# From Thought to Speech: An Overview of Language Production from a Psycholinguistic Perspective

**Lukačević, Karla**

Sveučilište u Zadru

Odjel za anglistiku
Sveučilišni prijediplomski studij Anglistike

**Karla Lukačević**

# From Thought to Speech: An Overview of Language Production from a Psycholinguistic Perspective

**Završni rad**

Zadar, 2023.

Sveučilište u Zadru

Odjel za anglistiku
Sveučilišni prijediplomski studij Anglistike

# From Thought to Speech: An Overview of Language Production from a Psycholinguistic Perspective

Završni rad

Student/ica:

Karla Lukačević

Mentor/ica:

Dr. sc. Emilija Mustapić Malenica, v. lektorica

Komentor/ica:

Doc. dr. sc. Frane Malenica

Zadar, 2023.

# Izjava o akademskoj čestitosti

Ja, **Karla Lukačević,** ovime izjavljujem da je moj **završni** rad pod naslovom **From Thought to Speech: An Overview of Language Production from a Psycholinguistic Perspective** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 29. rujna 2023.

# Table of Contents

## 1. Introduction

Knowing how a word is produced and understanding the mechanism behind it is the basis for studying linguistics. The thesis comprises three main chapters. Explaining how linguistic information is stored in the mental lexicon is the topic of the first chapter. The second chapter discusses word selection and the mechanics of the “tip-of-the-tongue” phenomenon (TOT). There is a complex process behind each word spoken or written and various models of speech production were constructed in order to explain it. Two of those models will be analysed in this thesis, the first one being Levelt's model of speech production and the second one Paivio's Dual coding theory. The third main body chapter deals with those two aforementioned models of speech production; Levelt's model will be discussed in the chapter’s first five subchapters, while Paivio's Theory will be the main topic of the other two subchapters. This thesis deals with the multifaceted process of speech production and its phases in terms of psycholinguistics.

## 2. The Mental Lexicon

Frederiksen and Kroll (1976, p. 378) refer to the mental lexicon as a kind of storage of articulatory information which is part of one's linguistic knowledge, but they call it the internal lexicon. Frost (1998, p. 74) defines the concept of mental lexicon as the cognitive system's hypothesized structure in which a person's linguistic knowledge is contained. Nagy and Hiebert (2010, p. 392) call it "a representation of word knowledge in memory". Acha and Carreiras (2014, pp. 196-197) define it as the pool of words which are stored as part of one's cognitive system and which one uses for language comprehension. As can be seen, the definitions differ a little, but the main idea is there: the mental lexicon is a kind of information storage.

Acha and Carreiras (2014, p. 197) state that the mental lexicon is comprised of phonological and orthographic representations which ensure that a word is successfully identified, and which are formed from specific visual and auditory features. The code is first executed for visual orthographic input, which leads to the identification of letters and coding of their position in a word. The phonological computation process matches the activated letters with the suitable phonemes, which consequently connects the orthographic representation with the appropriate phonological representation. As described in Acha and Carreiras (2014, p. 198), two kinds of tasks have been used in order to study the time course of lexical access. The first task is lexical decision, in which the subjects are to determine whether a letter sequence shown to them on a screen is a word or a non-word, and their reaction times are measured in milliseconds to conclude which words can be more easily accessible. The second task is the reading aloud task, in which the subjects have to read presented words out loud. It is implied by both of them that the subjects have access to the mental lexicon. However, it is thought that the first task might primarily show the processes of lexical selection and decision, while the second task might do so with processes of phonological recoding and production (Acha, Carreira, 2014, p. 198). The results from a study conducted by Marslen-Wilson and Tyler (1981,

p. 326) show that it takes us around 450 milliseconds to conclude if a sequence of letters is a word. This suggests that speakers can quickly search through their mental lexicon in order to find a particular word (Aitchison, 1987, p. 7).

Ingram (2007, p. 179) states that the fact that the units of lexical representation are smaller than words is generally agreed upon, but the question arises as to how words are stored in the mental lexicon, specifically, in what form. Are words stripped of their inflectional and derivational affixes before being placed in the mental lexicon, or do they keep these elements and "enter" the mental lexicon as wholes? There have been some disputes between linguists and psycholinguists over some specifics of morphological decomposition and the units that make up lexical representations. As explained in Ingram (2007, p. 182), linguists claim that decomposition should be pursued into root morphemes of word forms (which is also referred to as maximal decomposition), while psycholinguists disagree, claiming that decomposition should be pursued only up to inflectional morphemes and the most productive derivational morphemes. The majority would agree that affix-stripping should apply to all inflectional suffixes (for example, *-s, -ing, -ed*). However, there is disagreement on how far decomposition should be pursued when it comes to derivational affixes, which is illustrated in the following example based on the one provided by Ingram (2007, p. 182). The word *replacement* can easily be divided into *replace*<verb> and *-ment*<noun-maker> and the same affix can be productively used to form other abstract nouns (such as *refreshment*, *disappointment*, *adjustment*). In contrast, not the same logic can be applied to words such as *apartment* and *environment*. Therefore, Ingram (2007, p. 182) argues that not all language (in this case English) users will be able to perceive even some simple morphological relations entailed in formation of these words, which suggests that this logic may not be applicable to derivational affixes. Aitchison (1987, p. 109) suggests a general rule for distinguishing these two kinds of affixation: one can add inflectional suffixes after derivational ones but not vice versa (*comput-er-s* but not

*compute-s-er*). She also claims that words mostly take inflectional affixes during the process of speaking, but some words already contain the inflectional suffix in the mental lexicon as they are mainly used in that form (for instance, the words *lips* and *peas*). As she states, when it comes to slips of the tongue, prefixed words and unprefixed words keep their beginnings just as often, and prefixed words are often used in place of unprefixed words. This suggests that prefixed words are not considered to belong to a different category in the mental lexicon. The following sentences (1) and (2) are taken from Aitchison (1987, p. 113) to exemplify the interchange of prefixed words with unprefixed ones:

(1) The emperor had many *porcupines* (concubines).

(2) Those lovely blue flowers - *concubines* (columbines).

The word *concubines* contains the prefix *con-* (although not everyone would recognise the prefix since it originates from Latin), while the words *porcupines* and *columbines* are unprefixed. This indicates that prefixes are attached, or fixed, onto their stems in the mental lexicon, and consequently, that words are stored as wholes (Aitchison, 1987, p. 116). This is also the case with suffixes, as can be seen in Aitchison and Straf (1982, p. 203), who provided a list of malapropisms in which suffixes usually remained attached to words. As defined by Aitchison (1987, p. 120), malapropisms are instances in which a word which sounds similar has been wrongly chosen instead of a target word. In the aforementioned list, among others, are words such as *deterrent* for 'detergent', *flaw* for 'flair' and *transitional* for ‘transistor’ (Aitchison and Straf, 1982, pp. 223-230). The suffix remains the same, but other parts of the word are changed. These links between suffixes and stems, and the fact that the suffixes remain attached to stems in malapropisms suggest that words are placed into the mental lexicon as wholes made of stems and affixes (Aitchison, 1987, pp. 115-116).

