МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра англійської філології та міжкультурної комунікації

"EGGCORNS" AS A DISTINCT TYPE OF SPEECH ERROR IN ENGLISH

Кваліфікаційна робота
освітнього рівня "магістр"
студентки II курсу,
спеціальності
Сучасна англомовна комунікація та переклад
і дві західноєвропейські мови,
галузь знань 03 – гуманітарні науки
спеціальність 035 «Філологія»

Митрофанкіної Марії Михайлівни

науковий керівник:
канд.філол.н.
Карабан А.В.

«Допущено до захисту»
Протокол засідання кафедри англійської філології
та міжкультурної комунікації
Протокол № 8 від 18.04.2023
Завідувач кафедри____________проф. Бєлова А.Д.

КИЇВ - 2023

CONTENTS

INTRODUCTION 4

CHAPTER I. WORD PROCESSING AND RETRIEVAL ERRORS AS PSYCHOLINGUISTIC PHENOMENA 7

1.1. Mental lexicon: storage and retrieval processes 7

1.1.1. Mental lexicon structure 7

1.1.2. Word recognition processes 11

1.1.3. Models of word production 15

1.1.4. Word processing deficiencies 21

1.1.5. Speech production effects and biases 22

1.2. Types of speech errors 26

1.2.1. Types of speech production errors 26

1.2.2. Types of speech perception errors 34

1.3. Verbal blunders vs malapropism 38

Chapter conclusion 39

CHAPTER II. RESEARCH DESIGN 43

2.1. Quantitative approaches to the investigation of eggcorns 43

2.1.1. Computational and corpus-based approaches 43

2.2. Quantitative approaches to the investigation of eggcorns 47

2.2.1. Descriptive qualitative approach 47

2.2.2. Functional semantic field approach 48

Chapter conclusion 49

CHAPTER III. LINGUISTIC PROPERTIES OF EGGCORN AS A SPEECH ERROR 50

3.1. Defining eggcorns 50

3.2. Formal aspect of eggcorns 51
3.2.1. Neighborhood characteristics of eggcorns 51
3.2.2. Frequency effects. Correlation of frequency and neighborhood density 58
3.2.3. Phonotactic probability effect 61
3.2.3. Morphosyntactic characteristics of target words and phrases 67
3.2.4. Formal classification of eggcorns 73
3.3. Etymology 77
3.3. Semantic aspect of eggcorns 81
3.3.1. Semantic fields 81
3.3.2. Semantic classification 83
Chapter conclusion 88
CONCLUSION 93
SUMMARY 97

INTRODUCTION

Language errors have long been a subject of interest for linguists and cognitive scientists alike since they illuminate the intricate mechanisms underlying human language perception and production. The study of language errors sheds light on the cognitive processes involved in language use, as well as various linguistic and extralinguistic factors that lead to the production of errors. Over the years, such prominent scholars as Sigmund Freud, Rudolf Meringer, Carl Mayer, Victoria Fromkin, and Gary S. Dell stressed the significance of researching speech errors in order to comprehend the nature of language and how it is acquired and utilized by humans, believing that breakdowns in a complex system can be most revealing of its inner workings. For that reason, the study of speech errors is of paramount importance in the construction of language processing and production models, since it provides insight into how language functions, as well as how it is stored in the so-called mental lexicon.

Since Meringer and Mayer established the tradition of speech error research in the late 19th century by publishing their corpus, alongside their still influential classification of speech errors, they have been studied extensively, with scholars proposing various models and classifications drawing on Meringer and Mayer's seminal contribution and attempting to explain the processes behind these phenomena. Nevertheless, it remains a contentious field represented by multiple competing theories, and a consensus on many issues regarding the study of speech errors is yet to be found. Among all the topical issues in the field of speech error studies, the so-called "eggcorns" have recently gained attention as a particularly interesting type of language error, since they combine a number of unique characteristics. The most distinctive feature of eggcorns is the concatenation of both phonological similarity with the original expressions and semantic plausibility, and they have been recognized as a valuable source of data for investigating the complex interplay between sound, meaning, and context in language processing. Nevertheless, there has been very little scientific coverage of the eggcorn phenomenon due to it

being a fairly young concept (the term itself was coined as early as 2003). Therefore, in this paper, we have aimed to closely examine the concept of eggcorns, ascertain its defining characteristics and patterns, and ultimately gain a better understanding of its place in the broader framework of speech errors.

