A Model for Assessing the Difficulty Rate of Lexical Retrieval from Long-Term Memory

By:
Ali Jani

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree
of MA in TEFL

Supervisor: Dr. Amerian

Adviser: Dr. Ahmadian

Department of English Language and Literature

Arak University

December, 2009

Abstract

The purpose of this study was to test the validity of a model by which it has been claimed that the difficulty levels associated with lexical retrieval from long-term memory is estimated. The difficulty rates of words are assessed through a simple calculation in which the number of syllables, their structures, as well as the leaner's familiarity with initial and final morphemes of the words are taken into account as the criteria of this calculation. The model has been devised and developed using some lines of evidence from the available literature (mainly on short-term memory) and analyzing English language learners' errors made while recalling words in language classes. To achieve this purpose of testing the validity of the model, 40 words were taught to 40 advanced language learners of 2 institutes in Kermanshah in 5 sessions during a three-week treatment programme. The 40 target items were 8 sets of 5 words – each set represented a difficulty rate (ranging from the easiest to the hardest as assessed by the model). In each session, 8 words of various rates of difficulty were taught. According to the model's predictions, the words estimated as easy ones, or those having low difficulty rates, should have been recalled more frequently than those with higher difficulty rates. After a two-week interval, they were tested on a simple task which included recalling the words on their definitions or meanings. Data analysis (t-tests and rank-order correlations) revealed some positive and negative relationships between the difficulty rates assessed for the words and the numbers of the times they were retrieved by the participants, implying that words should be treated as unequal linguistic units learned (retrieved) at various levels of effort. The modest degree of the obtained validity of the model needs to be further investigated.

Key words: *Lexical retrieval, long-term memory, morpheme, syllable, vowel, consonant, coda, difficulty rate*

Dedication

To my mother, brothers and sisters for their patience and support, and in memory of my father and brother.

Acknowledgements

Entering into a novel (or an *apparently* novel) research area demands receiving sustainable support from an academic individual. I am deeply grateful to my supervisor, Dr Amerian, for his encouragement and invaluable comments that I received at every stage of the study.

Many thanks also go to Dr Ahmadian, my adviser; I am appreciative of his insightful guidance and suggestions.

I would also like to express my gratitude to Dr Yazdani, for his constructive criticisms upon reading the proposal of this thesis.

I also wish to thank Mr Mohammadi, for always being extremely helpful and considerate.

My special appreciation also goes to Dr Khodadadi, from Statistics Department of Razi University of Kermanshah, for his statistical help.

In addition, I would like to thank the principals and teachers of Nosrat and Novin Language Institutes (in Kermanshah), for being so cooperative, and all the language learners of these two institutes who willingly participated in this study.

I also want to thank my friend, Mr Khosro Bahramloo, for helping me with some of the statistical calculations.

I am also grateful to my brothers, Hassan Ali and Hamid Reza, who helped me a lot with the provision of the sources, and also for being directly involved in the study as two of the participants.

Table of Contents

Appendices

List of Tables

Table Page

List of Figures

List of Abbreviations

C: consonant

DR: difficulty rate

L1: first language

L2: second language

LTM: long-term memory

ρ: rho, rank-order correlation coefficient

RS: retrieval score

RSVP: rapid serial visual presentation

SD: standard deviation

SLA: second language acquisition

SLVA: second language vocabulary acquisition

STM: short-term memory

TOEFL: Test of English as a Foreign Language

TRS: total retrieval score or the number of the times a set of 5 words or the 8 words presented in each session have been retrieved – it has been specified to which group of the words it belongs

V: vowel

Chapter I

Introduction

In this opening chapter, an overall description of the study is provided. This is accomplished via shortly providing an introduction to and a background of the subject at hand, along with introducing the model on which the study is based, presenting the statement of the problem, purpose and significance of the study, research assumptions, questions and hypotheses. Finally, key terms are defined, and the organization of the study is sketched out.

1.1 General overview

Vocabulary learning has always been a major concern of language learners in the long and difficult process of acquiring a second language. In spite of this, except for some unsustainable attempts by those whose names come under the much heard name, the Grammar-Translation Method, and also what Michael West did (see Howatt & Widdowson, 2004), language teaching theoreticians almost overlooked the key role of vocabulary in their approaches until the late 1970s; this ignorance, “in large part”, arose out of putting too much emphasis on “grammatical and phonological structure” by dominant American linguistic theories during the 1940s, 1950s and 1960s (DeCarrico, 2001, pp. 285286). In fact, neither Chomsky’s transformational linguistics nor the emergence of communicative competence by Hymes in 1972 could appropriately improve the inferior status of vocabulary in language teaching. The former criticized behaviourists and proposed a set of abstract rules for generating an unlimited number of sentences (Carter & McCarthy, 1998, cited in DeCarrico, 2001). The latter, though did not