Aitchison (1987, p. 7) estimates that an educated adult is familiar with between 50,000 and 250,000 words. These high numbers as well as the fact that words can be located very fast

suggest that they are systematically organized in the mental lexicon. Marslen-Wilson and Tyler (1980, p. 26) proved that one can recognize a word in a sentence within 200 milliseconds from its onset. One can quickly distinguish words from non-words and locate the words in the mental lexicon when one needs them for speaking. However, the phases of the speech production process cannot be measured in the same way. Some have argued that it can be done because pauses in speech can be measured, and they often happen before large lexical items, but there is insufficient evidence to support this claim (Aitchison, 1987, p. 8). In Aitchison (1987, p. 9), the mental lexicon is referred to as the "human word-store" and the "mental dictionary", although words in the mental lexicon are not organised in alphabetical order. She proves this by supplying examples of slips of the tongue, such as "The doctor listened to her chest with his periscope", in which the word "stethoscope" is accidentally replaced by *periscope*. As can be seen, these two words, periscope and stethoscope, do not have the same initial letter; therefore, they are not listed next to each other in a dictionary. She claims that some other features of the word's sound structure are the key factors in organizing words in the mental lexicon. Those factors may include the word's stress pattern, ending and the location of the stress (Aitchison, 1987, p. 10). The content of the mental lexicon is not fixed, as one is able to constantly add new words (word entries, if we wish to extend the "dictionary" metaphor) and new meanings for already existing words. Furthermore, the mental lexicon keeps a greater amount of information about each word entry than a dictionary because a dictionary cannot possibly contain all of the meanings a particular word can carry and possible contexts a word can have. Fodor (1981, p. 287) discusses this issue by providing an example definition taken from a dictionary. The transitive verb *paint* means "to cover the surface of something with paint". However, as Fodor (1981, p. 287) states, one can knock over the paint bucket and cover the floor with paint, but we would not say that they have painted the floor. Stille et al. (2020, p. 6) provided a new view on the mental lexicon by producing their own neural model, which consists of two pathways:

the first one is a visual and auditory input pathway, while the second one is the production and articulatory output pathway. They connected the concept of the mental lexicon to Levelt's model of speech production (Levelt, 1989) by claiming that the mental lexicon is a part of long-term memory. As Stille et al. (2020, p. 6) explain, the input pathway produces a phonological form of a word by converting an auditory signal, a concept and a lemma, while the visual input directly transforms into a concept. The output pathway converts a concept into a lemma and a phonological form. Afterwards, motor plans become activated in order to produce speech.

## 3. Generating Words

The previous chapter provided an analysis of the mental lexicon, which acts as a source of linguistic knowledge. This knowledge is accessed in the process of word generation when it comes to word selection and the consequential activation of meaning-related words. As Griffin and Ferreira (2006, p. 23) mention, the speaker must first select a word which is best fitting in a given context. This is then followed by the activation of semantically related words which then have to compete to be selected. For the sake of simplicity, I will only provide an analysis of the generation of words; the generation of phrases and sentences will not be discussed in this thesis. The "tip-of-the-tongue" (TOT) phenomenon, described in the second subchapter, is a state in which a person is unable to remember a word which is ordinarily known to them, but they can remember similar words and some features of the word. Two kinds of recall are possible: complete and generic recall. The recall of the complete target word or the whole letter sequence is referred to as complete recall. On the other hand, generic recall happens when one can only recall word parts and attributes; it is available when a person cannot completely recall the word. I will begin the chapter with the analysis of the word selection process.

### 3.1 Word Selection

Kheder and Kaan (2019, p. 569) define lexical (or word) selection as the mechanism which is responsible for choosing a lexical item in the process of production or recognition and

which intends this item for further processing. According to Kheder and Khan (2019, p. 570), during lexical selection in bilinguals, the degree of semantic, orthographic and phonological similarity between the shared words (or cognates) in both languages affects the cognate effect and the resolution of competition. The word *cognate* refers to "words that look or sound similar in both languages" and this similarity, or dissimilarity can, for instance, slow down the naming process (Kheder, Khan, 2019, p. 570).

According to Griffin, Ferreira (2006, p. 23), speakers select words before assembling their sounds. This claim is largely based on two studies, one conducted by Fromkin (1971) and the other by Garrett (1975), in which they observed errors made during spontaneous speech. Fromkin (1971) analysed a list of more than six hundred speech errors which she collected over the course of three years. She wanted to find the reason behind these errors, and discover how they can explain the process of speech production. Fromkin (1971, p. 30) found that in most cases, speech errors in words and across word boundaries are realized through transposition (changed order of sound segments), substitution, omission, or addition of segments of the size of a phone. This can be seen on the following examples taken from Fromkin (1971, p. 30):

(3) also share → alsho share

(4) delayed auditory feedback → ... audif -auditory ...

(5) week long race → reek long race

(6) M-U values [ɛm juw væljuwz] → [ɛm vjuw] values

The first three errors, (3), (4) and (5), are examples of substitution, while the error (6) is an example of segment addition (Fromkin, 1971, p. 30). Garrett (1975) observed speech errors by focusing on syntactic variables and constraints on particular error types. Aside from the error realizations listed by Fromkin (1971), Garrett (1975, p. 138) found that speech errors can also be realized through, for instance, the following varieties: deletion, complex addition, complex

deletion, fusion and exchange. Examples of these varieties, taken from Garrett (1975, p. 138) are provided below in the corresponding order:

(7) “I’ll just get up and mutter -intelligibly.” (intended *unintelligibly)*

(8) “The one exPosner experiment that. . . .” (intended *Posner)*

(9) “That would be -having like Harry.” (intended *behaving)*

(10) “At the end of todays lection. . . .” (intended *lecture* or *lesson*)

(11) “Fancy getting your model renosed.” (intended nose *remodeled)*

These studies display that speech errors mostly include linguistic units “that can be most conservatively considered to correspond to whole words, morphemes, or individual speech sounds” (Griffin, Ferreira, 2006, p. 23).