The topicality of this research lies in the fact that the study of language errors is crucial for understanding the inner workings of language processing, while eggcorns, in particular, present a novel opportunity to explore how speakers use and interpret language.

The purpose of the research is to determine the defining characteristics of eggcorns, as well as various factors and processes that lead to their occurrence.

The following **tasks** are required to be solved to reach the aim of this research:

- To outline the concept of the mental lexicon, describe its main features and models of its structure and organization;
- To provide a comprehensive characterization of processes involved in word recognition and production;
- To consider and evaluate existing classifications of speech production and speech perception errors;
- To investigate the concept of competence error by comparing the notions of malapropism and eggcorn;
- To provide a comprehensive analysis and categorization of eggcorns, as well as to determine common patterns leading to their occurrence by examining their formal characteristics;
- To investigate the etymology of lexical items involved in the formation of eggcorns and to determine their semantic fields;
- To provide a semantic classification of eggcorns and outline their prominent semantic peculiarities.

The object of our research is speech errors and their role in language production and comprehension.

The subject of our research is eggcorn as a specific type of speech error.

The materials of the research are a sample of 100 eggcorns selected from the Eggcorn Database using random sampling technique.

The methodology of the research is a combination of qualitative and quantitative approaches, namely the elements of computational phonology approach, corpus-based approach, descriptive qualitative approach, and functional semantic field approach.

The scientific novelty of this research lies in the fact that while speech errors have been studied extensively, eggcorns are a relatively new notion that has only recently gained attention from linguists. By conducting a detailed analysis of the selected eggcorn sample, this research contributes to the growing body of knowledge on this phenomenon.

The practical value of the research is determined by the potential application of its findings in further studies on speech errors and word processing in the fields of psycholinguistics and cognitive linguistics.

The Master's paper is divided into three parts. The first part outlines existing scientific perspectives on word processing mechanisms and speech errors and constitutes the theoretical foundation of the study. The second part provides a detailed account of the methodology applied during the research. The third part comprises the description, analysis, and systematization of the research results. The Master's paper also contains an introduction, conclusions, appendices, a list of illustrative materials, and a list of references.

CHAPTER I. WORD PROCESSING AND RETRIEVAL ERRORS AS PSYCHOLINGUISTIC PHENOMENA

1.1. Mental lexicon: storage and retrieval processes

1.1.1. Mental lexicon structure

The mental lexicon is a concept used in linguistics and psycholinguistics and is generally understood as a mental dictionary containing information about the word stock of a speaker [53, p. 340]. According to Clark, the mental lexicon contains all the words known to a speaker that they can draw from when producing or processing speech. At the very least, a single word entry contains information on its meaning, syntactical features, morphology, and phonology [19, p. 2]. Since the mental lexicon is organized according to different properties (e.g., semantic, syntactic, phonological, and pragmatic) and contains various information about its entries, the metaphor of a dictionary is most often used in relation to it. However, this comparison might be too simplistic to account for what the mental lexicon actually represents. For instance, unlike a classical dictionary, the mental lexicon does not provide clear-cut, isolated definitions, but rather presents meanings in the form of use patterns, connections with sensory experience, and logical relations [27, p. 548]. Moreover, according to the consensus reached among the researchers, the mental lexicon is a dynamic system, with the lexical representations within it and connections between them constantly changing, evolving, and adapting depending on experience and context [84, p. 22]. Aitchison claims that the biggest difference between a dictionary and the mental lexicon is the amount of information stored in a single entry, with speakers possessing a much deeper knowledge about the meaning of words, possible contextual use, word ranks, frequency of usage, syntactic patterns, accents, different pronunciations, etc. Despite the amount of information stored, entries in the mental lexicon can be accessed by the speaker momentarily, which also serves as an argument for the complex organization of the mental lexicon. This is apparent from numerous experiments which showed that the standard speed of speech is six syllables a second (i.e., approximately three words) and that native speakers are able to recognize a word in approximately 200 milliseconds from its onset (i.e., around

