”})^{\frac{1}{10}}. \quad (2)$$

Here the exponent “10” of the right hand side of (1) comes from $\binom{5}{2}$ in “five-word task,” which is the number of all possible pairs on one question consisting of 5 words chosen randomly. Thus, if we use a questionnaire such that each question consists of “ n -words” ($n \in \mathbf{N}$) chosen randomly, that is, “ n -word task,” each exponent of the right hand sides of (1) and (2) is replaced with $\binom{n}{2}$ and $1/\binom{n}{2}$, respectively.

Before stating the next relation, we give a definition in terms of graph theory. The set of all words correspond to the set of vertices $V(G)$ of a given graph G . Similarly the set of all associations among words corresponds to the set of edges $E(G)$ and the density does to the ratio of the number of edges of G in that of the complete graph with size of $V(G)$.

Definition 2.2 Let $G = (V(G), E(G))$ be a finite simple undirected graph and let A be a non-empty proper subset of $V(G)$ and $B = V(G) \setminus A$. Then we define four types of densities p_{A+B}, p_{AA}, p_{BB} and p_{AB} as follows:

$$p_{A+B} = \frac{\#E(G)}{\binom{n_A + n_B}{2}}, p_{AA} = \frac{\#E(\langle A \rangle)}{\binom{n_A}{2}}, p_{BB} = \frac{\#E(\langle B \rangle)}{\binom{n_B}{2}} \text{ and } p_{AB} = \frac{\#E(A, B)}{n_A n_B}, \quad (3)$$

where n_U is the number of elements of the set U , $\langle U \rangle$ is the subgraph of G induced by the set of vertices U and $E(A, B)$ is the set of edges connecting two set of vertices A and B .

We find the following relation is trivial from Definition 2.2 but important in this paper. It may be said to be a kind of mean field theoretic approximation; the left hand side of the following equation implies a “one-body system” $A + B$ and the right hand a “two-body system” A and B .

Proposition 2.3 *We have the relation among four densities as*

$$\binom{n_A + n_B}{2} \cdot p_{A+B} = \binom{n_A}{2} \cdot p_{AA} + n_A \cdot n_B \cdot p_{AB} + \binom{n_B}{2} \cdot p_{BB}. \quad (4)$$

Let us express the symbols shown in the above in the context of the word association, which will be discussed in Section 3: n_A and n_B are the numbers of words in Group A and Group B, respectively; thus $n_A + n_B$ indicates the numbers of the whole set of all words, Group A + B. Moreover, p_{AA} , p_{BB} , p_{AB} and p_{A+B} imply the densities connecting two words in Group A, in Group B, one word in Group A and the other in Group B, and in Group A + B, respectively. In actual experiments, we may say (4) becomes a kind of assumption for derived/underived “small” densities satisfying in the sense of mean field theoretic approximation. In particular, we discuss in Section 3 the relationship among densities in our “five-word task” experiments under the relation (4) for simplicity. For more details about the relation (4) in the “ k -word task” experiments, refer to Section A3 in **Appendix**.

Remark 2.4. Throughout this paper, we treat “undirected” random graphs only. As stated in the previous subsection, “undirected” random graphs seemed to be treated in (Wilks & Meara, 2002). However, by seeing the differences between our estimations and their results and referring to the sequential works of Meara et al. (Wilks, Meara, & Wolter, 2005; Meara, 2007), we confirm they treated “directed” graphs, or, “digraphs” in their simulations. Now let us show the relationship between a hit rate and a density in the case of “digraphs” in the similar way to that in the above. Let $G = (V(G), A(G))$ be a di-graph, where $A(G)$ is the set of all directed edges, or, arcs: $[x, y] \in A(G)$ implies $[x, y]$ is an arc from x to y . In this setting, it is natural that x and y in $V(G)$ are considered to be “associated” if and only if $[x, y] \in A(G)$ or $[y, x] \in A(G)$. We set the density of a digraph $G = (V(G), A(G))$ as

$$p = \frac{\#A(G)}{\#V(G)(\#V(G)-1)}. \quad (5)$$

Under the assumption that such a probability p is identically and independently distributed, the probability of no-association occurring between x and y is estimated as $(1-p)^2$. Moreover we can easily find, for a randomly selected n -words experiment, that is, “ n -word task,”

$$\text{“hit rate”} = 1 - (1-p)^2 \binom{n}{2}. \quad (6)$$

Thus, once the hit rate is given from actual experiments, we soon obtain the density p as

$$p = 1 - (1 - \text{“hit rate”})^{\frac{1}{n(n-1)}}. \quad (7)$$

When $n = 5$, we can see the formula (6) almost recovers the results of simulations in Table 1 and Figure 3 in (Wilks & Meara, 2002). In their sequential works (Wilks, Meara, & Wolter, 2005; Meara, 2007), they treat digraphs without ambiguity and introduce a modified model considering associations linked “indirectly” in some way. We give mathematical interpretations on some of their models and results in simulations in Section A6 in **Appendix**. Repeatedly we remark we treat “undirected” random graphs only in this paper.

3. Main Study

3.1. The Study of Takahashi and Tanaka

The studies in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) report on a word association task for Japanese learners who learn English as an L2. The English words used in our word association task were chosen from The New JACET List of 8000 Basic Words (JACET, 2016), which is a list of essential English vocabulary for university students in Japan. In total, we chose 1090 words essential for junior high school students and 1744 words essential for senior high school students; hereinafter the former set of words is called “Group A”, the latter “Group B”, and the whole set of words “Group A + B.” The format of our word association task was similar to the one performed in Wilks and Meara (Wilks & Meara, 2002). However, we performed it using English words instead of French words. In the experiment, we presented a set of five words in each trial. See also **Figure 2**. There were 80 word sets in a word association test as follows: in the word set 1 - 40, we selected 5 words randomly from Group A + B; in the word set 41 - 80, we chose 5 words only from Group B; the two types of words sets appeared in a random order in each questionnaire. Participants were asked to identify a single association, if any, in a given set of randomly chosen five words in the word association tests that we devised. Then we measured the “hit rate”, the ratio of word associations. An example and the flow of our actual experiment can be found in Section A1.

In (Takahashi & Tanaka, in preparation), the participants were five Japanese undergraduate students whose proficiency level of English was intermediate (their TOEIC Listening and Reading score was approximately 600), and five Japanese researchers who used English on daily basis in their research. The data of the students showed that the mean hit rate was higher when the words from Group A were included, suggesting that the words in Group A might be functioning as hub vertices that help link the words in the lexical network in their L2 mental lexicon. See also **Figure 3**. This tendency was not observed in the test results of the researchers. See also **Table 1**.

In Tanaka & Takahashi (2019), the participants were thirty-two native

Table 1. Result in (Takahashi & Tanaka, in preparation).

		Students (# = 5)	Adults (# = 5)
Group A + B (2834 words)	Mean hit rate (# of answered pairs/40 × 5)	0.705 (141/200)	0.780 (156/200)
	Estimated density p_{A+B} [# of each type in answered pairs]	0.1149 AA = 30, AB = 67, BB = 44	0.1405 AA = 32, AB = 66, BB = 58
Group B (1744 words)	Mean hit rate (# of answered pairs/40 × 5)	0.550 (110/200)	0.800 (160/200)
	Estimated density p_{BB}	0.0767	0.1487

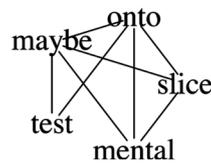


Figure 2. An example of associations in given five words.

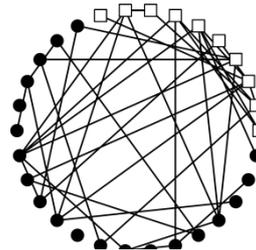


Figure 3. Sample of another kind of random graph: Group A (white-squared 10 vertices) and Group B (black-circled 20 vertices); Edge probabilities in two vertices within A, within B, and between A and B, are set as 0.15, 0.08, and 0.11, respectively.

Table 2. Result in (Tanaka & Takahashi, 2019).

		Students (# = 32)
Group A + B (2834 words)	Mean hit rate (# of answered pairs/40 × 32)	0.6805 (871/1280)
	Estimated density p_{A+B} [# of each type in 871 answered pairs]	0.1078 AA = 162, AB = 427, BB = 282
Group B (1744 words)	Mean hit rate (# of answered pairs/40 × 32)	0.5961 (763/1280)
	Estimated density p_{BB}	0.0867

speakers of Japanese whose proficiency level of English was also intermediate and their TOEIC Listening and Reading score was approximately 400 in average. We found a significant difference between the mean hit rate in Group A + B (68%) and that in Group B (60%, $p < 0.001$). This suggests that the ratios of how one word is associated with another are different in Group A and Group B. See also **Table 2**. In addition, we found that estimated densities for word associations were 0.11 for Group A + B, whereas 0.09 for Group B, which might indicate that the word associations involving Group A were denser than those in Group B.