During the word production process, the speaker must select a content word depending on what they wish to express. The intention of producing a word initiates the activation of a word family, whose words are semantically related (Griffin, Ferreira, 2006, p. 25). Meaning, or semantic representation of words, can be described through decompositional and non-decompositional views (Griffin, Ferreira, 2006, pp. 25-26; Bierwisch, Schreuder, 1992; Levelt et al., 1999). Decompositional views (as described in Griffin, Ferreira, 2006; Bierwisch, Schreuder, 1992) illustrate words as being made up of smaller entities that carry meaning; for example, the meaning of *dog*[1] can include *HAS A TAIL* and *BARKS*. However, some other words can share those features, and they become activated when the speaker activates those shared features. Consequently, words are similar to one another depending on how many semantic features they share; their similarity increases the more features they share. In contrast, non-decompositional views (described in Levelt et al., 1999) argue that the meanings of words and their representational bases have a one-to-one relationship, thus meaning that, for example,

[1] Words that refer to lexical entries or words are written in lowercase italics, while those which refer to semantic representations are marked by uppercase italics.

the word *dog* is supplied by an atomic meaning representation (often called *lexical concept*) of *DOG* (Griffin, Ferreira, 2006). Levelt et al. (1999) proposed the Weaver++ model to portray how multiple meanings become activated. The Weaver++ model is a computational model describing the course of lexical selection and word-form encoding using nodes to illustrate a word's syntax (Levelt et al., 1999, p. 6). According to Levelt et al. (1999, p. 11), if, for instance, the concept *DOG* becomes activated, the concept *CAT* is also activated because a semantic network connects the concept *DOG* to the concept *ANIMAL*, which will, consequently, activate the concept *CAT*. Finally, the concept *CAT* will then activate the word *cat*, and other words similar in meaning will also become active.

During word generation, a speaker must choose the most suitable word to use in a given context. This word is often referred to as a target or intended word (Griffin, Ferreira, 2006, p. 27). At the same time, other words, contextually similar in meaning, also compete to be selected. Dell et al. (1997, p. 810) found semantically related errors to be the most common cause of error appearing in word selection. To specify, this happens when a speaker chooses a semantically similar word instead of the target word (for example, calling an apple *pear*). In a study conducted by Cutting and Ferreira (1999), it was proved that only words which are semantically similar to the target word compete for selection, while words associated with the target word, that is, words that are only related in meaning, do not tend to do this. In the same example of calling an apple *pear*, the word *tree* would not compete for selection, although it is related to the word *apple*. Furthermore, Griffin and Ferreira (2006, p. 28) state that semantically unrelated words do not interfere with word production the same way semantically related words do. They claim that the lexical representations of semantically related distractor words are activated both through distractors themselves and through their semantic relation to the target word. In contrast, the representations of semantically unrelated distractor words receive activation solely by distractors. I will illustrate this by using my own example which is based

on the one provided by Griffin, Ferreira (2006, p. 29). If one were to name a picture of an apple, this would lead to the activation of the lexical representation of *pear* by the distractor word *pear* and by the semantic representation of *APPLE*, that is, through *pear's* semantic relation to *apple*. In contrast, the distractor word *car* would be the sole trigger for the activation of the lexical representation of *car*. However, Noteboom (1969, p. 155) deduces that "a mistakenly selected word always or nearly always belongs to the same word class as the intended word". According to Noteboom (1969, p. 155), a phrase's grammatical structure dictates which words can be selected in that particular situation. Brehm (2023, p. 8) also highlights the same fact that the word class usually remains the same in both target words and errors, and implies that all of the linguistic entities on which speech errors occur, that is, phonemes, sounds, lemmas, verbs, nouns, vowels and consonants, are "psychologically real[2], decomposable parts of language that are used in processing".

### 3.2 The "Tip-of-the-Tongue" Phenomenon

According to Griffin, Ferreira (2006, p. 33), the name of this phenomenon comes from the expression "to have a word on the tip of one's tongue". Abrams and Davis (2016, p. 13) define the TOTs as "temporary word-finding problems, characterized by an inability to retrieve a word at an intended time despite a strong feeling of knowing the word" and refer to it as a "production failure". Brown and McNeill (1966, p. 325) define the "tip-of-the-tongue" (TOT) phenomenon as a state in which a person cannot remember a word which is well known to them, but they can remember structurally and semantically similar words. A vivid description of the state is provided in Brown, McNeill (1966, p. 326): "[a subject] was, in fact, "seized" by a TOT state. The signs of it were unmistakable; he would appear to be in mild torment, something like the brink of a sneeze, and if he found the word his relief was considerable". Brown and McNeill

[2] Smith (2004, p. 98) refers to the notion of psychological reality as an indication that a theory is true or correct.

were generally the first to conduct experiments in order to study the phenomenon. In their 1966 experiment, the aim was to study the phenomenon and investigate its frequency and the accuracy of generic recall. The study was conducted on fifty-six undergraduate students who were asked to write down words which match definitions they hear. They used low frequency[3] words such as *nepotism* and *apse*. The students were to write down the following: if the word they were thinking of was the target word, the number of syllables, the first letter of the word they were thinking of, phonologically and semantically similar words, and the word they were thinking of. They recorded 360 cases of TOT in total, 233 of which were positive TOTs. Positive TOTs are those in which the target word matches the word a person was thinking of, while in negative TOTs these two words are not the same; a person was thinking of a different word and not the target word. They found that similar-sounding words matched the number of syllables of the target word in 48% of instances, while words similar in meaning consisted of the same number of syllables in 20% of cases. The subjects successfully guessed the initial letter of the intended word in 57% of instances of positive TOTs, and the first letters in similar-sounding words were the same as the ones in target words in 49% of all cases. Words similar in meaning had matching first letters with target words in 8% of instances. These results show that one can correctly guess the number of syllables and the first letter of the target word when they are in the TOT state. These pieces of information become more accurate the closer a speaker is to successfully recall a word, and also if the word they are thinking of sounds similar to the target word, rather than if the words have similar meanings. Furthermore, Brown and McNeill (1966, p. 326) differentiate between complete and generic recall. Complete recall is the recall of the target word as a whole, or the complete letter sequence. In contrast, generic recall is the recall of word parts and attributes. When a person cannot recall the whole word,

[3] According to a study conducted by Astell and Harley (1996, p. 204), target words with lower imageability and frequency have a higher chance of inducing the TOT state than those with higher imageability and frequency.

generic recall may be successful. Accurate generic recall is probable if complete recall of a word has not yet happened but is "felt to be imminent" (Brown, McNeill, 1966, p. 333). Generic recall exists in two varieties: partial recall and abstract form recall (Brown, McNeill, 1966, p. 326). Partial recall is the recall of only a part of the target word, that is, its letter, affix or syllable, while abstract form recall means recalling the number of syllables of the target word or the location of the primary stress. The resolution of TOT can be delayed by alternates, or words that come to mind and are phonologically and grammatically similar to the target word (Burke et al., 1991, p. 568). These words reduce the rate of recall, which consequently slows the resolution of TOT. Levelt (1989, pp. 320-321) claims that the lemma of the target word is available on semantic grounds, but the speaker is unable to completely access form information; the issue arises when it comes to the transition from the lemma to the sound form.