one-fifth of a second). Furthermore, the idea of the mental lexicon as a dictionary implies it is organized according to the orthographical aspect. However, studies of speech errors show that, although they often involve orthographically or phonologically similar words, the words produced instead of the target are usually not direct alphabetic neighbors [1, pp. 7-14].

The organization of the mental lexicon has been a central topic of debate in psycholinguistics. The consensus among the researchers seems to be that the mental lexicon is organized into a complex, structured network of interrelated words and concepts. For example, Aitchison refers to the mental lexicon as "a gigantic multi-dimensional cobweb" [1, p. 72]. Merea and Wolter hold a similar view, noting that "vocabulary knowledge is rather more than the sum of the learners' knowledge of the individual words in their vocabulary" [77, p. 88]. More specifically, it is believed that lexical entries are interconnected based on shared features in different language domains [1, p. 72].

Graph theory has been increasingly used by researchers to examine the network structure of the mental lexicon. This mathematically rigorous approach presents the mental lexicon as the collection of words represented by nodes with connections between them represented by paths or edges [103]. An especially influential theory called the Spreading Activation Model was developed by Collins and Loftus who argued that a complex network of nodes (i.e., concepts) sharing connections based on different types of relationships and with varying degrees of strength could serve as a psychologically plausible representation of the mental lexicon. In their model, the strength of the connection between concepts is represented by the length of the edge linking them and is determined by the number of their common features, while each feature is a separate concept (and, hence, a node in the network) in and of itself [20]. The connections in the semantic network are motivated both by word association, semantic similarity, and lexical cooccurrence [80; 97; 61]. Moreover, the strength of connections between the concepts and their features depends on the frequency with which a speaker uses the concept.

It has been suggested that the mental lexicon exhibits a so-called “small world structure” [61, p. 182]. Two key small-world characteristics are high clustering and short path length between nodes. High clustering refers to the high degree of interconnectedness in the mental lexicon: words connected to another word are likely themselves related to each other. On the other hand, a short path length means that two words can be connected through only a few connections. Small-world characteristics explain why the speaker is able to retrieve words rapidly and efficiently despite the fact the mental lexicon contains an extensive amount of information [11; 22; 60; 96].

Aside from the conceptual semantic network, the Spreading Activation Model also includes a lexical network, which stores the phonological and orthographic properties of words. Since the semantic network includes only abstract concepts, the actual names that speakers attribute to them, their pronunciation, and their written form are represented in the lexical network of the Spreading Activation Model. The strength of the connections between the nodes of the lexical network depends on their phonological and orthographic similarity [20, p. 407].

When it comes to the morphological structures in the mental lexicon, it has been debated whether morphologically complex words are stored wholly (full listing), as a number of morphemes with distinct representations (full parsing), or they can be accessed either way depending on the regularity of forms (dual-route) [112, p. 396]. Finally, it has been proposed that both complete forms and separate morphemes are activated in a parallel fashion [34, p. 370].

The criteria of lexical organization at the semantic level can be understood by examining the qualitative characteristics of separate lexical groups. Whether these criteria are taxonomically or thematically based has been a topic of debate in the scientific community [70; 51]. However, numerous studies of lexical clusters indicated the existence of an extensive thematic structure [23; 67]. For instance, it has been shown that the taxonomic category of “birds” might include not only different bird species but also words like “nest”, “beak”, and “egg” [8, pp. 435-436]. In a

language, the majority of words are taxonomically connected to a very limited number of words, yet may transpire in various thematic situations, suggesting that a thematic arrangement is its intrinsic feature [61, p. 185].