For interpretation and analysis in terms of (psycho)linguistics for these results, refer to (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019). We should remark that the young participants selected for our experiments in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) are intermediate-level learners of English in a Japanese university. We believe participants must not be too high-grade nor low-grade in investigating the standard structure of the L2 mental lexicon for the majority. In this sense, our participants might be quite suitable.

3.2. Taste of Scale-Free Network

The concept of complex networks, or, scale-free network has relied on the construction of network evolution models. Most of studies on complex network, which are based on (Barabási, 2009; Barabási & Albert, 1999), discuss the topological statistics of the evolving network.

In our studies, we cannot evaluate such properties since the number of vertices, or equivalently, words, is very large but fixed. However we imagine the mental lexicon network of L2 learner may evolve like a complex network evolution model as they acquire new words more and more. In this sense, we can say that we could see some snapshot in their evolution and some trend as complex network. Thus, through the actual experiment, we would find the existence of “hubs”, which correspond to “important words” in the process of acquiring words.

Let us assume that Group A is the set of “more important” words in the whole set of words Group A + B. If this assumption is true, we can find the significant differences between the hit rates of Group A + B and Group B; moreover the density involving Group A is denser than that involving Group B. In other words, the significant differences can be said to be in favour of complex network in the process of L2 learners’ mental lexicon evolution model. However, we cannot state that the mental lexicon forms the scale-free network due to lack of showing the scale-free properties (Barabási, 2009; Barabási & Albert, 1999; Schur, 2007; Vitevitch, Goldstein, Siew, & Castro, 2014), for example, a power law distribution of degree of vertices. At this stage, all we can say is that the L2 mental lexicon network has a possibility to be a kind of scale-free network. Here let us recall our main purpose is to show that the whole set of words can be divided into more and less important groups and to propose some modelling for our experiment.

3.3. Latent Mathematical Ideas

As is seen in Section 3.1, the data of the students shows that the mean hit rate in Group B is lower than that in the whole set of words Group A + B. Thus the density of word associations in Group A is expected to be higher and affect that of Group A + B. However, by this result only, it is not sufficient to state that “the set of basic words”, Group A, plays an important role in the whole set and act like “hubs”.

A natural question which may arise is why they did two types of experiments, Group B only and Group A + B, and did not another type, Group A only. There exist two types of answers: one is that the total number of questions becomes over 100 and seems to be too much to keep the qualities in participants’ answering (Dörnyei & Taguchi, 2010; Gillham, 2008). Another is that, by using only the data we obtained so far, we can derive some qualitative expressions exhibiting words on Group A are more important. These expressions can be seen in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) without de-

tails. Here let us state their explicit expressions; their proofs are given in Section A4. First, let us introduce the “densities of association” including at least one word in Group A and at least one word in Group B denoted by p_{A^*} and p_{B^*} , respectively.

Definition 3.1. We define two probabilities p_{A^*} and p_{B^*} as follows:

$$p_{A^*} = \frac{\binom{n_A}{2} \cdot p_{AA} + n_A \cdot n_B \cdot p_{AB}}{\binom{n_A}{2} + n_A \cdot n_B} \quad \text{and} \quad p_{B^*} = \frac{\binom{n_B}{2} \cdot p_{BB} + n_A \cdot n_B \cdot p_{AB}}{\binom{n_B}{2} + n_A \cdot n_B} \quad (8)$$

Remark 3.2. As are seen in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) or in Table 1 and Table 2, the densities we can obtain are just p_{A+B} and p_{BB} . Thus we cannot evaluate p_{B^*} . On the other hand, thanks to (4) in Proposition 2.3, which is, so to say, the general assumption, we can fortunately evaluate p_{A^*} via n_A , n_B , p_{A+B} and p_{BB} as follows:

$$p_{A^*} = \frac{\binom{n_A + n_B}{2} \cdot p_{A+B} - \binom{n_B}{2} \cdot p_{BB}}{\binom{n_A}{2} + n_A \cdot n_B}. \quad (9)$$

Proposition 3.3. Assume that $n_A < n_B$ and that p_{A+B} and p_{BB} are given. Then, if $p_{AA} > R_{A,B}$, we have

$$p_{A^*} > p_{B^*}, \quad (10)$$

where

$$R_{A,B} = R(n_A, n_B, p_{A+B}, p_{BB}) = \frac{p_{BB} \binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right) - p_{A+B} \binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B \right)}. \quad (11)$$

In particular, if

$$\frac{p_{A+B}}{p_{BB}} > \frac{\binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)}, \quad (12)$$

then $p_{A^*} > p_{B^*}$ always holds. Here, for any fixed $n_A > 0$, the right hand side of (12) monotonically decreases as $n_B (> n_A)$ increases; moreover it tends to 1 as $n_B \rightarrow \infty$.

Next let us introduce the average numbers of associations of each word in Group A and Group B denoted by d_A and d_B , respectively.

Definition 3.4. We define two expected degrees d_A and d_B as follows:

$$d_A = (n_A - 1) \cdot p_{AA} + n_B \cdot p_{AB} \quad \text{and} \quad d_B = (n_B - 1) \cdot p_{BB} + n_A \cdot p_{AB} \quad (13)$$

Proposition 3.5. Assume that $n_A < n_B$ and that p_{A+B} and p_{BB} are given. Then, if $p_{AA} > S_{A,B}$, we have

$$d_A > d_B, \quad (14)$$

where

$$S_{A,B} = S(n_A, n_B, p_{A+B}, p_{BB}) = p_{A+B} - \frac{n_B(n_B - 1)}{n_A(n_A - 1)}(p_{A+B} - p_{BB}). \quad (15)$$

In particular, if

$$\frac{p_{A+B}}{p_{BB}} > \frac{n_B(n_B - 1)}{n_B(n_B - 1) - n_A(n_A - 1)}, \quad (16)$$

then $d_A > d_B$ always holds. Here, for any fixed $n_A > 0$, the right hand side of (16) monotonically decreases as $n_B (> n_A)$ increases, moreover it tends to 1 as $n_B \rightarrow \infty$.

Under the assumption $p_{A+B} > p_{BB}$, we should remark that the restriction for p_{AA} becomes weaker for the conclusions as n_B is bigger than any fixed n_A in both Propositions 3.3 and 3.5. Furthermore, the next corollary shows the condition in Proposition 3.5 is stronger than that in Proposition 3.3, that is, $p_{A*} > p_{B*}$ always holds if $d_A > d_B$ but the converse does not.

Corollary 3.6. Assume that $n_A < n_B$. Then $R_{A,B} < S_{A,B}$ if and only if $p_{A+B} > p_{BB}$.

Let us apply these estimates to results in experiments in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019). Here we use a statistical programming language R (R Core Team, 2019) for numerical calculations throughout this paper.

3.3.1. In Preliminary Study of Takahashi and Tanaka

In the preliminary study of Takahashi and Tanaka (Takahashi & Tanaka, in preparation), for students, there was a significant difference between mean hit rate in Group A + B and that in Group B via a statistical test (Welch t-test): $t = 3.2399$ and $p\text{-value} = 0.001297 < 0.01$. See also **Table 1**. For students in this case, we will show $p_{A*} > p_{B*}$ and $d_A > d_B$; in other words, Group A is more important for students.

Let us calculate $R_{A,B}$ in Proposition 3.3 and $S_{A,B}$ in Proposition 3.5 for students. Here $n_A = 1090$, $n_B = 1744$, $p_{A+B} = 0.1149$ and $p_{BB} = 0.0767$. Thus we have, by (11) and (15),

$$R_{A,B} = -0.01925604 \quad \text{and} \quad S_{A,B} = 0.01707433 \quad (17)$$

By Proposition 3.3, we immediately have

$$p_{A*} = 0.1381755 > p_{B*},$$

where we apply (9) to obtaining the value of p_{A*} . Also we should remark that

$$\frac{p_{A+B}}{p_{BB}} = 1.498044 \quad \text{and} \quad \frac{\binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)} = 1.398099,$$

which satisfy (12) in Proposition 3.4.

We recall a kind of exact value of p_{AA} or p_{AB} cannot be obtained. However, by using (34), (36) in **Appendix** and the information on “the number of each type in answered pairs” in **Table 1**, we can estimate p_{AA} and p_{AB} below:

Claim 3.7. We have

$$p_{AA} \times \text{“total number of all possible associations per question (= 10)”} \\ \times \text{“total number of questions (= 40)”} \times \text{“total number of participants”} \quad (18)$$

$$\times \frac{\binom{n_A}{2}}{\binom{n_A + n_B}{2}} \geq \text{AA},$$

$$p_{AB} \times \text{“total number of all possible associations per question (= 10)”} \\ \times \text{“total number of questions (= 40)”} \times \text{“total number of participants”} \quad (19)$$

$$\times \frac{n_A n_B}{\binom{n_A + n_B}{2}} \geq \text{AB},$$

$$p_{BB} \times \text{“total number of all possible associations per question (= 10)”} \\ \times \text{“total number of questions (= 40)”} \times \text{“total number of participants”} \quad (20)$$

$$\times \frac{\binom{n_B}{2}}{\binom{n_A + n_B}{2}} \geq \text{BB},$$

where AA, AB and BB are the numbers of answered pairs in the experiment connecting both two words in Group A, between one word in Group A and the other in Group B, and both two words in Group B, respectively.