Metcalfe et al. (2017) performed an experiment to find if the "tip-of-the-tongue" phenomenon and curiosity correlated. The experiment was performed on forty-six undergraduate students by presenting them with general information questions, for instance, "What is the name of the ancient warrior who was dipped in the River Styx?" (Metcalfe et al., 2017, p. 3). The students were instructed to attempt to answer questions and while doing so, to state whether or not they were in the TOT state, and whether or not they wished to see the answer later. However, they would only be available to see the answers of up to ten questions. The questions were preceeded by four-digit numbers half of the time, and the other half by asterisks. The subjects were asked to recall the numbers. Metcalfe et al. (2017, p. 4) found that those participants who were in the TOT state wished to see the answers more than those who were not affected by the state, and also that some participants claimed to have been in a TOT state, although they had answered the question correctly. Since the examiners have not provided students with feedback about correctness of their answers, this perhaps caused uncertainty in some students, which is why they were convinced they were in the TOT state (Metcalfe et al.,

2017, p. 6). It may be assumed that those who were thinking of a different word than the target word feel greater relief after hearing the target word than those who feel they are in the TOT state, but have the correct answer. However, there is virtually no evidence to support this hypothesis, thus further research should be conducted to confirm it.

## 4. Models of Speech Production

In this chapter, I am going to offer a comprehensive overview of two chosen models of speech production, illustrate their processing systems using schemas and tables, and explain the relevance of elements involved in two theories of language production. Willem Levelt's model of speech production assumes that language production proceeds in three main phases, which are accompanied by various subprocesses, and that these phases require inputs and outputs for successful execution of utterances. The second model which will be presented is Paivio's Dual coding theory, or DCT, which argues that two interconnected, but functionally independent processing systems are responsible for production of speech. The thesis will display the mechanism behind these processing systems of DCT, and make a connection with sensorimotor systems. Levelt's model of speech production is the first to be analysed.

### 4.1 Levelt's Model of Speech Production

According to Willem Levelt (1989), the process of language production is based on three phases: conceptualization, formulation and articulation. As illustrated in Figure 1, language production begins with conceptualization, described as a phase in which the speaker decides what to say, chooses relevant information from their knowledge store and orders its expression, and afterwards monitors what they are saying. In formulation, grammatical and phonological encoding translate the conceptual information into a linguistic structure. Finally, articulation is the phase of the motor execution of the phonetic plan, which ensues after formulation. *Overt speech* is the result of articulation.

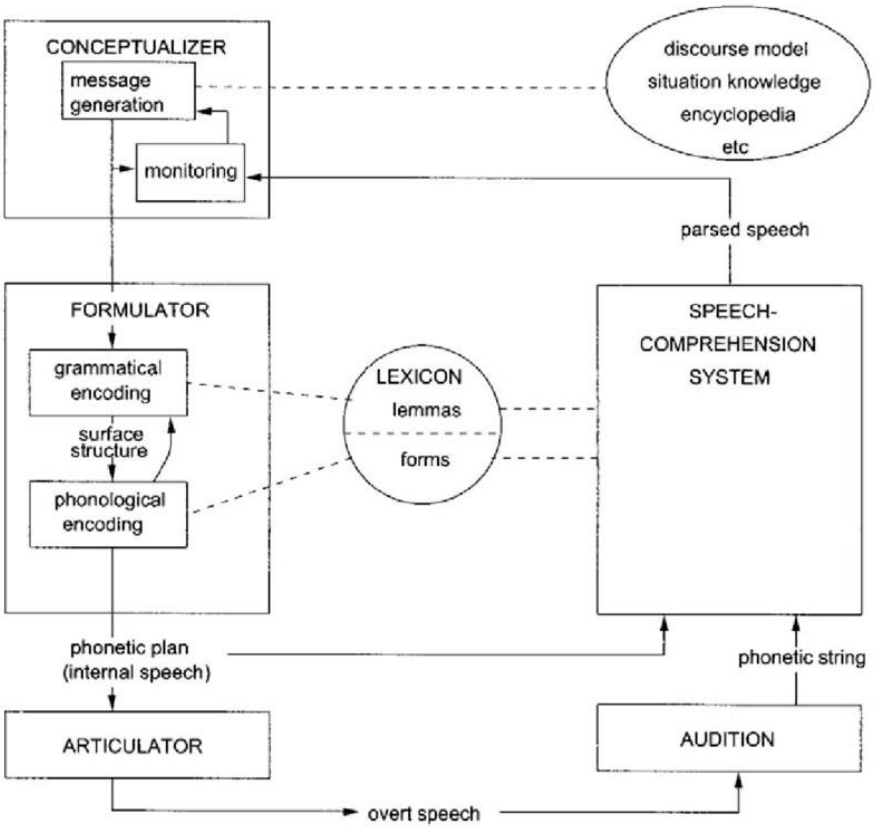


**Figure 1.** Levelt's Model of Speech Production (Levelt, 1989, p. 9)

### 4.1.1 Conceptualization

The process of talking begins with an intention. Setting an intention marks the beginning of the process of conceptualization, a set of mental activities whose outcome is the *preverbal message*, or the proposition, which is based on our knowledge. The processing system responsible for the successful execution of conceptualization is occasionally called the *Conceptualizer* (Levelt, 1989, p. 9). Levelt (1989, p. 11) distinguishes two kinds of processes involved in conceptualization: microplanning and macroplanning. Macroplanning is the stage of developing a communicative goal into a string of subgoals and retrieving the information needed to realize those subgoals. Microplanning aims to assign the right propositional shape to these pieces of information and provide the topic and focus of the utterance.

The speaker uses two types of knowledge in order to linguistically encode conceptual information: *procedural* and *declarative* knowledge. Procedural knowledge is in the format IF X THEN Y. To illustrate, Levelt (1989, p. 10) uses the following sentence as an example: "IF the intention is to commit oneself to the truth of p, THEN assert p." In this case, *p* represents a proposition and the speaker's aim is to express it as being true by asserting that proposition. Working memory, sometimes called a *buffer* (Ingram, 2007, p. 267), serves as a repository of information obtained by the Conceptualizer and currently available to the speaker. Declarative knowledge is the second type of knowledge used by the speaker; it is available in *Long-Term Memory*, the speaker's knowledge acquired throughout his life, and as declarative knowledge in the moment of speaking. In Figure 1, the ellipse and the circle actually represent declarative knowledge, while procedural knowledge is within the processors represented by rectangles. Messages generated by each of the three stages of language production serve both as the output of the preceding stage and as the input of the following one. For instance, the same message generated by the Conceptualizer acts both as its output, or *preverbal* message, and as the input of the Formulator.