reject Chomsky's ideas, placed great emphasis on communicative competence – the ability to use language correctly and properly. Both the generative transformational theory and the communicative competence approach gave vocabulary secondary status (Schmitt, 2000). Several promising events helped to change this view: the appearance of comprehension approaches in language teaching arena that proved the key role of vocabulary in communication, applied linguists' attempts to carry out research projects on lexical related issues, and finally, developing computerized analysis of language corpora, enabling researchers to provide a great deal of information about the frequency of words, and how differently they act in written and spoken samples (Sinclair & Renouf, 1988, cited in Nunan, 1999, pp. 102–103).

For different skills different properties of words are involved. In reading, for example, language learners must recognize the words, using their orthographic and morphological knowledge of words, while in listening they have to rely mainly on phonological properties of the acquired words. In productive skills, they need to have a good repertoire of vocabulary to retrieve from long-term memory (LTM). Nevertheless, the words are not equally difficult to acquire. Depending on their lengths, structures, and pronunciations, and some other factors, they are learned at various rates. Another factor pertaining to word learning is the frequency of the words – the rate of their occurrences or uses. In the case of high-frequency words, language learners face relatively fewer problems. But acquiring low-frequency words is the main challenge in second language vocabulary acquisition (SLVA). The acquisition of both low and high-frequency words can be studied from different perspectives because various factors are involved in SLVA, making it a multidimensional phenomenon. Thus, each study focuses on one, or at most on a few, of these factors (see e.g., Barcroft, 2004 for an 'overview' of the important studies). Each individual study provides some information to deepen our understanding of SLVA. Out of these findings, cognitive models – e.g., Forster's 1976 search

model of lexical access (see Garman, 1990, pp. 266 –276), and Levelt et al.'s 1999 theory of spoken word production (presented in Chapter II of this thesis) – have been developed to investigate lexical-related issues from a more internal point of view, representing some underlying mechanisms in the process of L1 and even L2 learning – what goes in the mind of a language learner.

Like other aspects of language, vocabulary learning of young children has been used as one commonly applied procedure, but not a completely right way, because of the differences, to acquire information about SLVA. Studies investigating early stages of vocabulary acquisition of very young children have shown that acquiring "skills in recognizing and storing details of phonological input" precedes word production. These skills enable them to be sensitive and aware of the various properties of phonological input such as "rhythmic patterns and segmental combination", and to store words as segments even during the first year of their life (Chiat, 2006, p. 553). It has also been hypothesized that during the second year of their life, children have the ability to acquire a large number of new words rapidly after just little exposure. Various terms have been used to label this phenomenon: "*the vocabulary spurt*", "*the vocabulary burst*", and "*the naming explosion*" (Dromi, 1987; Nelson, 1973). The vocabulary spurt, though has been proved to exist among many children, has not yet been supported as a universal phenomenon (Ganger & Brent, 2004). Another ability attributed to young children of around 2–10 years old is "*fast mapping*" which has been defined as, quickly mapping "a novel word [on]to a novel object" (Horst & Samuelson, 2008, p. 129). Acquiring a great number of words is absolutely necessary for children to be able to survive various speech events. This necessity is felt to be overwhelming when the learner is an adult foreign language learner who lacks such an innate ability and the language in question is English which has a formidably large vocabulary.

The number of words in English as a unique characteristic of this language is strikingly large and expanding constantly. There are two main reasons behind this largeness. First, historically, many Roman French words entered into English after England was invaded by the Norman in 1066. Again, in the nineteen century, a large number of words based on the two highly prestigious languages, Latin and Greek, were introduced into English by scholars to enrich and expand it further, forming a vocabulary composed of "layers of words which are heavily marked from the stylistic point of view". Second, as an inherent property, English tends "to use rare and unusual words" whereas other languages adopt circumlocutory methods that require "simpler" and more common words. This means that English learners have to know sufficient words to act efficiently in various areas of discourse. With regard to its lexicon, English is a difficult language. Learners who accomplish this obligatory task are rewarded with success in the main skills, especially in reading and listening (Nation & Meara, 2002, pp. 48–52). In general, the importance of vocabulary is well reflected in a quotation from Chomsky (1995, p. 131), where he sees language acquisition mainly as "a matter of determining lexical idiosyncrasies".