Depending on the requirement (2) in Subsection 2.1 for experiments, “hidden” associations, which may exist, cannot be counted. Thus we only have such inequalities stated in the above.

Applying “total number of participants” = 5, AA = 30, AB = 67 and BB = 44 in **Table 1** to Claim 3.7, we can obtain

$$p_{AA} \geq 0.1014573 \quad \text{and} \quad p_{AB} \geq 0.07074378 \quad (21)$$

Here we can also get $p_{BB} \geq 0.05810657$, which does not contradict the value $p_{BB} = 0.0767$ obtained by the actual experiment. By (21) and (17) with Propositions 3.5, we can obtain

$$d_A > d_B.$$

As a further observation, combining Proposition 2.3 and (21), we have the upper bound estimates for p_{AA} and p_{BB} . Actually we can see

$$\begin{aligned}
 p_{AA} &= \frac{p_{A+B} \binom{n_A + n_B}{2} - p_{BB} \binom{n_B}{2} - p_{AB} n_A n_B}{\binom{n_A}{2}} \\
 &\leq \frac{0.1149 \cdot \binom{n_A + n_B}{2} - 0.0767 \cdot \binom{n_B}{2} - 0.07074378 \cdot n_A n_B}{\binom{n_A}{2}} \quad (22) \\
 &= 0.3541553.
 \end{aligned}$$

Similarly, we have

$$\begin{aligned}
 p_{AB} &= \frac{p_{A+B} \binom{n_A + n_B}{2} - p_{BB} \binom{n_B}{2} - p_{AA} \binom{n_A}{2}}{n_A n_B} \\
 &\leq \frac{0.1149 \cdot \binom{n_A + n_B}{2} - 0.0767 \cdot \binom{n_B}{2} - 0.1014573 \cdot \binom{n_A}{2}}{n_A n_B} \quad (23) \\
 &= 0.1496395.
 \end{aligned}$$

Consequently, by the experiment in (Takahashi & Tanaka, in preparation), we have

$$p_{A+B} = 0.1149, \quad p_{BB} = 0.0767, \quad 0.1014573 \leq p_{AA} \leq 0.3541553$$

and $0.07074378 \leq p_{AB} \leq 0.1496395$

for students.

On the other hand, for adults, no such significant difference can be found between mean hit rate in Group A + B and that in Group B: $t = -0.48995$ and p -value = 0.6244. Refer to **Table 1** and (Takahashi & Tanaka, in preparation). Thus we cannot expect $p_{A^*} > p_{B^*}$ or $d_A > d_B$ holds. Let us discuss the densities for adults by the same arguments as in the above. Here $n_A = 1090$, $n_B = 1744$, $p_{A+B} = 0.1405$ and $p_{BB} = 0.1487$, Thus we have, by (11) and (15),

$$R_{A,B} = 0.1692979 \quad \text{and} \quad S_{A,B} = 0.1614992. \quad (24)$$

Furthermore we have

$$p_{A^*} = 0.1355037.$$

Next Applying “total number of participants” = 5, AA = 32, AB = 66 and BB = 58 in **Table 1** to Claim 3.3.1, we can obtain

$$p_{AA} \geq 0.1082211 \quad \text{and} \quad p_{AB} \geq 0.0696879 \quad (25)$$

Here we can also get $p_{BB} \geq 0.07659502$, which does not contradict the value $p_{BB} = 0.1487$ obtained by the actual experiment. Obviously, by (25) and (24) as far, we cannot conclude whether

$$d_A > d_B \quad \text{or} \quad p_{A^*} > p_{B^*}$$

holds; in fact, our intuition in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) is that $d_A = d_B$ and $p_{A^*} = p_{B^*}$ hold approximately, which does not contradict the discussion above.

Again, by the same argument as in the case for students with the results for adults in (Takahashi & Tanaka, in preparation), we consequently have

$$p_{A+B} = 0.1405, p_{BB} = 0.1487, 0.1082211 \leq p_{AA} \leq 0.3463076$$

and $0.0696879 \leq p_{AB} \leq 0.1440217$ for adults.

3.3.2. In Further Study of Tanaka and Takahashi

In a further study of Tanaka and Takahashi (Tanaka & Takahashi, 2019), for 32-students, there was also a significant difference between mean hit rate in Group A + B and that in Group B via classical Welch t-test: $t = 4.4578$ and $p\text{-value} = 8.637e-06 < 0.00001$. Also in this case, we will show $p_{A^*} > p_{B^*}$ and $d_A > d_B$ by the same arguments as in the previous subsection 3.3.1.

Let us calculate $R_{A,B}$ in Proposition 3.3 and $S_{A,B}$ in Proposition 3.5. Referring to Table 2, we find $n_A = 1090$, $n_B = 1744$, $p_{A+B} = 0.1078$ and $p_{BB} = 0.0867$. Thus we have, by (11) and (15),

$$R_{A,B} = 0.0336981 \text{ and } S_{A,B} = 0.0537654. \tag{26}$$

In this case, we see $p_{A^*} = 0.1206564$ and

$$\frac{p_{A+B}}{p_{BB}} = 1.243368 \text{ and } \frac{\binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)} = 1.398099,$$

which does not satisfy (12) in Proposition 3.4.

Now applying “total number of participants” = 32, AA = 162, AB = 427 and BB = 282 in Table 2 to Claim 3.7, we can obtain

$$p_{AA} \geq 0.0856046 \text{ and } p_{AB} \geq 0.07044681 \tag{27}$$

here we can also get $p_{BB} \geq 0.05818911$, which does not contradict the value $p_{BB} = 0.0867$ obtained by the actual experiment. By (27) and (26) with Propositions 3.3 and 3.5, we can obtain

$$p_{A^*} > p_{B^*} \text{ and } d_A > d_B.$$

Again, by the same argument as in the previous subsection, we consequently have

$$p_{A+B} = 0.1078, p_{BB} = 0.0867, 0.0856046 \leq p_{AA} \leq 0.2814746$$

and $0.07044681 \leq p_{AB} \leq 0.1316$ for this experiment.

3.4. Modelling towards Actual Experiments

Assumption 3.8. (The Lorentz-Berthelot rules) We assume the relation among three densities as

$$p_{AB} = \sqrt{p_{AA} \cdot p_{BB}}, \tag{28}$$

where p_{AA} , p_{BB} and p_{AB} implies the density (probability) connecting two words in Group A, Group B and between Group A and B, respectively.

Let us give a brief explanation of Assumption 3.8 stated above. The relation (28) corresponds to the Lorentz-Berthelot rules or the Berthelot rule, which is well known as one of basic combining rules in computational chemistry and molecular dynamics. In our context, it is natural to consider every word in Group A and B corresponds to a particle of two categories of identical particles A and B, respectively. Therefore three densities p_{AA} , p_{BB} and p_{AB} can be also considered as the interaction force within A, B and between A and B, respectively. In this sense, it is natural to give the relation (28) among p_{AA} , p_{BB} and p_{AB} by the Berthelot rule in our setting. We should remark that we put $n_A = 1090$, $n_B = 1744$ in the experiments of Takahashi and Tanaka. In addition, once densities p_{A+B} and p_{BB} are obtained by actual experiments, we can calculate other densities p_{AA} and p_{AB} by the relations (28) and (4). The explicit expressions of p_{AA} and p_{BB} in terms of n_A , n_B , p_{A+B} and p_{BB} , which are somewhat complicated forms, can be seen in Subsection A5 in **Appendix**. Calculated results are shown concretely in **Table 3** and **Table 4** in Subsections 3.4.1 and 3.4.2, respectively.

As stated before Claim 3.7, we should take care in dealing associated pairs in the participants' answers. The requirement asked for all participants, which is suitable in the actual experiments discussed in Subsection 2.1, is that the participants must indicate at most one-pair with association per question. By virtue of this requirement, when two words are chosen as one pair in some item, we cannot judge whether it may be only one pair in this item, or it may be stronger one pair than other one or more pairs, or it may be one pair in a triangle or more complicated subgraph. In other words, we cannot know how strong the answered pair is, and whether other hidden pairs are within Group A, Group B, or, between Group A and Group B. Namely, this requirement might break the structure of associations which may exist potentially in answering. For these

Table 3. Results summary for (Takahashi & Tanaka, in preparation).