The proposed method of processing, spreading activation, is reflected in the model's title. A word becomes activated when it is heard, thought about, or read by the speaker. The activation then spreads to other connected nodes, which consequently lend activation to their respective neighboring nodes. However, the strength of the activation is proportional to the strength of connections and weakens with its outward movement from the source. Thus, frequently used words have stronger connections, which allows for quicker access to neighboring concepts or features [Collins and Loftus]. This can be exemplified by a study conducted by Rosch, where participants made a decision that a "canary" is a "bird" more quickly than in the case of "penguin". According to the Spreading Activation Model, the connection between "bird" and "canary" is stronger because "canary" is a more typical member of a category and, thus, it is more easily accessible – this is known as the "typicality effect" [88, p. 224].

The Spreading Activation Model is also used to account for the tip-of-the-tongue (TOT) phenomenon, a state in which a speaker cannot retrieve a familiar word. Brown and McNeill's study of the phenomenon revealed that participants in the tip-of-the-tongue state were nevertheless able to accurately determine the initial letter, the number of syllables, and the primary stress of the target word [13]. This finding points to the fact that phonological structure also plays a role in the organization of the mental lexicon, which is reflected in the lexical network of the Semantic Activation Model.

As for the enrichment of the mental lexicon, according to Carey and Bartlett, entries of the mental lexicon are acquired by mapping word forms to meaning [17, p. 18]. The process of fast mapping allows one to learn a word through minimal exposure to it, whereas speakers expand existing lexical entries with additional information with increasing frequency of exposure, i.e., through slow mapping [85, p.

378]. This is supported by the results of the study conducted by Richter et al. which suggests word form representations are established first, while semantic properties of the word are added later in the process of acquisition [65].

Finally, important features of the mental lexicon are its size, also known as vocabulary breadth, and the quality of its lexical entries (or vocabulary depth). Lexical quality is the measure of how well a person's knowledge about a word reflects its meaning, form, usage, and pragmatic features. Both vocabulary breadth and average lexical quality of entries vary among individuals and may depend on different factors, such as age, experience, and the level of expertise in various fields [85, p. 359].

1.1.2. Word recognition processes

It is possible to utilize studies on the structure of the mental lexicon to explain the processes behind language processing. In speech, the message is decoded through the mapping of the auditory information onto the word representations stored in the mental lexicon. Although listeners usually perceive this process as fairly easy and natural, the actual procedure behind it is far more complex due to some inherent features of spoken language. Firstly, vocabularies are developed from a small number of phonemes, which leads to a resemblance between words and short words being part of longer ones. The second feature is a high degree of variation due to such reasons as different phonological contexts, speaking styles, and rates. Finally, speech is transient, being dispersed in time, swiftly fading from perception, and lacking distinct word boundaries [112, p. 387].

Spoken word recognition is realized through a combination of bottom-up (data-driven) processes activated by the auditory signal and top-down (conceptually driven) processes caused by linguistic context [55, p. 338]. Studies of the phonemic restoration effect serve as evidence of the involvement of top-down, context-based processes in speech recognition. They have shown that context has an impact on the perception of the ambiguous element in the sentence. For example, participants exposed to the sentence "It was found that the *eel was on the axle", where the

asterisk signifies a removed fragment, heard "*eel" as "wheel", while those listening to the sentence "It was found that the *eel was on the shoe" heard "heel" [111, p. 32].

During speech perception, various prosodic cues, such as stress and intonation, facilitate the recognition of the syntactic or grammatical structure of a sentence by the listener. Studies have shown that listeners are usually successful in predicting the general structure of a sentence from a short fragment, which indicates the facilitating effect of prosodic information [9]. Interestingly, lip-reading has also been shown to be extensively utilized to understand speech, even among normal-hearing listeners. For instance, the McGurk effect, which is when the auditory information producing one sound combined with the visual information indicating the production of another sound leads to the perception of a third sound different from the original ones [76]. Visual information from lip movements facilitates the recognition of speech since the information provided by speech sounds alone is frequently insufficient [55, pp. 339-340].