### 4.1.2 Formulation

The Formulator's task is to encode the *chunks* of conceptual information into a linguistic structure. Therefore, in this case, the conceptual information is the input, and the output, that is, the product of the Formulator, is a *phonetic* or *articulatory plan*. During formulation, the preverbal message is shaped by two subprocesses: *grammatical* and *phonological encoding*. Grammatical encoding starts with accessing lemmas, stored in the speaker's mental lexicon. Bock and Levelt (1994, p. 951) argue that a lemma is a word as a syntactic entity and that lemmas are directly opposed to lexemes, which reflect the word's morphological and phonological form. Lemmas are activated when their semantics partially match the preverbal message, which then makes its syntax available. As a result, phrases and clauses are built based

on the syntactic category of the lemma; for instance, if a lemma is a noun, then the *surface structure*, or the product of grammatical encoding, will become a noun phrase (Levelt, 1989, p. 11). For instance, the lemma *beautiful* is in the network model of lexical access[4] linked to the adjective node. However, when observed on the lexeme level, the word *beautiful* is polymorphemic as it is formed from two morphemes (a stem and a suffix), and it comprises seven phonological segments, /b/, /j/, /uː/, /t/, /ɪ/, /f/, /l/.

Phonological encoding is the phase in which the surface structure of the lemmas, which resulted from grammatical encoding, converts into a phonetic or articulatory plan for each of the lemmas and for the whole utterance. This is done by seeking information from the *lexical form* about the item's morphological and phonological structure. To illustrate, the word *beautiful* is comprised of a root (beauty) and a suffix (-ful), and it contains three syllables, the first of which is stressed. *Phonetic plan* can sometimes be referred to as *internal speech*, although they are not synonymous. As Levelt (1989, p. 12) states, *internal speech* can have the same meaning as the *phonetic plan* when it is *parsed*, or interpreted, by the speaker.

### 4.1.3 Articulation

Articulation, or articulating, represents the "execution of the phonetic plan" (Levelt, 1989, p. 12), which simultaneously becomes both the output of the process of formulation and the input of articulation. The Articulator operates by coordinating the work of the laryngeal, supralaryngeal and respiratory muscles. Because internal speech is formed before the process of articulation, it needs temporary storage; the storage space called the *Articulatory Buffer* acts as information reserves and allows the Articulator access when it is needed. The result of articulation is *overt speech.*

[4] The network model of lexical access is a model representing a system of word relations organized on three levels: the conceptual, the lemma and the lexeme level. Nodes are representations of lexical concepts (Bock, Levelt, 1994).

### 4.1.4 Self-Monitoring

As we listen to ourselves speaking, we can notice dysfluencies or errors in our overt speech. Self-monitoring is a mechanism whose role is to enable the speaker to monitor their own speech and if necessary, correct it (Levelt, 1989, p. 459). In order to correct ourselves, we must first notice the mistake, and then interrupt our own speech which consequently interrupts the interlocutor's interpretation of the ongoing speech. To do this, we often use words and phrases such as *er*, *I mean*, *well* and *no*, which Levelt (1989, p. 459) calls *editing expressions*. According to Levelt (1983, p. 41), noticing the mistake and interrupting the ongoing speech are the first two phases of self-repairing in speech, while the third and final phase is "making the repair proper", i.e., correcting ourselves. Levelt (1983) analyses sentences from a corpus of 959 repaired utterances, which were recorded in adult Dutch speakers, to determine the mechanism behind self-repairing. The subjects were presented with visual patterns made of coloured dots and connected with lines drawn at right angles, and were asked to describe these patterns in such a way that would enable the listener to draw the pattern from the description, provided they had seen some examples of these patterns.

(12) Go from left again to, uh . . . , from pink again to blue (Levelt, 1983, p. 44)

The example (12) provided above represents the most common type of repair, which consists of three parts: the original utterance, the editing phase and the repair proper. In this case, the original utterance is *Go from left again to*, and *left* is considered to be *reparandum*, or the spot that needs to be repaired. The moment of interruption, *again to*, appears three syllables after the reparandum; this value (in this case 3 syllables) is referred to as the *delay of interruption*. The editing phase is marked by a period of hesitation which can be signified by editing expressions or *editing terms*, as Levelt (1983, p. 44) calls them. The repair proper, as defined by Levelt (1983, p. 44), is the variant of an incorrect original utterance, now without mistakes. In the provided example, the repair proper is *from pink again to blue*. The question arises as to

how the speaker manages to detect the mistake in their utterance, and Levelt (1983, p. 46) offers two possible answers. The first possibility is that the speaker can directly access components of the speech production process, which implies that the outputs of the components have to meet certain criteria. In contrast, the second possibility suggests that the components are not accessible to the speaker; he can only access the final result of speech production, which is parsed and compared with the speaker's original intent. This is known as *the perceptual theory of monitoring*, while the first possibility is called *the production theory of monitoring*. Levelt prefers the second of the two theories, and the validity of his choice was confirmed by the results of the analysis of colour repairs, or cases in which the speaker realized he had made a mistake and consequently replaced one colour name with another one. Essentially, Levelt and his associates (1983, pp. 59-60) wanted to find the correlation between different positions in the constituent, and the chance of detecting errors. They found that error detection increases towards the end of the constituent. The Main Interruption Rule should be mentioned here, It is a thesis that the speaker stops speaking instantly after he notices an error and decides to repair it while simultaneously ignoring the linguistic structure of the utterance (Levelt, 1983, p. 56). This means that the speaker can interrupt his speech at any moment, before, after or during overt production of the reparandum. The example (13) below provides an illustration of the reparandum interruption.