		Students	Adults
Results from experiments	cf. Table 1	AA = 30, AB = 67, BB = 44	AA = 32, AB = 66, BB = 58
	(normalized ratio)	(0.212766:0.475177:0.312057)	(0.205128:0.423077:0.371795)
1) Ratio without densities	$P_{AA} : P_{AB} : P_{BB}$ (normalized)	0.14785:0.47354:0.37861	
	Result of chi-square test	X-squared = 5.6692, df = 2, p-value = 0.05874	X-squared = 4.3197, df = 2, p-value = 0.1153
	calculated densities	$p_{AA} = 0.192023$, $p_{AB} = 0.12136$	$p_{AA} = 0.12786$, $p_{AB} = 0.13789$
2) Ratio via our modelling	$Q_{AA} : Q_{AB} : Q_{BB}$ (normalized)	0.24709:0.50017:0.25274	0.13455:0.46474:0.40071
	Result of chi-square test	X-squared = 2.8114, df = 2, p-value = 0.2452	X-squared = 6.6839, df = 2, p-value = 0.03537

Table 4. Results summary for (Tanaka & Takahashi, 2019).

Results from experiments	cf. Table 2	AA = 162, AB = 427, BB = 282
	(normalized ratio)	(0.185993:0.490241:0.323766)
1) Ratio without densities	$P_{AA} : P_{AB} : P_{BB}$ (normalized)	0.14785:0.47354:0.37861
	Result of chi-square test	X-squared = 16.004, df = 2, p-value = 0.0003348
	Calculated densities	$p_{AA} = 0.14634$, $p_{AB} = 0.11264$
2) Ratio via our modelling	$Q_{AA} : Q_{AB} : Q_{BB}$ (normalized)	0.20070:0.49479:0.30451
	Result of chi-square test	X-squared = 2.0358, df = 2, p-value = 0.3613

reasons, it is dangerous to investigate the details among all the answered “associations”, nevertheless they have fruitful information. Let us illustrate an example using the concrete density $p_{A+B} = 0.1149$ in the result for student in (Takahashi & Tanaka, in preparation); refer to also **Table 1**. Here we should remark the following holds:

$$\text{Expected number of associated pairs per question} = \binom{5}{2} \cdot \text{density}.$$

Then there is expected to be 45.96, derived from $\binom{5}{2} \times p_{A+B} \times 40$, associations for one participant in average, but the number of answered association, which are asked to be answered as one association per question, is 28.2, derived from $141/5$, per participant in average. The difference per participant between them, about 18-associations, are considered to be hidden due to the requirement in one person’s answering. On the other hand, the impression for almost all participants is that almost all answered association were unique and there were three questions at most having two or three associations (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019). The interpretation about this difference between the measured quantities and their impressions is a further problem for (psycho)linguistics.

Although there exist some problems like stated above, it is plausible to consider the ratio among three types of answered associations, AA:AB:BB, as a “criterion” for the ratio among three types of associations which might be appeared in minds. Moreover, each expected number of pairs per question derived from our experiments (cf. **Table 1** and **Table 2**) is at most 1.5 fortunately; this suggests the hidden association are somewhat small. In any case, the ratio of pairs within Group A, Group B and between Group A and Group B among all the answered associations should be treated very carefully.

Under such discussion above, let us compare the ratio AA:AB:BB with two types of ratios: one is the ratio $P_{AA} : P_{AB} : P_{BB}$, which does not contain any densi-

ties and another is the ratio $Q_{AA} : Q_{AB} : Q_{BB}$, which contains densities derived from our experiments and modelling. Their details are as follows.

1) The ratio $P_{AA} : P_{AB} : P_{BB}$ among the numbers of pairs of both in Group A, between Group A and Group B and in Group B becomes

$$P_{AA} : P_{AB} : P_{BB} = \binom{n_A}{2} : n_A n_B : \binom{n_B}{2} \quad (29)$$

by Equation (38) in **Appendix**. For details, refer to Subsection A2 in **Appendix**; here we write $P_{AA} : P_{AB} : P_{BB}$ for $P(E_{AA}) : P(E_{AB}) : P(E_{BB})$ in Subsection A2. In this case, we do not consider any densities. Since $n_A = 1090$ and $n_B = 1744$ in all experiments we treat, we have the normalized ratio $P_{AA} : P_{AB} : P_{BB}$ as

$$P_{AA} : P_{AB} : P_{BB} = 0.14785 : 0.47354 : 0.37861. \quad (30)$$

2) Considering the densities p_{AA} and p_{AB} , which are calculated via Claim 5.5, and p_{BB} derived from the experiments, we define the ratio $Q_{AA} : Q_{AB} : Q_{BB}$ as

$$Q_{AA} : Q_{AB} : Q_{BB} = \binom{n_A}{2} p_{AA} : n_A n_B p_{AB} : \binom{n_B}{2} p_{BB}. \quad (31)$$

If we assume that all densities would be equal, $p_{AA} = p_{AB} = p_{BB} = p_{A+B}$ in (31), then we have (29).

3.4.1. For Preliminary Study of Takahashi and Tanaka

From experiments (cf. **Table 1**), we have the total numbers of answered pairs connecting two words in Group A, between Group A and B, and in Group B are as follows:

Students: AA = 30, AB = 67, BB = 44; AA + AB + BB = 141 (total 200 items);

Adults: AA = 32, AB = 66, BB = 58; AA + AB + BB = 156 (total 200 items).

Also we recall that $p_{A+B} = 0.1149$ and $p_{BB} = 0.0767$ for students; $p_{A+B} = 0.1405$ and $p_{BB} = 0.1487$ for adults.

We can see a slight difference of the ratio AA:AB:BB of student for the ratio among P_{AA}, P_{AB} and P_{BB} (without their densities); in addition, we can see a slight difference of adults for the ratio among Q_{AA}, Q_{AB} and Q_{BB} (with their densities). Here we use a statistical programming language R (**R Core Team, 2019**) for numerical calculations and the chi-square test. See **Table 3**.

On the other hand, we can see no significant difference of the ratio AA:AB:BB of student for the ratio among Q_{AA}, Q_{AB} and Q_{BB} (with their densities) and that of adults for the ratio P_{AA}, P_{AB} and P_{BB} (without their densities).

Therefore we may say the density of adults would be uniform in all the words A and B and our densities derived in our modelling are fit on the data for students in the sense that there exists no significant difference from the ratio among AA, AB and BB.

3.4.2. For Further Study of Tanaka and Takahashi

From experiments (cf. **Table 2**), we have the total numbers of answered pairs connecting two words in Group A, between Group A and B, and in Group B are as follows:

$$AA = 162; AB = 427; BB = 282.$$

Also we recall that $p_{A+B} = 0.1078$ and $p_{BB} = 0.0867$.

We derived, from the chi-square test, a significant difference of the ratio AA:AB:BB for the ratio among P_{AA}, P_{AB} and P_{BB} in 1). On the other hand, we cannot derive a significant difference of the ratio AA:AB:BB for the ratio among Q_{AA}, Q_{AB} and Q_{BB} in 2). See **Table 4**.

Then, combining with the discussion in the previous subsection 3.4.1, we may say our densities for students in our modelling are fit on the data in the sense there exists no significant difference from the ratio among AA, AB and BB.

Consequently, we presume the Lorentz-Berthelot rules in (28) can be applied to the densities in the mental lexicon network of L2 “young” learners, which is considered to be in the process of evolution; for L2 “matured” learners, their mental lexicon network are considered to be almost completed. For details, these topics should be discussed in the context of linguistics or psychology.

4. Concluding Remarks

The main topics discussed in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019) are explorations of the structure of the mental lexicon of L2 learners’ in the context of (psycho)linguistics. Though a concept of random graph is introduced and applied essentially, only rough ideas derived mathematically are provided and details are omitted. Thus one of main topics in this paper is to give the explicit formulae and their proofs, which are seen in Section 3.3 and in Section A4, respectively. In addition, we propose a kind of modelling for actual experiments in (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019). We know that there exist many types of restriction in doing actual experiments although we always want to get sufficiently many kinds of data. In the sense that much information is derived by the actually obtained data so far, our proposal may be useful. For example, the “hit rate” in the n -word task brings us the “density” via a generalized form of Proposition 2.1 and Remark 2.4.

Here we discussed almost all topics in terms of mathematics, but many results should be interpreted and discussed in the context of (psycho)linguistics (cf. Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019), that is, we believe many fruitful and further studies are brought.

