# One Hundred Years of Linguistics and
# Word Assocation（WA）Research

Thomas H. Caton
（2019年11月27日受理）

## Introduction

For over a century, Word Association (WA) tasks have been used to investigate the content and organization of words and semantic concepts. WA was first developed as a research instrument by Francis Galton who introduced the first WA test to psychology (Galton 1879). He used a list of 75 stimulus words which he read and noted his subjects' responses and suggested that there might be a link between these and a person's IQ (Intelligence Quotient). Carl Jung (1910) subsequently developed the WA test as a clinical diagnostic tool, attempting to identify abnormal patterns of response as a means to identify psychological complexes. Many other early studies (eg, Kent and Rosanoff 1910), also focused on basic concepts, interpreting responses as indicators of general behaviour and going on to identify and diagnose intellectual and emotional deficiencies and other forms of psychological abnormality. By the 1950s market researchers began to realize that WA could be used to ensure the proper message is conveyed by names or adjectives when promoting a company's products (eg. Vicary 1948). The WA map in Fig.1 shows an example of this by illustrating the strength of



**Fig.1　A map showing the strength of word association with a well-known brand of chocolate (from: https://godivasmu.wixsite.com/godiva)**

brand awareness and brand image of the well known chocolate brand, Godiva, in consumers' minds. The map shows the most strongly associated words in larger size with consumers preferring the words "Chocolates", "Women", "Expensive", "Gift" and "Prestige."

## Word Association (WA) and Linguistics

This paper focusses on a different area, that of linguistics and second language acquisition (L2) research and describes the contribution that WA tasks have made to this field. Using two specific examples nearly a century apart, it will attempt to trace the development of WA as a research tool which can analyze how L2 learners organize vocabulary that they have learnt and the influence that specific variables can have on lexical access. It might also help with our understanding how the mental lexicon can develop over time (Fitzpatrick, 2007:319). In applied linguistics, the main interest has been on how L2 items are acquired by learners and the ways in which WA responses might reflect the development of L2 proficiency (e.g. Kruse et al. 1987).

This first paper, which is a replication study carried out in 1918, helps strengthen earlier evidence that WA becomes more effective the more frequently occurring a word appears in response to a stimulus (or cue) word. Furthermore, this effectiveness can increase as the time taken for the association to occur decreases. The study is one of the first attempts to explain the associational basis of the English language and goes on to make comparisons with a second language, German. The second paper which will be discussed is an exploratory study undertaken in 2011 and again examines the type and speed of responses given to stimulus words in two

languages, English and this time, Spanish. Unlike the first paper it also explores if both variables are affected by the language proficiency of participants and whether L2 responses are mediated through L1.

## E.A. Esper's 1918 paper on Word Association (WA) and Language

The paper, "*A contribution to the experimental study of analogy,*" by E.A. Esper is a good example of an early study on word association and attempts to describe the characteristics of stimulus and associate words which enable them to become more linguistically effective. It first replicates and then builds on an earlier word association experiment completed by Thumb and Marbe (1901) using English instead of German materials. It succeeds in confirming many of the original experiment's findings and further expands the boundaries of this area of research.

Thumb and Marbe established that the more frequently a response word occurs with word association then the more rapidly it is likely to take place. In their original experiment 60 words are chosen from six categories and given one at a time in random order. There are eight observers who give a single response to each of these stimulus words. A second experiment uses 80 verbs with a similar number of observers. Esper's 1918 experiment uses the same number (60) of mostly identical words from six almost similar categories. The main difference, however, is in the total number of observers (126) which are divided into three groups: A. Educated adults (100), B. Children (11) and C. Less educated adults (15). The stimulus words largely correspond with the earlier experiment as far possible with 50 out of the original 60 words being used. Reaction times are measured with a Hipp chronoscope and a stopwatch with participants being encouraged to produce single response words in a spontaneous manner.

There is agreement between the German and English results with 35 out of 50 responses being identical. It is also found that stimulus-words of a given category are responded to predominantly with words of the same category. For example 85.3% of Esper's responses to adjective stimulus words are adjectives while for Thumb and Marbe this figure is 87.5%. The exception is with responses with adverbs of place and verbs with far lower figures occurring. Of particular note is that with responses to adjectives many are adjectives of opposite meaning and there are many cases where reciprocal associations take place. With reaction times both German and English-speaking groups confirmed that the more frequently a response occurs then the shorter the reaction time. Except with reaction times there are only small differences between the three groups. Groups B and C only display small differences in response words used from group A. 86.4% of the most frequent responses from group B and 83.1% from group C are the same as group A. This shows that the associative processes of children and less educated adults are similar to those of educated adults.