(13) We can straight on to the ye.. , to the orange node (Levelt, 1983, p. 56)

**4.2 Dual Coding Theory**

Allan Paivio first proposed his Dual coding theory in 1971, according to which human cognition is based on two "functionally independent but interconnected multimodal systems" (Paivio, 2010, p. 207). He calls them internalized verbal and nonverbal processing systems. The two systems are regarded as *symbolic systems*, and the theory also assumes the existence of *sensorimotor systems* which exist orthogonally to symbolic systems. The latter claim refers to

the qualitative difference between the subsets of mental representations within the two mental codes; this is due to the fact that they originate from different sensory experiences, as explained by Sadoski and Paivio (2013, p. 887). The symbolic systems are therefore thought to be developed from sensorimotor systems, while keeping their functional properties (Paivio, 1986, p. 57). Table 1 exemplifies each system's information specified by their modalities. The right side of the table shows the difference between the verbal and nonverbal internal representations, while the left side contains the sensorimotor modalities from which the former are developed. As shown in Table 1, both words and objects have visual, haptic and auditory representations, but words lack olfactory, affective and gustatory representations, as those are solely activated by nonverbal processes. To illustrate, humans create visual representations in the verbal code for linguistic units that they have seen (for instance, words or phrases), and in the nonverbal code for non-linguistic objects. They create auditory and haptic representations in a similar way, following the same logic. On the other hand, they do not produce verbal representations for gustatory and olfactory modalities, typically only nonverbal representations, but they can produce verbal representations for emotional states, although they are by definition nonverbal (Sadoski & Paivio, 2013, 887). According to Paivio (1991, p. 257), olfactory, gustatory and affective modalities are nonverbal, because linguistic symbols cannot be developed from smells, tastes and emotions. However, according to Paivio (2010, p. 208), words *can* activate memories of events related to a certain taste or smell.

**Table 1.** A Visual Representation of the Orthogonal Conceptual Connection between Symbolic Systems and Sensorimotor Systems (adapted from Paivio, 2010, p. 208)

| | Symbolic Systems | |
|---|---|---|
| **Sensorimotor** | **Verbal** | **Nonverbal** |
| Visual | Visual words | Visual objects |
| Auditory | Auditory words | Environmental sounds |
| Haptic | Writing patterns | "Feel" of objects |
| Taste | - | Taste memories |
| Smell | - | Olfactory memories |
| Emotion | - | Affective reactions |

Furthermore, the theory suggests that an internalized verbal processing system deals with language, or linguistic stimuli, while an internalized nonverbal processing system is tasked with processing information regarding nonverbal content, as explained by Paivio (1986, p. 53). These processing systems are functionally independent, meaning they can be activated separately, or they can both be active at the same time. Furthermore, they are functionally independent in the stages of processing information; however, it is usually thought that the activity in one system is triggered by the activity in the other one. This leads to the idea that the systems' representations are interconnected (Paivio, 1986, p. 62), as exemplified in Figure 2. According to Paivio (2010, p. 208), functional independence also incorporates the idea that different sensorimotor systems, for instance, the auditory speech system, can operate independently within nonverbal or verbal systems. Studies have shown that concrete and abstract words are not processed equally; while the verbal processing system processes both the abstract and the concrete words in the same way, the nonverbal system mainly processes, or encodes, concrete words, as they evoke imagery quicker than abstract words (Paivio, 2010, pp. 220-221). Figure 2 depicts the mechanism behind the verbal and nonverbal processing systems. The verbal processing system requires verbal stimuli as an input and produces verbal responses as an output, while its representational unit is called a logogen. The scheme exemplifies three

types of logogens: visual, auditory and motor, while there are also haptic logogens. In this case, *phone* is the visual logogen, *ring* is the auditory logogen, and *gossip* is the motor logogen. Similarly, the input of the nonverbal processing system (also called the imagery processing system) are nonverbal stimuli, and the processes involving imagens, i.e., the system's representational units, result in nonverbal responses. Based on the same example shown in Figure 2, the imagens of *telephone* are the visualization and the "feel" of the object telephone, as well as the sound of ringing.

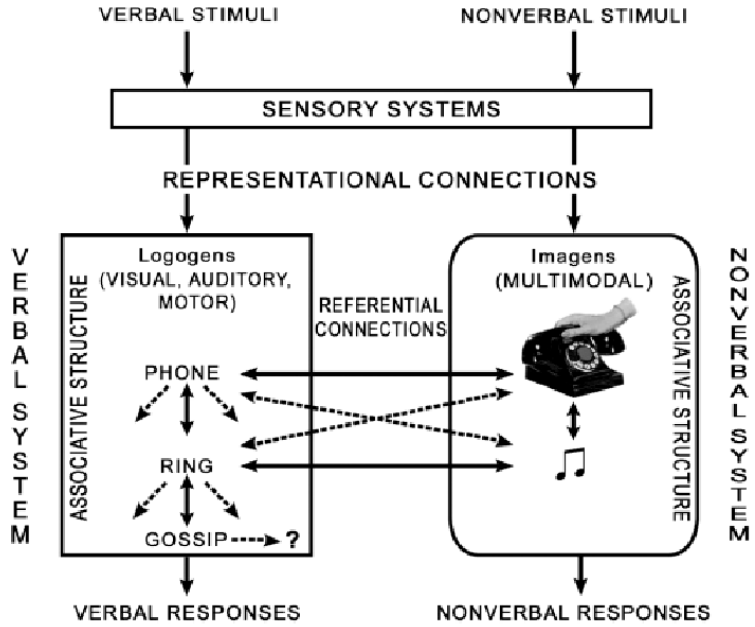


**Figure 2.** A Schematic Representation of Verbal and Nonverbal Processing Systems Based on the Object Telephone (Paivio, 2010, p. 209)

### 4.2.1 Levels of Processing

Paivio (1986, p. 69) identifies three levels of processing: *representational*, *referential* and *associative* (Figure 2). Representational processing begins when verbal representations become activated by verbal stimuli, and nonverbal representations by nonverbal stimuli. The activation is relatively direct because the perceptual analysis of verbal stimuli of different modalities faces complications, which is why reading printed words takes more time than

repeating spoken words. Referential processing is the level on which the verbal system is activated by non-linguistic stimuli or the nonverbal system is activated by linguistic stimuli. It is considered indirect because the activity is transmitted from one processing system to the other. Sadoski et al. (2000, p. 85) mention the importance of familiarity on this processing level. Associative processing is the level of processing at which representations within a processing system are activated by other representations belonging to the same system; for instance, nonverbal situations can activate nonverbal memories. I will attempt to illustrate these processes by examining sentence (14) based on the analysis provided by Mustapić Malenica (2021, pp. 42-43).

(14) A dark-haired boy is picking fresh strawberries.

The sentence above acts as input to a listener or an interlocutor, and the verbal code is translated into nonverbal code. This results in creating mental representations of the words *boy* and *strawberries*. The verbal code is activated once the information is processed, and it is translated into output. As previously stated, associative processing includes the activation of representations related to other representations of the same system. In this case, the word boy can activate other semantically similar words such as *girl* and *toy*.

When discussing the complex process of language production, the terms concreteness and abstractness must be mentioned. Concreteness is an abstract term which is of relevance in psychology and linguistics; it is used to explain some mechanisms behind memory stores and speech production. It refers to the degree to which the idea expressed by a certain word can be perceived using one's senses (Brysbaert et al., 2013). Paivio (1971, p. 16) defines the term abstractness as "the directness with which the stimulus denotes particular objects or events". Hayakawa (1949, p. 167) uses the metaphor "abstraction ladder" to illustrate his claim that the concept of abstraction is organized as levels, with higher levels referring to objects with the least reference to the features of a particular object.