Acknowledgements

The authors wish to thank the anonymous reviewers for comments on earlier version of this paper. In addition, they wish to thank Professor Jenny Zhang, Managing Editor, and the staff members of OJML for their invaluable editing task. The authors also would like to express their sincere gratitude to Professors Hiroshi OGURA, Etsuo SEGAWA and Yusuke YOSHIKAWA for their helpful suggestions. YuH’s work was supported in part by Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research Grant Numbers (C)25400208, (A)15H02055 and (C)18K03401. MT acknowledges financial supports from the

Grant-in-Aid for Scientific Research (C) Japan Society for the Promotion of Science (Grant Number 17K02762).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Aitchison, J. (1987, 2012). *Words in the Mind: An Introduction to the Mental Lexicon* (1st ed. 1987, 4th ed. 2012). Hoboken, NJ: Wiley.
- Barabási, A. L. (2009). Scale-Free Networks: A Decade and Beyond. *Science*, 325, 412-413. <https://doi.org/10.1126/science.1173299>
- Barabási, A. L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286, 509-512. <https://doi.org/10.1126/science.286.5439.509>
- Boyd, R. (1993). Metaphor and Theory Change. In A. Ortony (Ed.), *Metaphor and Thought* (2nd ed.). Cambridge: Cambridge University Press.
- Csardi, G., & Nepusz, T. (2006). The Igraph Software Package for Complex Network Research. *InterJournal, Complex Systems*, 1695, 1-9. <http://igraph.org>
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. New York: Routledge. <https://doi.org/10.4324/9780203864739>
- Erdős, P., & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6, 290-297.
- Gaskell, G. (2009). *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Gillham, B. (2008). *Developing a Questionnaire* (2nd ed.). London: Continuum.
- Gruenenfelder, T. M., & Mueller, S. T. (2007). *Power Law Degree Distributions Can Fit Averages of Non-Power Law Distributions*. Research on Spoken Language Processing Progress Report No. 28, Bloomington, IN: Indiana University.
- Gruenenfelder, T. M., & Pisoni, D. B. (2005). *Modeling the Mental Lexicon as a Complex System: Some Preliminary Results Using Graph Theoretic Measures*. Research on Spoken Language Processing Progress Report No. 27, Bloomington, IN: Indiana University.
- Gruenenfelder, T. M., & Pisoni, D. B. (2009). The Lexical Restructuring Hypothesis and Graph Theoretic Analyses of Networks Based on Random Lexicons. *Journal of Speech, Language, and Hearing Research*, 52, 596-609. [https://doi.org/10.1044/1092-4388\(2009/08-0004\)](https://doi.org/10.1044/1092-4388(2009/08-0004))
- Harley, T. A. (2010). *Talking the Talk: Language, Psychology and Science*. Hove: Psychology Press. <https://doi.org/10.4324/9780203853030>
- JACET (2016). *The New JACET List of 8000 Basic Words*. Tokyo: Kirihara-Shoten.
- Kiss, G. (1968). Words, Associations and Networks. *Journal of Verbal Learning and Verbal Behaviour*, 7, 707-713. [https://doi.org/10.1016/S0022-5371\(68\)80130-6](https://doi.org/10.1016/S0022-5371(68)80130-6)
- Marcus, G. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/1187.001.0001>
- Meara, P. (1992). Network Structures and Vocabulary Acquisition in a Foreign Language. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 62-70). New York: Macmillan. https://doi.org/10.1007/978-1-349-12396-4_6

- Meara, P. (2007). Simulating Word Associations in an L2: Approaches to Lexical Organisation. *International Journal of English Studies*, 7, 1-20.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org>
- Rapaport, A., Livant, W. P., & Boyd, J. (1966). A Study of Lexical Graphs. *Foundations of Language*, 2, 338-376.
- Schur, E. (2007). Insights into the Structure of L1 and L2 Vocabulary Networks: Intimations of Small Worlds. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 182-204). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667268.013>
- Scott, J. (1991). *Social Network Analysis: A Handbook*. London: Sage.
- Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex Model of Mental Lexicon Reveals Explosive Learning in Humans. *Scientific Reports*, 8, Article No. 2259. <https://doi.org/10.1038/s41598-018-20730-5>
- Takahashi, R., & Tanaka, M. (in preparation). *An Exploration of Density in L2 Word Association Networks in the Mental Lexicon of Japanese Learners of English*.
- Tanaka, M., & Takahashi, R. (2019). *The Dynamic Characteristics in the L2 Mental Lexicon* (pp. 61-66). IEICE Tech. Rep. 119, No. 151, TL2019-23.
- Vitevitch, M. (2008). What Can Graph Theory Tell Us about Word Learning and Lexical Retrieval? *Journal of Speech, Language, and Hearing Research*, 51, 408-422. [https://doi.org/10.1044/1092-4388\(2008/030\)](https://doi.org/10.1044/1092-4388(2008/030))
- Vitevitch, M. S., Goldstein, R., Siew, C. S. Q., & Castro, N. (2014). Using Complex Networks to Understand the Mental Lexicon. In *Yearbook of the Poznan Linguistic Meeting* (Vol. 1, pp. 119-138). Berlin: De Gruyter Open. <https://doi.org/10.1515/yplm-2015-0007>
- Wilks, C., & Meara, P. (2002). Untangling Word Webs: Graph Theory and the Notion of Density in Second Language Word Association Networks. *Second Language Research*, 18, 303-324. <https://doi.org/10.1191/0267658302sr203oa>
- Wilks, C., & Meara, P. (2007). Implementing Graph Theory Approaches to the Exploration of Density and Structure in L1 and L2 Word Association Networks. In *Modelling and Assessing Vocabulary Knowledge* (pp. 167-181). Cambridge: Cambridge University Press.
- Wilks, C., Meara, P., & Wolter, B. (2005). A Further Note on Simulating Word Association Behaviour in a Second Language. *Second Language Research*, 21, 359-372. <https://doi.org/10.1191/0267658305sr251oa>

Appendix

A1. Flow of Actual Experiment-Five-Word Task

- 1) INPUT: Participant looks at 5 words per one question. See also **Figure A1**.
- 2) Example of possibilities in a participant's mind: total number of possibilities is $1024(=2^{10})$. See also **Figure A2**.
- 3) Required condition: Choose the at most one association (or two words) with the strongest link, even if two or more associations exist.
- 4) OUTPUT: Answer one association of two words or no association. See also **Figure A3**.
- 5) Go to the next question, that is, return to 1) in the next one.

A2. Counting and Probability

Let n_A and n_B be the numbers of words in Group A and Group B, respectively. The event that both two associated words belong to Group A in a given word set consisting of five words selected randomly from Group A + B is denoted by E_{AA} . Let us show the probability $P(E_{AA})$ can be given as

$$P(E_{AA}) = \frac{\binom{n_A}{2}}{\binom{n_A + n_B}{2}}$$

Let X be the random variable such that the number of words belong to Group A in a word set, that is, randomly selected five words. Then we have the probability $P(X = k)$ as

slice	onto	maybe	test	mental
-------	------	-------	------	--------

Figure A1. Example of a questionnaire in Step (1).

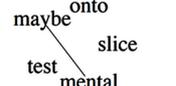
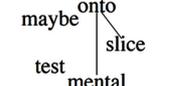
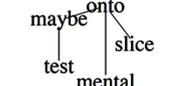
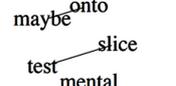
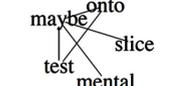
No association (only one pattern)	Association exists (1,023 patterns)		
<div style="display: flex; flex-direction: column; align-items: center;"> <div style="display: flex; justify-content: space-around; width: 100%;"> maybeonto</div> <div style="display: flex; justify-content: space-around; width: 100%;"> testslice</div> <div style="display: flex; justify-content: space-around; width: 100%;"> testmental</div> </div>			
			
			Many other types...

Figure A2. Example of Step (2): latent patterns of association.

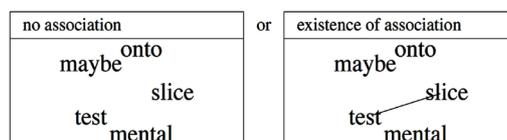


Figure A3. Example of Step (3): answer.

$$P(X = k) = \frac{\binom{n_A}{k} \binom{n_B}{5-k}}{\binom{n_A+n_B}{5}}.$$

Thus $P(E_{AA})$ can be expressed by

$$P(E_{AA}) = \sum_{k=2}^5 P(X = k) \frac{\binom{k}{2}}{\binom{5}{2}} = \sum_{k=2}^5 \frac{\binom{n_A}{k} \binom{n_B}{5-k} \binom{k}{2}}{\binom{n_A+n_B}{5} \binom{5}{2}}. \tag{32}$$

Here we remark the following relation: for positive integers $n \geq \ell \geq m$,

$$\binom{n}{\ell} \binom{\ell}{m} = \binom{n}{m} \binom{n-m}{\ell-m}. \tag{33}$$

Thus we can express the numerator and denominator in (32) as follows by putting $(n, \ell, m) = (n_A, k, 2)$ and $(n, \ell, m) = (n_A + n_B, 5, 2)$ in (33), respectively:

$$\binom{n_A}{2} \binom{n_A-2}{k-2} \binom{n_B}{5-k} \quad \text{and} \quad \binom{n_A+n_B}{2} \binom{n_A+n_B-2}{3}.$$

Then we obtain

$$\begin{aligned} P(E_{AA}) &= \sum_{k=2}^5 \frac{\binom{n_A}{2} \binom{n_A-2}{k-2} \binom{n_B}{5-k}}{\binom{n_A+n_B}{2} \binom{n_A+n_B-2}{3}} \\ &= \frac{\binom{n_A}{2}}{\binom{n_A+n_B}{2}} \sum_{\ell=0}^3 \frac{\binom{n_A-2}{\ell} \binom{n_B}{3-\ell}}{\binom{n_A+n_B-2}{3}} = \frac{\binom{n_A}{2}}{\binom{n_A+n_B}{2}}, \end{aligned} \tag{34}$$

where we use the total mass for the hypergeometric distribution is equal to 1:

$$\sum_{0 \leq k \leq n} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = 1. \tag{35}$$

The argument stated above similarly gives

$$P(E_{BB}) = \frac{\binom{n_B}{2}}{\binom{n_A+n_B}{2}} \quad \text{and} \quad P(E_{AB}) = \frac{n_A n_B}{\binom{n_A+n_B}{2}}, \tag{36}$$

where $P(E_{BB})$ and $P(E_{AB})$ are probabilities such that both two associated words belong to Group B and one word belongs to Group A and the other does to Group B in a given word set consisting of five words selected randomly from Group A + B, respectively.