In almost all respects Esper's results confirm Thumb and Marbe's findings: (1) word association becomes more effective if it occurs frequently and (2) word association becomes more effective with the decrease in the average time taken for the word association to take place. The paper also shows that response words tend to be predominantly from the same word category and that a high degree of agreement can be found between the German and English languages. Finally, it demonstrates that educated, less educated and child participants are likely to favour similar association characteristics although the speeds at which associations take place are likely to differ.

## Fitzpatrick and Izura's 2011 paper on Word Association (WA) in L1 and L2

The second paper, "*Word Association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation,*" by Fitzpatrick and Izura is an example of a modern approach to the use of WA and applied linguistics. Its main aims are to firstly, establish whether some response types are produced more rapidly and more often than others which is similar to what Esper's paper sets out to achieve. Secondly, it compares the same participants' responses to stimulus words in their L1 (Spanish) and L2 (English) as well investigating whether their L2 responses were mediated through L1 or affected by L2 proficiency. This second aim helps demonstrate that essentially the same WA

tool can be used in new ways to explore differences that may occur through multiple languages and proficiency levels.

The first experiment uses 24 Spanish L1 speakers who are proficient in English (L2). Their proficiency was measured using the Eurocentres Vocabulary Size Test (EVST; Meara and Jones 1990) with the group's average level found to be upper intermediate. The participants completed two WA tasks: one in Spanish (L1) and one in English (L2) and were asked to produce a single associate to each cue word as rapidly as possible. 95 Spanish cue words and 95 English cue words were used for the task and these two word lists were matched on word length, imageability (or the ability to invoke a mental image), word frequency, syntactic class and age of acquisition. Two sets of data were produced, one Spanish (94) and one English (94) and the response word and response time were noted. WA responses from both tasks were divided into six categories (1) form and meaning (2) meaning and collocation (3) collocation (4) form (5) equivalent meaning and (6) nonequivalent meaning (see P.384 for examples).

With comparing the L1 and L2 and the responses given per category there were generally fewer responses to L2 cues than L1 cues. The interaction between language and response category was significant showing that subjects tended to choose words belonging to similar categories for both tasks. Response time comparisons per category revealed a significant main effect of language with faster responses in the L1 than the L2. There were also faster responses in some categories compared to others. Meaning and collocation category had the fastest response time and nonequivalent meaning the slowest. As a group L2 responses were all slower. The difference between L1 and L2 response times generally decreased with subjects who had higher L2 proficiency and a larger vocabulary size.

The second experiment describes is a lexical decision task where 36 real and 36 invented Spanish words were presented to participants. Real words were divided into sets A and B (18 words in each). Set A consisted of 18 primed words or translation equivalents which had been used as cues in the English word association task. Set B consisted of non-primed words that were not

translation equivalents of any items in the English WA task. Findings from the lexical decision task showed that participants were generally faster at recognizing primed Spanish words. These 'primed words' are the translation equivalents of English words used as cues in the previously completed English WA task. This suggests that in some cases, the Spanish L1 words were somehow activated while participants were doing the English L2 WA task. This 'priming effect' was only really evident with lower proficiency participants.

## Discussion

With the earlier study, the focus is more conceptual, with responses interpreted as indicators of general behaviour but it still represents an important step in the investigation into the development and organization of the mental lexicon. Nonetheless there are a number of points about Esper's paper which the modern reader should be aware of. These are concerned with differences in methodologies, overall aims of the research and the fact that many earlier studies failed to understand the complexity of WA and human behavior.

Firstly, there seems to be a lack of a clear hypothesis or clear aim. On P.469 the author describes a 'problem' (in fact two research questions) which the paper sets out to resolve but expresses these in very general terms. The notion of a null hypothesis, which is accepted or rejected on the basis of a series of investigations, is absent. A more productive approach might be to create a list of questions requiring simple yes or no answers which then can be supported or refuted with the provided evidence. In fact, the aims of the research gradually become easier to appreciate later but some early concrete questions would enhance the average reader's understanding of the concepts involved. The 2011 paper clearly avoids any misconceptions at the start by setting out three basic research questions which addressed in order. Secondly, in the first paper some of the terms used and concepts considered are rather dated particularly the use of the word "uneducated" to describe members of one of the three comparison groups participating in the WA tasks. It might be more informative to provide descriptions on the actual level education achieved by the different participants although in a modern setting

the concept of testing participants on this basis might be controversial. This problem is avoided altogether by perhaps differentiating between participants using language proficiency. Similarly there could be concerns about the use of long German quotations on P.479 and 485. There seems to be the assumption that the reader has a sufficient knowledge of German to understand these, which might possibly be true with an intended contemporary readership, but is not really appropriate for the modern age.