Paivio (1991, pp. 260-261) emphasizes the connection the so-called *conceptual-peg hypothesis* has with the terms concreteness and abstractness. The hypothesis served as the basis for the dual coding theory as it emerged from the *one-bun, two-shoe* rhyming mnemonic technique, which was used to remember lengthy lists of items. The principle behind it can be illustrated by taking the word *table* as an example. If we wanted to remember this item, we could connect it with the peg word *bun* and consequently create a nonverbal image of a table inside a sandwich bun. During recall, we remember the generated image, retrieve the component *table,* and decode it back into the word. The reason why the method is successful is due to concreteness of the words used. It is assumed that nouns are more concrete than adjectives, which is why they are quicker at inducing recall. Paivio (1971, p. 359) finds that the frequency of imagery use is proportional to concreteness.

To conclude, I have provided an outline of both Levelt's model of speech production and Paivio's dual coding theory, and now I will examine the main differences between the two. Unlike Levelt's model whose single processing system carries the entire process of language production, Paivio's model relies on *two* multimodal systems. Levelt's processing system comprises conceptualization, formulation and articulation as its three main phases while it also requires grammatical and phonological encoding to build a linguistic structure. Paivio's theory, on the other hand, involves verbal and nonverbal systems, while their representational units are called logogens and imagens. Also, sensorimotor modalities provide various information to these systems, for instance, knowledge based on something the speaker has heard. To simplify, Paivio claims that in speech production, one processing system is required to process language, while the other processes non-linguistic stimuli. This kind of task division does not exist as part of Levelt's model, and neither does the notion of sensorimotor information.

## 5. Conclusion

To summarise, the first main chapter of the thesis analyses the notion of the mental lexicon as a storage of linguistic knowledge which is not fixed as a dictionary but can always accept new words and meanings. In the second main chapter I have discussed the process of word selection, explained the meaning behind compositional and decompositional views and finally, examined the "tip-of-the-tongue" phenomenon. Decompositional views assume that words are made up of smaller entities that have meaning, while non-decompositional views argue that meanings of words and their representational bases share a one-to-one relationship. The "tip-of-the-tongue" state is one in which a person is unable to recall a word which they are familiar with, however, they can remember structurally and semantically similar words. Furthermore, speech production is a complex process which is usually described through various models. I have provided the analysis for two models of speech production in the third main chapter. Firstly, I have examined Levelt's model of speech production and explained the processes conceptualization, formulation and articulation, as well as the process of self-monitoring. In conceptualization, the speaker decides what to say, chooses relevant information and orders the information for expression. In formulation, the conceptual information is becomes a linguistic structure, while in articulation the phonetic plan is executed. Self-monitoring enables the speaker to monitor and correct their own speech. Secondly, I have defined Paivio's Dual coding theory and analysed its multimodality portrayed through verbal and nonverbal processing systems. Paivio's Dual coding theory is based on verbal and nonverbal processing systems, which function independently, but they are interconnected. I have provided figures and tables in order to demonstrate the models and the flow of speech production. To conclude, this thesis has provided a detailed analysis of the speech production process, and I intend to do more research on some of the mentioned concepts and terms, as I wish to know more and enrich my mental lexicon.

**6. Works Cited**

Abrams, L.,& and Davis, D. K. (2016). The Tip-of-the-Tongue Phenomenon: Who, what, and why. In Wright, H. H. (Ed.), *Cognition, Language, and Aging* (pp. 13-54). John Benjamins Publishing Company.

Acha, J., & Carreiras, M. (2014). Exploring the Mental Lexicon: A Methodological Approach to Understanding how Printed Words are Represented in our Minds. *The Mental Lexicon*, 9(2), 196–231. https://doi.org/10.1075/ml.9.2.03ach

Aitchison, J. (1987). *Words in the Mind: An Introduction to the Mental Lexicon.* Basil Blackwell Ltd.

Aitchison, J., & Straf, M. (1982). Lexical Storage and Retrieval: a Developing Skill. In Cutler, A. (Ed.), *Slips of the Tongue and Language Production* (197-241). Mouton.

Astell, A. J., & Harley, T. A. (1996). Tip-of-the-Tongue States and Lexical Access in Dementia. *Brain and Language*, 54, 196–215. https://doi.org/10.1006/brln.1996.0071

Bierwisch, M., & Schreuder, R. (1992). From Concepts to Lexical Items. *Cognition*, 42, 23-60. https://doi.org/10.1016/0010-0277(92)90039-K

Bock, K., & Levelt, W. J. M. (1994). Language Production. Grammatical Encoding. In Traxler, M. J., & Gernsbacher, M. A. (Eds), *Handbook of Psycholinguistics* (pp. 945-984). Academic Press Inc.

Brehm, L. (2023). What's an Error anyway? Speaker- and Listener-Centered Approaches to Studying Language Errors. In Federmeier, K. D., Montag, J. L. (Eds.), *Psychology of Learning and Motivation* (pp. 1-39). Academic Press.

Brown, R., & McNeill, D. (1966). The "Tip of the Tongue" Phenomenon. *Journal of Verbal Learning and Verbal Behaviour*, 5, 325-337. https://doi.org/10.1016/S0022-5371(66)80040-3

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas. Behavior Research Methods, 45(3). https://doi.org/10.3758/s13428-013-0403-5

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the Tip of the Tongue: What Causes Word Finding Failures in Young and Older Adults? Journal of Memory and Language, 30, 542-579. https://doi.org/10.1016/0749-596X(91)90026-G

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and Phonological Information Flow in the Production Lexicon. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2), 318-344. https://doi.org/10.1037/0278-7393.25.2.318

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and Sound: Approaches to the Internal Lexicon. Journal of Experimental Psychology: Human Perception and Performance, 2(3), 361-379. https://doi.org/10.1037/0096-1523.2.3.361

Frost, R. (1998). Toward a Strong Phonological Theory of Visual Word Recognition: True Issues and False Trails. *Psychological Bulletin*, 123(1),71-99. https://doi.org/10.1037/0033-2909.123.1.71

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of Spoken Language Production. In Traxler, M. J., & Gernsbacher, M. A. (Eds.), *Handbook of Psycholinguistics* (pp. 21-59). Academic Press Inc.

Hayakawa, S. I. (1949). *Language in Thought and Action*. George Allen & Unwin Ltd.