Then the ratio of numbers of association pair that both two words are in Group A, one word is in Group A and the other in Group B or both two words are in Group B is estimated as

$$P(E_{AA}) : P(E_{AB}) : P(E_{BB}) = \frac{\binom{n_A}{2}}{\binom{n_A+n_B}{2}} : \frac{n_A n_B}{\binom{n_A+n_B}{2}} : \frac{\binom{n_B}{2}}{\binom{n_A+n_B}{2}} \quad (37)$$

$$= n_A(n_A - 1) : 2n_A n_B : n_B(n_B - 1).$$

In experiments in Takahashi and Tanaka (Takahashi & Tanaka, in preparation; Tanaka & Takahashi, 2019), we have approximately the normalized ratio

$$P(E_{AA}) : P(E_{AB}) : P(E_{BB}) = 0.14785 : 0.47354 : 0.37861, \quad (38)$$

since $n_A = 1090$ and $n_B = 1744$.

A3. Mean Field Theoretic Approximation

If we consider the mental lexicon network model in which a unique probability of association occurring p , then the relations (1) and (2) hold and its “density” coincides with p via the “hit rate” derived from word association tasks. However, in the setting where two or more kinds of densities, the hit rate from “ k -word task” ($k \geq 2$) is estimated in another expression. Let us show it in our model exhibited in Section 2.2.2. The “hit rate” can be expressed explicitly as follows:

$$1 - (1 - p_{A+B})^{\binom{k}{2}}$$

$$= \frac{1}{\binom{n_A+n_B}{k}} \sum_{\ell=0}^k \binom{n_A}{k-\ell} \binom{n_B}{\ell} \left(1 - (1 - p_{AA})^{\binom{k-\ell}{2}} (1 - p_{AB})^{\ell(k-\ell)} (1 - p_{BB})^{\binom{\ell}{2}} \right), \quad (39)$$

where we set $\binom{s}{2} = 0$ for $s < 2$. In the above, the probability p_{A+B} from (2) may be said to be the averaged “probability” in the whole set Group A + B, that is, two bodies system of Group A and Group B with p_{AA} , p_{AB} and p_{BB} is translated into one body system Group A + B with p_{A+B} . We should remark that p_{A+B} , which is expressed by p_{AA} , p_{AB} and p_{BB} , cannot coincide with the ratio of number of edges in G as is shown in (3), in general, and that the relation (39) coincides with the general assumption (4) when $k = 2$. Referring to (35) in Section A2, we obtain

$$\binom{n_A+n_B}{k} (1 - p_{A+B})^{\binom{k}{2}}$$

$$= \sum_{\ell=0}^k \binom{n_A}{k-\ell} \binom{n_B}{\ell} (1 - p_{AA})^{\binom{k-\ell}{2}} (1 - p_{AB})^{\ell(k-\ell)} (1 - p_{BB})^{\binom{\ell}{2}}, \quad (40)$$

where $(1 - p_{AA})^{\binom{k-\ell}{2}} (1 - p_{AB})^{\ell(k-\ell)} (1 - p_{BB})^{\binom{\ell}{2}}$ implies the probability of any association not occurring in “ k words” selected randomly, which consist of

“($k - \ell$) words” in Group A and “ ℓ words” in Group B. Let p be $\max(p_{AA}, p_{AB}, p_{BB}, p_{A+B})$ and p be sufficiently small. Then, the left hand side of (40) is

$$\binom{n_A + n_B}{k} - \binom{n_A + n_B}{k} \binom{k}{2} p_{A+B} + o(p) \tag{41}$$

and the right hand side of (40) is

$$\sum_{\ell=0}^k \binom{n_A}{k-\ell} \binom{n_B}{\ell} \left(1 - \binom{k-\ell}{2} p_{AA} - \ell(k-\ell) p_{AB} - \binom{\ell}{2} p_{BB} \right) + o(p), \tag{42}$$

where $o(\cdot)$ is the asymptotic little-o notation. Referring to (35) again in Section A2, we obtain

$$\begin{aligned} & \binom{n_A + n_B}{k} \binom{k}{2} p_{A+B} \\ &= \sum_{\ell=0}^k \binom{n_A}{k-\ell} \binom{n_B}{\ell} \left(\binom{k-\ell}{2} p_{AA} + \binom{\ell}{1} \binom{k-\ell}{1} p_{AB} + \binom{\ell}{2} p_{BB} \right) + o(p). \end{aligned} \tag{43}$$

Moreover, referring to (33) in Section A2, we can see that the left hand side of (43) is

$$\binom{n_A + n_B}{2} \binom{n_A + n_B - 2}{k-2} p_{A+B} \tag{44}$$

and the right hand side of (43) is

$$\begin{aligned} & \sum_{\ell=0}^k \left(\binom{n_A}{2} \binom{n_B}{\ell} \binom{n_A - 2}{k-2-\ell} p_{AA} + n_A n_B \binom{n_A - 1}{k-1-\ell} \binom{n_B - 1}{\ell-1} p_{AB} \right. \\ & \left. + \binom{n_B}{2} \binom{n_A}{k-\ell} \binom{n_B - 2}{\ell-2} p_{BB} \right) + o(p), \end{aligned} \tag{45}$$

where we set $\binom{s}{t} = 0$ if $t < 0$. Thanks to (35) again in Section A2, (45) can be simplified as

$$\binom{n_A + n_B - 2}{k-2} \left(\binom{n_A}{2} p_{AA} + n_A n_B p_{AB} + \binom{n_B}{2} p_{BB} \right) + o(p). \tag{46}$$

Consequently we have

$$\begin{aligned} & \binom{n_A + n_B}{2} p_{A+B} \\ &= \binom{n_A}{2} p_{AA} + n_A n_B p_{AB} + \binom{n_B}{2} p_{BB} + o(p), \end{aligned} \tag{47}$$

which implies the exact relation (40) in the “ k -word task” for $k \geq 3$ can be well approximated by the relation (4) if all densities are sufficiently small. This is a reason why the relation (4) is set as the plausible general assumption throughout

this paper. However, if some density in the networks is not small, the relation (40) may no longer be approximated by the relation (4), that is, the averaged probability p_{A+B} may be quite different from the “density” in the network, the ratio of number of edges in a graph G as is shown in (3). See an example in Section A6.2; see also (Meara, 2007).

A4. Proofs of Inequalities for Our Estimation

Proof of Proposition 3.3. Let us evaluate $p_{A^*} - p_{B^*}$. By Definition 3.1 and Proposition 2.3, we have

$$\begin{aligned}
 p_{A^*} - p_{B^*} &= \frac{\binom{n_A}{2} p_{AA} + n_A n_B p_{AB}}{\binom{n_A}{2} + n_A n_B} - \frac{\binom{n_B}{2} p_{BB} + n_A n_B p_{AB}}{\binom{n_B}{2} + n_A n_B} \\
 &= \frac{\left(\binom{n_B}{2} + n_A n_B\right) \left[\binom{n_A}{2} p_{AA} + n_A n_B p_{AB}\right] - \left(\binom{n_A}{2} + n_A n_B\right) \left[\binom{n_B}{2} p_{BB} + n_A n_B p_{AB}\right]}{\left(\binom{n_A}{2} + n_A n_B\right) \left(\binom{n_B}{2} + n_A n_B\right)} \\
 &= \frac{\left(\binom{n_B}{2} + n_A n_B\right) \left[\binom{n_A + n_B}{2} p_{A+B} - \binom{n_B}{2} p_{BB}\right] - \left(\binom{n_A}{2} + n_A n_B\right) \left[\binom{n_A + n_B}{2} p_{A+B} - \binom{n_A}{2} p_{AA}\right]}{\left(\binom{n_A}{2} + n_A n_B\right) \left(\binom{n_B}{2} + n_A n_B\right)} \\
 &= \frac{\left(\binom{n_B}{2} - \binom{n_A}{2}\right) \binom{n_A + n_B}{2} p_{A+B} - \left(\binom{n_B}{2} + n_A n_B\right) \binom{n_B}{2} p_{BB} + \left(\binom{n_A}{2} + n_A n_B\right) \binom{n_A}{2} p_{AA}}{\left(\binom{n_A}{2} + n_A n_B\right) \left(\binom{n_B}{2} + n_A n_B\right)}.
 \end{aligned}$$