A third point is that Esper's paper sets out to establish, 'an associational basis of the English language' (P.469) but even early on there are questions about how far the data collected from the 126 respondents might represent the overall population. More recent studies have questioned the validity of the assumption that there is coherent norm behaviour in native speakers, with Fitzpatrick (2007) finding that, "not only do [native speakers] vary in the actual words they produce, they also seem to vary in the types of association they make" (2007: 327). To decide if the data collected is indeed representative it is possible to make an interesting comparison with the University of South Florida norms (USF norms, Nelson, McEvoy, & Schreiber, 1998), a large collection of single word association responses for more than 5,000 cue words. For many words the responses of both Esper and the USF norms are comparable. For instance the cue word father elicits the most frequently occurring response as mother in both cases (Esper 72.6% and USF 70.6%) but in other cases the same does not hold true. The cue word big elicits the response little (Esper 55.9%) but a different word *small* with the second group (USF 63.5%). Many other disparities are likely to exist between the two sets of data revealing the fact that perhaps Esper perhaps was not yet fully aware of the underlying complexities of human cognitive behaviour during the earliest days of WA research. Fitzpatrick and Izura have the advantage of modern resources like the USF norms which can be used to assess the suitability of cues for WA tasks.

Finally, there are questions relating to sample size and reliability used in both papers. One would expect general agreement with the fact that the more observations there are then the better the chances are that a sample will predict something about the language

and behaviour of a population under study. With only eight observers taking part in Thumb and Marbe's research there is likely to be a concern on the problem of reliability with the data produced and this could cast some doubt whether the English and German results that Esper gives are indeed comparable. Also there might be some risk in making comparisons between different-sized groups of observers (A, B and C) involved in the study although Esper appears to appreciate some of the difficulties with doing so, stating that, "the supplementary groups (B and C) are not large enough, however, to permit us to make comparisons between the results for single words." (P.484). Ideally these three groups, along with those involved in Thumb and Marbe's earlier experiments, should be approximately equal in size as this would help provide data with a higher degree of confidence and reliability. With Fitzpatrick and Izura the sample size (24) was the same in both experiments allowing for a greater degree of confidence in their results.

## Conclusion

Comparing two papers from very different periods of research using a similar measurement tool revealed a number of interesting features. Similarities between the two are evident. They both used WA tasks with carefully chosen cue words that required a single response for each. They also attempted to categorize each response although the way in which this was done was very different in each case. They measured the time that responses took for each cue and found similar results in each case, namely that when there was a stronger association between words, the time taken for the response decreased. The differences in each approach are also significant. The 1918 paper compares its results gained from English L1 participants with those obtained from German L1 ones in an earlier experiment and is a good example of a replication study. The later paper uses a single sample of highly proficient foreign language speakers and requires them to complete two WA tasks in both L1 and L2s. This allows for the opportunity to examine the different ways in which both L1 and L2 acquisition processes work by seeing how learners, who may have different language proficiency levels, approach WA tasks.

# References

Esper, E. A. (1918)  A contribution to the experimental study of analogy  *Psychological Review* 25(6), 468-487

Fitzpatrick, T. (2007) Word association patterns: Unpacking the assumptions *International Journal of Applied Linguistics* 17(3), 319–331.

Fitzpatrick, T. and Izura, C. (2011) Word Association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation *Studies in Second language Acquisition* 33, 373-398

Galton, F. (1879). 'Psychometric experiments,' *Brain* 2 149–62.

Godiva Chocolate Company, "Word Association Map."

Retrieved from https://godivasmu.wixsite.com/godiva/word-association-map (1st October 2019)

Jung, C. G. (1910) The association method *The American Journal of Psychology* 21/2: 219–69.

Kent, G. H. and A. J. Rosanoff. 1910. 'A study of association in insanity,' *American Journal of Insanity* 67: 37–96, 317–90.

Kruse, H., Pankhurst, J., and Sharwood Smith, M. (1987). A multiple word association probe in second language acquisition research *Studies in Second Language Acquisition*, 9, 141 –154.

Meara, P.M and Jones, G. (1990) *Eurocentres vocabulary size test 10KA* Zurich: Eurocentres

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998) The University of South Florida word association, rhyme, and word fragment norms

Retrieved from http://w3.usf.edu/FreeAssociation/ (1st October 2019)

Thumb, A. & Marbe, K. (1901) *Experimentelle Untersuchungen über die1 psychologischen Grundlagen der sprachlichen Analogiebildung* Leipzig

Vicary, J.M. (1948) Word Association and Opinion Research: "Advertising"  — an Illustrative Example *Public Opinion Quarterly* 12(1) 81-98 https://doi.org/10.1086/265923