Research in the field of psycholinguistics has produced some other important findings on spoken-word recognition. One of the most important discoveries is that comprehension is incremental: the process of comprehension begins before the utterance is finished, and several candidates compatible with the incoming auditory information get activated simultaneously. Specifically, experiments conducted by Marslen-Wilson in the 1970s revealed that a word can be recognized even before the speaker finishes uttering it (i.e., before its acoustic offset) [75].

Parallel activation of multiple words is evident from priming experiments, where a word onset common for two different words (e.g., both 'captain' and 'captive' beginning with /capt/) makes it easier to recognize words with meanings similar to those of both possible alternatives (e.g., 'ship' and 'slave'). This suggests that all words consistent with the available auditory input get activated instantly. However, lexical competitors begin to deactivate when their compatibility with the input ceases: in experiments, only words related to the target were primed after the offset of the word was heard (e.g., the word 'slave' related to the word 'captive'

received no priming when the offset of 'captain' was heard) [116, p. 34]. According to Marslen-Wilson, competitors of the target get deactivated when the word reaches its "uniqueness point", a place where there is no longer any overlap with other words in the mental lexicon (e.g., the uniqueness point of the word 'captain' is /n/) [73, p. 134-135]. The priming effect has also been observed among semantically related words: experiment participants who were shown a certain word responded to words semantically related to it more quickly than to unrelated words [79].

Parallel activation has also been observed in the case of words embedded in longer words. For example, 'bone' in 'trombone' facilitates the recognition of the semantically similar word 'rib', which means that the word related to 'bone' is activated even if 'bone' is presented as part of a longer word [93, p. 25]. The mechanism of parallel activation has more recently been demonstrated through eye-tracking experiments: listeners presented with several objects need more time to look at the target designated in a spoken statement when objects with names similar to that of the target are also present (for example, a candle is looked at later if candy is also shown). Therefore, for a period of time, listeners consider both similarly named objects as potential targets [52, p. 1633].

Furthermore, numerous studies show that phonologically related words have an impact on how quickly lexical decisions are made. In addition to being activated concurrently, word candidates also compete with one other, meaning that the increase of activated candidates can lead to the higher degree of inhibition they enact on one another [109]. In the studies on the mental lexicon, neighborhood effects refer to the influence of the target word's neighbors (i.e., lexical entries that share features with the target), which can be either facilitatory or inhibitory [4, p. 123]. The phonological neighborhood is usually defined as the number of words that differ from another word by one phoneme. A number of studies measuring the time of processing and the degree of accuracy have demonstrated that words with large phonological neighborhoods are recognized slower and less accurately than words with small phonological neighborhoods [42; 71]. Since in speech perception, a selection from

numerous partially activated items should be made, the increased number of similar neighbors makes it more difficult to single out the target from other activated candidates [104, p. 224].

Moreover, research has demonstrated that semantic neighbors can also influence the processing of a word. For instance, Buchanan et al. found that the larger number of semantically related neighbors corresponds to the quicker response to the target word by the participant [15]. Subsequent research has shown evidence that the semantic neighborhood effect is more pronounced for low-frequency words, the meanings of which are less well-established in the mental lexicon [72]. Furthermore, it has been demonstrated that the number of semantic neighbors facilitates the activation of abstract concepts [86].

Another facet of the neighborhood effect is the neighborhood frequency, i.e., the frequency of the word's neighbors. As has been verified experimentally, a word with a high neighborhood frequency is likely to be recognized slower and less accurately than a word with a low neighborhood frequency [106, p. 215].

As has been demonstrated in the discussion of neighborhood effects above, word frequency also plays a role in the process of word recognition. Multiple studies have demonstrated that high-frequency words are more readily recognized than low-frequency words both visually and auditorily [99; 7; 72; 31; 74]. Moreover, the study by Savin established that high-frequency words are recognized not only faster but also with better quality than low-frequency words in the conditions of a noisy signal [91].