Therefore we obtain $p_{A^*} > p_{B^*}$ if and only if

$$p_{AA} > \frac{p_{BB} \binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B\right) - p_{A+B} \binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2}\right)}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B\right)} = R_{A,B}. \quad (48)$$

If the inequality (12) holds, then $R_{A,B} < 0$; thus $p_{AA} > R_{A,B}$ holds for any $p_{AA} \geq 0$. Let us show the monotonically decreasing of the right hand side of the inequality (12), that is,

$$\frac{\binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B\right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2}\right)}.$$

Then we have, by remarking the relation $\binom{n_A + n_B}{2} = \binom{n_A}{2} + n_A n_B + \binom{n_B}{2}$,

$$\begin{aligned}
 & \frac{\binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)} \\
 &= \frac{\binom{n_B}{2} \left(\binom{n_A + n_B}{2} - \binom{n_A}{2} \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)} = 1 + \frac{\binom{n_A}{2} \left(\binom{n_A + n_B}{2} - \binom{n_B}{2} \right)}{\binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)} \tag{49} \\
 &= 1 + \frac{n_A(n_A - 1)}{(n_A + n_B)(n_A + n_B - 1)} \frac{n_A(n_A + 2n_B - 1)}{(n_B - n_A)(n_A + n_B - 1)} \\
 &= 1 + \frac{n_A^2(n_A - 1)}{(n_A + n_B)(n_A + n_B - 1)^2} \left(2 + \frac{3n_A - 1}{n_B - n_A} \right),
 \end{aligned}$$

which implies that this monotonically decreases as $n_B (> n_A)$ increases and tends to 1 as $n_B \rightarrow \infty$. Thus the proof is completed.

Proof of Proposition 3.5. Let us evaluate $d_A - d_B$. By Definition 3.4 and Proposition 2.3, we have

$$\begin{aligned}
 d_A - d_B &= ((n_A - 1) \cdot p_{AA} + n_B \cdot p_{AB}) - ((n_B - 1) \cdot p_{BB} + n_A \cdot p_{AB}) \\
 &= (n_A - 1)p_{AA} + (n_B - n_A) \frac{\binom{n_A + n_B}{2} p_{A+B} - \binom{n_A}{2} p_{AA} - \binom{n_B}{2} p_{BB}}{n_A n_B} - (n_B - 1)p_{BB} \\
 &= \frac{(n_A + n_B) \binom{n_A}{2} p_{AA} + (n_B - n_A) \binom{n_A + n_B}{2} p_{A+B} - (n_A + n_B) \binom{n_B}{2} p_{BB}}{n_A n_B} \\
 &= \frac{n_A + n_B}{2n_A n_B} [n_A(n_A - 1)p_{AA} + (n_B - n_A)(n_A + n_B - 1)p_{A+B} - n_B(n_B - 1)p_{BB}].
 \end{aligned}$$

Therefore we obtain $d_A > d_B$ if and only if

$$p_{AA} > p_{A+B} - \frac{n_B(n_B - 1)}{n_A(n_A - 1)}(p_{A+B} - p_{BB}) = S_{A,B}. \tag{50}$$

If the inequality (16) holds, then $S_{A,B} < 0$; thus $p_{AA} > S_{A,B}$ holds for any $p_{AA} \geq 0$. The remaining part is trivial from the form of the right hand side of the inequality (16). Thus the proof is completed.

Proof of Corollary 3.6. Let us recall Equation (11):

$$R_{A,B} = \frac{p_{BB} \binom{n_B}{2} \left(\binom{n_B}{2} + n_A n_B \right) - p_{A+B} \binom{n_A + n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B \right)}.$$

In addition, we can express $S_{A,B}$ in Equation (15) as

$$S_{A,B} = \frac{\left(\binom{n_A}{2} - \binom{n_B}{2} \right) p_{A+B} + \binom{n_B}{2} p_{BB}}{\binom{n_A}{2}}.$$

Then we have, by remarking the relation $\binom{n_A + n_B}{2} = \binom{n_A}{2} + n_A n_B + \binom{n_B}{2}$,

$$\begin{aligned}
 & S_{A,B} - R_{A,B} \\
 &= \frac{p_{A+B} \left(\binom{n_A}{2} - \binom{n_B}{2} \right) \left[\binom{n_A}{2} + n_A n_B - \binom{n_A + n_B}{2} \right] + p_{BB} \binom{n_B}{2} \left[\left(\binom{n_A}{2} + n_A n_B \right) - \left(\binom{n_B}{2} + n_A n_B \right) \right]}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B \right)} \\
 &= \frac{\left(\binom{n_A}{2} - \binom{n_B}{2} \right) \left(-p_{A+B} \binom{n_B}{2} + p_{BB} \binom{n_B}{2} \right) \binom{n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B \right)} = \frac{\binom{n_B}{2} \left(\binom{n_B}{2} - \binom{n_A}{2} \right)}{\binom{n_A}{2} \left(\binom{n_A}{2} + n_A n_B \right)} (p_{A+B} - p_{BB}),
 \end{aligned}
 \tag{51}$$

which completes the proof.

A5. Exact Expressions of Densities under Assumption 3.8

Claim 5.1. Let n_A and n_B be positive integers. Moreover let p_{A+B} , p_{BB} , p_{AB} and p_{AA} be probabilities. Under the assumption that (4) and (28) hold, we have

$$p_{AA} = \frac{\binom{n_A + n_B}{2} p_{A+B} + p_{BB} \frac{n_B}{2(n_A - 1)} \left(n_A n_B + n_A + n_B - 1 - 2\sqrt{n_A(n_A + n_B - 1)} \sqrt{n_B + \frac{p_{A+B}}{p_{BB}}(n_A - 1)(n_A + n_B)} \right)}{\binom{n_A}{2}},
 \tag{52}$$

$$p_{AB} = \frac{p_{BB}}{n_A(n_A - 1)} \left(-n_A n_B + \sqrt{n_A(n_A + n_B - 1)} \sqrt{n_B + \frac{p_{A+B}}{p_{BB}}(n_A - 1)(n_A + n_B)} \right)
 \tag{53}$$

for given p_{A+B} and p_{BB} .

Proof. It follows from (4) and (28) that

$$\binom{n_A + n_B}{2} p_{A+B} = \binom{n_A}{2} p_{AA} + n_A n_B \sqrt{p_{AA} p_{BB}} + \binom{n_B}{2} p_{BB}.
 \tag{54}$$

Let us consider the following function $f(x)$ on x :

$$f(x) = \binom{n_A}{2} x^2 + (n_A n_B \sqrt{p_{BB}}) x + \binom{n_B}{2} p_{BB} - \binom{n_A + n_B}{2} p_{A+B}.
 \tag{55}$$

Here, remarking (4), we easily see that

$$\binom{n_B}{2} p_{BB} - \binom{n_A + n_B}{2} p_{A+B} \leq 0 \quad \text{and that} \quad \binom{n_B}{2} p_{BB} - \binom{n_A + n_B}{2} p_{A+B} = 0$$

if and only if $p_{AA} = 0$. In addition, again remarking (4), we easily see that $f(1) \geq 0$; thus there exists a unique solution $x_0 \in [0, 1]$, which corresponds to $\sqrt{p_{AA}}$, such that $f(x_0) = 0$. Solving the quadratic equation $f(x) = 0$, we obtain

$$x_0 = \frac{-n_A n_B \sqrt{p_{BB}} + \sqrt{(n_A n_B)^2 p_{BB} - 4 \binom{n_A}{2} \left(\binom{n_B}{2} p_{BB} - \binom{n_A + n_B}{2} p_{A+B} \right)}}{2 \binom{n_A}{2}}. \quad (56)$$

Simplifying (56) leads to

$$\sqrt{p_{AA}} = \frac{\sqrt{p_{BB}}}{n_A (n_A - 1)} \left(-n_A n_B + \sqrt{n_A (n_A + n_B - 1)} \sqrt{n_B + \frac{p_{A+B}}{p_{BB}} (n_A - 1)(n_A + n_B)} \right). \quad (57)$$

Combining (57) with

$$p_{AB} = \sqrt{p_{AA}} \sqrt{p_{BB}} \quad \text{and} \quad p_{AA} = \frac{\binom{n_A + n_B}{2} p_{A+B} - \binom{n_B}{2} p_{BB} - n_A n_B p_{AB}}{\binom{n_A}{2}},$$

we can get the desired expressions (52) and (53) in Claim 5.1.