The significance of vocabulary makes it mandatory for language teachers and syllabus designers to dedicate a great part of their time and energy to lexical items of the language. Thus, in a vocabulary teaching programme, widely investigated issues like vocabulary selection, methods of teaching words (e.g., Davies & Pearse, 2000; DeCarrico, 2001; Hunt & Beglar, 2005; Nation & Meara, 2002), depth and breadth of vocabulary (e.g., Laufer & Hill, 2000; Nation, 1990; Nassaji, 2004), vocabulary and grammar (e.g., Cook, 2001; Nunan, 1999), vocabulary related strategies (e.g., Cook, 2001; O'Malley & Chamot, 1990; Thomas & Wang, 1996), reading and vocabulary (e.g., Krashen, 1985, 1989, cited in Lightbown & Spada, 2006; Laufer, 2003; Pigada & Schmitt, 2006), effective vocabulary learning activities (e.g., Barcroft,

2004; Karpicke & Roediger, 2007; Nation & Meara, 2002; Pashler, Rohrer, Cepeda & Carpenter; 2007), and a good number of other issues deserve careful consideration. But, presumably, at the heart of these issues – or at least as important as them – is the notion that words are not equally difficult or easy to learn and that there are word-related factors that ease or make their learning more difficult. This study is an attempt to gain some information about some of those factors affecting lexical retrieval when long term retention is concerned. To systemize this inquiry, several of those factors which were supposed to affect lexical retrieval have been presented in the form of a model; thus, the focus of the study is on testing the model, and then providing beneficial suggestions for its later improvement. Since the model stands on the core of the study, as the initial step, the model should be introduced and discussed fully enough. But to begin with, a succinct background of the previously acquired insights into the pertinent topics seems imperative.

1.2 Background

Plenty of early studies exploring the relationship between memory and lexical recall have centered upon short-term memory (STM) and characteristics of words such as length and frequency. Baddeley, Thomson, and Buchanan (1975), claiming that the capacity of STM is limited, could show that lists of long words were more difficult to recall than lists of short ones. This was a starting point for introducing the notion of length-effect as a factor influencing the process of lexical retrieval. Several properties of words were regarded as length: the number of syllables, duration or the time needed to articulate a word, the number of phonemes, and the number of letters in each word (e.g., Baddeley, Chincotta, Stafford & Turk, 2002; Baddeley et al., 1975; Coltheart, Mondy, Dux & Stephenson, 2004; Logie, Sala, Wynn, & Baddeley, 2000; Meyer, Roelfs & Levelt, 2002; Narine, Neath & Serra, 1997; New, Ferrand, Pallier, &

Brysbaert, 2008; Nickels & Howard, 2004; Santos et al., 2006).

Another factor that has been proved to influence word recognition, and recall is the frequency of the word. High frequency is hypothesized to act as a word characteristic facilitating recognition (Macleod & Kampe, 1996), and recall (Alario, Costa & Caramazza, 2002; Jurafsky, 2003; Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002; Stemberger & MacWhinney, 1986).

Word syllable has also been shown as a factor affecting visual word recognition (Álvarez, Carreiras, & de Vega, 2000; Conrad, Grainger, Jacobs, 2007; Ferrand, Segui, & Humphreys, 1997; Hutzler, Conrad, & Jacobs, 2005). The number of syllables and onset complexity were also identified as variables influencing "vocal production latency" which is defined as the time-lag between seeing a word or a picture representing a word and starting its pronunciation (Santiago, Mackay, & Palma, 2002). Furthermore, frequency of syllables has been recognized to have effect on phonological STM tasks, so that high-frequency syllables facilitated performing such tasks (Nimmo & Rooderys, 2002).

Lexical processing is one of the other issues that have become the subject of many recent word-related studies. Researchers working on processing of words in the mind have confirmed the fundamental role of morphemic units in the process (Frost & Grainger, 2000; Frost, Kugler, Deutsch & Froster, 2005; Rastle, Davis, Marslen-Wilson & Tyler, 2000). Some other studies have also provided data for supporting the role of orthography in lexical access (Rapp, Benzing, & Caramazza, 1997) and silent reading (Taft & Kougious, 2004).

Word characteristics influencing recognition, recall and other aspects of vocabulary learning are not limited to those mentioned above. Among the other factors are: phonotactics, probabilistic phonotactics, neighborhood density, context variability, and priming. In the subsequent chapter they are dealt with briefly.

The model presented in this study was developed, making use of the findings of the relevant above-mentioned studies, and the data obtained via analyzing a huge number of errors made by hundreds of students during several educational years. The researcher's object of developing such a model has been to provide a valid and reliable tool for assessing the extent of difficulty that a learner is supposed to face when recalling a word from LTM.