Ingram, John C. L. (2007). *Neurolinguistics. An Introduction to Spoken Language Processing and its Disorders.* Austin, P., Bresnan, J., Comrie, B., Crain, S., Dressler, W., Ewen, C., Lass, R., Lightfoot, D., Rice, K., Roberts, I., Romaine, S., & Smith, N. V. (Eds.). Cambridge University Press.

Kheder, S., & Kaan, E. (2019). Lexical Selection, Cross-Language Interaction, and Switch Costs in Habitually Codeswitching Bilinguals. *Bilingualism: Language and Cognition,* 22(3), 569–589. https://doi.org/10.1017/S1366728918000500

Levelt, W. J. M. (1983). Monitoring and Self-Repair in Speech. *Cognition,* 14, 41-104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Joshi, A. (Ed.). Halliday Lithograph.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences*, 22, 1-75.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, l-71. https://doi.org/10.1016/0010-0277(80)90015-3

Marslen-Wilson, W., & Tyler, L. K. (1981). Central Processes in Speech Understanding. *Philosophical Transactions of the Royal Society*, 295, 317-332. https://doi.org/10.1098/rstb.1981.0143

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications*, 2(31). https://doi.org/10.1186/s41235-017-0065-4

Mustapić Malenica, E. (2021). *Uloga Koverbalnih Gesta u Jezičnoj Recepciji* [Doctoral dissertation, University of Zagreb, Faculty of Humanities and Social Sciences]. ODRAZ. https://doi.org/10.17234/diss.2021.8542

Nagy, W. E., & Hiebert, E. H. (2010). Toward a Theory of Word Selection. In Kamil, M. L., Pearson, P. D., Moje, E. B., & Afflerbach, P. (Eds.), *Handbook of Reading Research, Volume IV* (388-404). Routledge.

Noteboom, S. (1973). The Tongue Slips into Patterns. In A. G. Sciarone, Essen, A. J., & Raad, A. A. (Eds.), *Leyden Studies in Linguistics and Phonetics* (pp. 144-156). Mouton.

Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart and Winston, Inc.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Broadbent, D. E., McGaugh, J. L., Mackintosh, N. J., Posner, M. I., Tulving, E., & Weiskrantz, L. (Eds.). Oxford University Press. (https://archive.org/details/mentalrepresenta0000paiv/mode/2up)

Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology, 45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A. (2010). Dual Coding Theory and the Mental Lexicon. *The Mental Lexicon*, 5(2), 205–230. https://doi.org/10.1075/ml.5.2.04pai

Sadoski, M., & Paivio, A. (2013). A Dual Coding Theoretical Model of Reading. In R. R. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading* (5th ed., pp. 1329–1362). International Reading Association.

Sadoski, M., Goetz, E. T., & Rodriguez, M. (2000). Engaging Texts: Effects of Concreteness on Comprehensibility, Interest, and Recall in Four Text Types. *Journal of Educational Psychology*, 92(1), 85-95. https://doi.org/10.1037/0022-0663.92.1.85

Smith, N. (2004). *Chomsky: Ideas and Ideals.* Cambridge University Press.

Stille, C. M., Bekolay, T., Blouw, P.,& Kröger, B. J. (2020). Modeling the Mental Lexicon as Partof Long-Term and Working Memory and Simulating Lexical Access in a Naming Task Including Semantic and Phonological Cues. *Frontiers in Psychology*, 11(1594). https://doi.org/10.3389/fpsyg.2020.01594

# From Thought to Speech: An Overview of Language Production from a Psycholinguistic Perspective

**Abstract**

Language is fundamental for successful human communication, whose execution relies on access to the necessary data in the mental lexicon. The mental lexicon is a storage or structure which contains linguistic information. Words are stored in the mental lexicon as wholes consisting of stems and appropriate affixes. Before speaking, a speaker must first choose the most suitable word in a particular context, which leads to the activation of other similar words and the competition among them. The "tip-of-the-tongue" phenomenon, illustrated in the second main body chapter, is a state described as the unability to recall a word which is usually known to a person; however, they can remember other words, similar in structure and meaning. The mechanics of speech production were examined and explained through a number of models, two of which are analysed in this thesis. The first one, Levelt's model of speech production, argues that language is produced in three phases: conceptualization, formulation and articulation, all of which are part of the same processing system. On the other hand, Paivio's Dual coding theory assumes that language is produced as a result of verbal and nonverbal processing systems, which are functionally independent and interconnected. In Dual coding theory, representational units of the verbal processing system are called imagens, while logogens are the representational units belonging to the nonverbal processing system.

**Key words:** speech, mental lexicon, the "tip-of-the-tongue" phenomenon, Levelt's model, conceptualization, formulation, articulation, Paivio's Dual coding theory, verbal and nonverbal processing systems, imagen, logogen

Od misli do govora: pregled jezične proizvodnje iz psiholingvističke perspektive

**Sažetak**

Jezik je ključan za uspješnu komunikaciju ljudi, čije se ostvarenje oslanja na pristup potrebnim podatcima u mentalnom leksikonu. Mentalni je leksikon spremnik ili struktura koja sadrži lingvističke informacije. Riječi su u njemu spremljene kao cjeline sastavljene od baza i prikladnih afiksa. Prije govorenja, govornik prvo mora izabrati riječ koja je najprikladnija u određenom kontekstu, što dovodi do aktivacije sličnih riječi i natjecanja (eng. *competition*) između njih. Fenomen „na vrhu jezika", opisan u drugom poglavlju glavnog dijela, stanje je opisano kao nemogućnost prisjećanja riječi koja je osobi obično poznata, ali se ona može sjetiti drugih riječi, sličnih po strukturi i značenju. Mehanizam govorne proizvodnje proučili su i objasnili brojni modeli, od kojih su dva analizirana u ovome radu. Prvi od njih, Leveltov model govorne proizvodnje, tvrdi da se jezik proizvodi u tri faze: konceptualizacija, formulacija i artikulacija, a sve pripadaju istom sustavu procesiranja. S druge strane, Paivijeva Teorija dvostrukog kodiranja pretpostavlja da je jezik proizvod verbalnog i neverbalnog sustava procesiranja koji su funkcionalno neovisni i međusobno povezani. U Teoriji dvostrukog kodiranja, reprezentacijske jedinice sustava verbalnog procesiranja zovu se imageni, dok su logogeni reprezentacijske jedinice neverbalnog sustava procesiranja. Paivio također prepoznaje tri razine procesiranja: reprezentacijska, referentna i asocijativna.

**Ključne riječi:** govor, mentalni leksikon, fenomen „na vrhu jezika", Leveltov model, konceptualizacija, formulacija, artikulacija, Paivijeva Teorija dvostrukog kodiranja, verbalni i neverbalni sustav procesiranja, imagen, logogen