A6. Mathematical Interpretations on Some Models in (Wilks, Meara, & Wolter, 2005); (Meara, 2007)

A6.1. For a Model in Wilks, Meara, & Wolter 2005

In Wilks, Meara, & Wolter (2005), they modified the original model in Wilks & Meara (2002) by considering some association “indirectly” linked between two words. In terms of a random digraph $G = (V(G), A(G))$, two words x and y are associated if $[x, y] \in A(G)$ or $[y, x] \in A(G)$ in the original model; in addition, they consider two words x and y are also associated “indirectly” if there exists a word $z \in V(G)$ such that $[x, z] \in A(G)$ and $[y, z] \in A(G)$. In such a model, referring to Remark 2.4 for random digraphs, we can easily find, for a randomly selected n -words experiment,

$$\begin{aligned} \text{“hit rate”} &= 1 - (1 - p)^{n(n-1)} \left((1 - p)^n + np(1 - p)^{n-1} \right)^{\#V(G)-n} \\ &= 1 - (1 - p)^{(n-1)\#V(G)} (1 + (n-1)p)^{\#V(G)-n}. \end{aligned} \quad (58)$$

Here we should remark that the values $(1 - p)^{n(n-1)}$ and $(1 - p)^n + np(1 - p)^{n-1}$ imply the probabilities of no direct association occurring within a randomly selected n -words and no indirect association occurring in a randomly selected n -words via some word z outside n -words, where the total number of such words are $\#V(G) - n$, respectively. Following (Wilks, Meara, & Wolter, 2005), we put $\#V(G) = 1000$ and $n = 5$ in (58); we have

$$\text{“hit rate”} = 1 - (1 - p)^{4000} (1 + 4p)^{995}, \quad (59)$$

where we put $p = k/999$ and k is the “number of links per word”, in other words, the regularity of outer degree in digraph; this formula (59) almost recovers the results of simulations in Table 3 in (Wilks, Meara, & Wolter, 2005). See **Table A1**.

Table A1. Outputs of (59) for $p = k/999$ ($k = 1, 2, \dots, 20$).

Number of links per word	1	2	3	4	5	6	7	8	9	10
hit rate	0.0295	0.0767	0.1387	0.2121	0.2932	0.3782	0.4633	0.5456	0.6225	0.6923
Number of links per word	11	12	13	14	15	16	17	18	19	20
hit rate	0.7539	0.8068	0.8511	0.8874	0.9163	0.9390	0.9563	0.9693	0.9788	0.9856

A6.2. For a Model in Meara (2007)

In Meara (2007), as a kind of small world model, the following random digraph model is introduced: 1) 1000 words are grouped into 20 cluster, each consisting of 50 words; 2) within these clusters, every word has random k -out-degree ($k \geq 3$); 3) among these clusters, we give additional 50 arcs. For details, see Section III.2 in (Meara, 2007). In the sense of the probability of arc occurring, it is obvious that $q = k/49$ within every cluster is quite denser than $r = 50/(950 \times 50 \times 20) = 1/1900$ among clusters; this implies every cluster is almost isolated. For such a model, the results of “hit rates” in the task of five randomly selected words under the computer simulations are given. Now let us interpret these results in our context and methods. First we divide into 7 cases based on the number and type of clusters to which selected 5 words belong: case (1, 1, 1, 1, 1), case (2, 1, 1, 1), case (2, 2, 1), case (3, 1, 1), case (3, 2), case (4, 1), case (5).

1) For case (1, 1, 1, 1, 1): This implies there exist 5 clusters such that each of them includes just one word in randomly selected 5 words. Then, considering “direct” and “indirect” associations by the similar way to that in the previous subsection, we can obtain the probability of association occurring within 5 words in this case that

$$p_1 = \frac{1}{\binom{1000}{5}} \binom{20}{5} \cdot 50^5 \cdot \left(1 - (1-r)^{20} \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{750} \right). \quad (60)$$

here any association occurs among clusters only and never occurs within each cluster in this case.

2) For case (2, 1, 1, 1): This implies there exist 4 clusters such that just one of them includes two words and each of three others does just one word. Then, considering “direct” and “indirect” associations within cluster and among clusters, we can obtain the probability of association occurring within 5 words in this case that

$$p_2 = \frac{1}{\binom{1000}{5}} \binom{20}{4} \cdot 4 \cdot \binom{50}{2} \cdot 50^3 \cdot \left(1 - (1-q)^2 \left((1-q)^2 + 2q(1-q) \right)^{48} \cdot (1-r)^{20} \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{800} \right). \quad (61)$$

3) For case (2, 2, 1): This implies there exist 3 clusters such that one of them includes one word and each of two others does just two words. Similarly we have

$$p_3 = \frac{1}{\binom{1000}{5}} \binom{20}{3} \cdot 3 \cdot \binom{50}{2} \cdot \binom{50}{2} \cdot 50 \times \left(1 - (1-q)^4 \left((1-q)^2 + 2q(1-q) \right)^{96} \cdot (1-r)^{16} \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{850} \right). \quad (62)$$

4) For case (3, 1, 1): This implies there exist 3 clusters such that one of them includes 3 words and each of two others does just one words. Similarly we have

$$p_4 = \frac{1}{\binom{1000}{5}} \binom{20}{3} \cdot 3 \cdot \binom{50}{3} \cdot 50^2 \times \left(1 - (1-q)^6 \left((1-q)^3 + 3q(1-q)^2 \right)^{47} \cdot (1-r)^{14} \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{850} \right). \quad (63)$$

5) For case (3, 2): This implies there exist 2 clusters such that one of them includes 3 words and another does just 2 words. Similarly we have

$$p_5 = \frac{1}{\binom{1000}{5}} \binom{20}{2} \cdot 2 \cdot \binom{50}{3} \cdot \binom{50}{2} \cdot \left(1 - (1-q)^8 \left((1-q)^2 + 2q(1-q) \right)^{48} \times \left((1-q)^3 + 3q(1-q)^2 \right)^{47} \cdot (1-r)^{12} \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{900} \right). \quad (64)$$

6) For case (4, 1): This implies there exist 2 clusters such that one of them includes 4 words and another does just 1 word. Similarly we have

$$p_6 = \frac{1}{\binom{1000}{5}} \binom{20}{2} \cdot 2 \cdot \binom{50}{4} \cdot 50 \times \left(1 - (1-q)^{12} \left((1-q)^4 + 4q(1-q)^3 \right)^{46} \cdot (1-r)^8 \cdot \left((1-r)^5 + 5r(1-r)^4 \right)^{900} \right). \quad (65)$$

7) For case (5): This implies there exists just one cluster including 5 words selected randomly. Similarly we have

$$p_7 = \frac{1}{\binom{1000}{5}} \binom{20}{1} \cdot \binom{50}{5} \cdot \left(1 - (1-q)^{20} \cdot \left((1-q)^5 + 5q(1-q)^4 \right)^{45} \right). \quad (66)$$

Here any association occurs within one cluster only and never occurs among clusters in this case.

Thus we find, combining all formulae stated above,

$$\text{“hit rate”} = \sum_{\ell=1}^7 p_{\ell} \quad (67)$$

for a randomly selected 5-words experiment. This is valid for general probabilities q within each cluster and r among clusters. Furthermore it is obvious that the function (67) on q and r , both of which are defined on $[0,1]$, is monotonically

increasing as q or r increases. Incidentally, we remark that, if every cluster is completely isolated, equivalently $r = 0$, then the “hit rate” takes values in $[0.00000000, 0.41274800]$. On the other hand, the “hit rate” is bounded above and below for a fixed $r \in [0, 1]$. For example, if $r = 1/19000$, then the “hit rate” takes values in $[0.001021923, 0.413378042]$; these bounds are almost same as those in the case where $r = 0$, respectively, thus we may say every cluster is almost isolated even if $r = 1/19000 \neq 0$. Similarly, for a fixed $q \in [0, 1]$, the “hit rate” is also bounded; if $q = 8/49$, then the “hit rate” takes values in $[0.3425407, 1.0000000]$.

Now, in order to compare with the results of simulations in (Meara, 2007), let us put $q = k/49$ and $r = 1/1900$ and observe every output value of (67) for $k = 3, 4, \dots, 20$. We immediately find our **Table A2** almost recovers the results of simulations in Figure 7 in Meara (2007).

We may recover other results in simulation by similar discussion.

Table A2. Outputs of (67) for $q = k/49$ ($k = 3, 4, \dots, 20$) and $r = 1/1900$.

Number of links per word within cluster	3	4	5	6	7	8	9	10	11
hit rate	0.1233	0.1767	0.2279	0.2737	0.3123	0.3432	0.3666	0.3836	0.3952
Number of links per word within cluster	12	13	14	15	16	17	18	19	20
hit rate	0.4028	0.4075	0.4102	0.4118	0.4126	0.4130	0.4132	0.4133	0.4133