In the following section of the thesis, the model is introduced at some length, though deeper appreciation of its components entails reading Chapter II.

1.3 The Model

The model came into being via utilizing two main sources. First, and mostly as a source of inspiration, there existed dozens of recorded errors made by this researcher's beginning students during different exercises requiring lexical retrieval that led him to seek for a systematic pattern underlying word learning in general and lexical retrieval in specific. Analysis of such errors revealed that more errors occurred when learners were required to recall long words, and words beginning with a vowel or a diphthong (e.g., *vegetable*, *question*, *neighbour*, *careless*, as long words and *enjoy*, *arrive*, *eraser*, *ever*, *over*, *other*, *uncle*, *about*, *ear* as words beginning with a vowel or diphthong). Thus, word length and initial vowels or diphthongs were assumed to be factors causing difficulty in lexical recall. Subsequent experiences of teaching English to more advanced language learners yielded evidence that more or less gave support to such assumptions. Second, more reliable evidence to base the model on came from the findings of the studies on memory system, word length, syllables and the relationships between these word characteristics and vocabulary learning or lexical-related STM tasks. The main findings of these studies that were utilized were those that confirmed (a) morphemes and syllables as key units of words (e.g., Frost et al., 2005; Frost & Grainger 2000; Rastle et al., 2000; Levelt et al., 1999),

(b) the effect of word length on lexical retrieval in a way that longer words are claimed to be recalled with more difficulty than shorter ones (e.g., Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Santos et al., 2006), and (c) the facilitatory effect of LTM knowledge and familiarity with the target language on word recall or learning (Nairne et al., 1997; Masoura & Gathercole, 2005).

It is supposed that the model is capable of estimating the difficulty rate (DR) of each word to retrieve from LTM through a *simple* calculation. The basic unit for this calculation is the syllable; but it does not mean that the words with more syllables are necessarily more difficult than those with fewer numbers of syllables. The DR of a syllable having the **coda** is two. If it lacks the **coda**, it takes just one point (the results of the study by Nimmo and Roodenrys, 2002 helps to assume such a role for codas). Because the first syllable plays a more important role, if it begins with a vowel or diphthong, it should be added another point. In other words, such syllables at the beginning of a word have a DR of two if they do not have a **coda**, and three if they have it. The evidence used to support the validity of this element of the model comes from the classroom activities and tasks that required students to recall newly learned words in which students had great difficulty in remembering words beginning with such sounds. On the other hand, if the first syllable is a familiar prefix or a familiar free morpheme, depending on the learner's vocabulary knowledge, as it facilitates recalling, it receives a –1 point. This means that such syllables decrease the DR when they occur at the beginning of a word. This is because it seems that in the mental process of lexical access, the first syllable acts as "an access code", as Taft & Forster (1976, cited in Álvarez et al., 2000, p. 346) put it. Familiar suffixes and familiar free morphemes at the end of the words are assumed to have no effects on recalling; they are mainly predictable, and as they are decoded last, they have almost no influence on retrieving their previous syllables; thus, the DR for such chunks is zero.

Accordingly, based on this model, the words *landfall*, *ontology*, *lee*, *lop, heron, tandem, arbutus,* and *entourage* have the DRs of –1, 0, 1, 2, 3, 4, 5 and 6 respectively – their DRs were calculated regarding my own knowledge of words, suffixes and prefixes. Therefore, the DR of a word may change from one person to another. The minimum DR would be –1; the words having –1 DR are usually words containing two or more familiar morphemes. Words with more than 6 DR are very rare for advanced learners with a considerable repertoire of familiar morphemes.

Two important issues should be taken into consideration while studying the words using this model. First, calculating the DR of words presupposes that learning conditions for all words will be as equal as possible. For example, if a word is practised and encountered much more than the other words, it will normally be learned sooner, and will retain in LTM for a much longer period of time. Second, the L1, and L2 levels of participants would probably influence the validity and reliability of the model; it may decrease or even increase its reliability and validity. So, all the participants should necessarily have the same first language and L2 level as much as possible. The model includes the following 5 basic components:

1. Familiar prefixes or familiar initial morphemes with any number of syllables, and any structure have a DR of –1.
2. Familiar suffixes or familiar final morphemes with any number of syllables, and any structure have a DR of 0.
3. Every **closed** syllable (of unfamiliar morphemes), including at least VC (having the **coda)**, gets 2 DRs. (If the **coda** is the predictable /r/ sound which is not mainly pronounced in British English, that syllable should be treated as a syllable without the **coda**, taking 1 point of DR.)
4. Each ***open*** *syllable* (of unfamiliar morphemes), including (C)(C)(C)V (not having the