Another type of effect on word recognition is the concreteness effect, which presupposes that easily imageable words (e.g., 'chicken') are processed more quickly and accurately than abstract words (e.g., 'intuition') [58, p. 94]. There are two competing theories aiming to explain this effect. The dual-coding theory put forth by Paivio contends that lexical entries of concrete words in the mental lexicon contain a visual component, aside from linguistic information, while this element is absent

from abstract word entries [83]. On the other hand, the context-availability hypothesis brought forward by Bransford, McCarrell, and Kieras asserts that concrete words simply possess more associative links in the mental lexicon network [12; 56]. A thorough explanation of the concreteness effect, however, is likely to incorporate aspects of both views [98, pp. 109-110].

Finally, the role of the lexical status effect and non-word legality effect should be considered. The lexical status effect refers to the fact that the recognition and processing of real words are fast and more precise than that of non-words (i.e., meaningless strings of letters) [89, p. 492]. It is assumed that this observation reflects the fact that real words are more frequent and fixed in the mental lexicon, which leads to quicker and more efficient processing [32, pp. 693-694]. As for the non-word legality effect, it refers to the fact that non-words that conform to the legal patterns of the language and are similar to real words are recognized and processed faster than non-words that violate language patterns and do not resemble real words [94, pp. 34-35]. This effect is thought to be explained by the fact that non-words resembling real words are more likely to be mistaken for them, which leads to faster processing [31, pp. 632-634].

1.1.3. Models of word production

It has long been recognized by researchers that the mechanisms behind complex systems can be revealed by considering the breakdowns in such systems. For instance, throughout the history of psycholinguistic research such scholars as Sigmund Freud, Meringer, Mayer, and Fromkin argued that speech errors can shed light on various linguistic processes and laws governing speech formation [35, p. 71; 78, p. 10; 36, pp. 43-44].

Word retrieval models usually envision the lexicon as a structure consisting of three levels, namely conceptual, lemma (word) level, and phonological level. These levels are organized in a network, which contains nodes representing different types of linguistic information, such as concepts, words, morphemes, phonemes, and

phonemic features. Accordingly, these nodes are connected across different levels [24, p. 286; 18, p. 2].

Evidence obtained by psycholinguists, cognitive psychologists, and neuropsychologists suggests the existence of two stages in the process of language production. Presuming that the lexical concept is already identified, the first stage is the selection of the corresponding syntactically and semantically specified lexical representation (i.e., lemma), after which follows the ordering of the word form (selection of morphemes, phonemes, and phonemic features [16, pp. 309-310; 63, p. 224; 24, p. 286]. The two-stage model is supported by evidence obtained from studying speech errors. In 1975, Garrett discovered that words involved in word exchange errors can be separated by a certain distance and belong to the same grammatical category but possess different phonological structures (e.g., "he left it and forgot it behind"), while sounds involved in sound exchange errors usually emerge in words that are situated closely and have similar phonological environments but different grammatical classes (e.g., "rack pat" for "rat pack"). These constraint patterns reveal that the processing of linguistic information occurs at different stages of speech production, namely a stage of assigning syntactic functions and a stage of formal ordering [40, pp. 133-177]. Moreover, the investigation of the tip-of-the-tongue (TOT) phenomenon is used in support of the two-stage model, since studies show that speakers in a tip-of-the-tongue state are aware of the grammatical properties of the word they are unable to utter [25, p. 520]. Such evidence suggests the existence of two levels of word representations in the mental lexicon: syntactic/semantic and phonological. Additionally, the existence of morpheme errors and phoneme errors serves as evidence that word forms are not necessarily stored wholly but can be ordered from smaller units (i.e., morphemes and phonemes) [92, pp. 256-257].

Although word production models usually do not include any subsequent stages, it is generally assumed that after the sequencing of phonemes is completed, phonemes are given jolts of activation that lead to the transformation of phonemes

into articulatory codes [24, p. 806]. For instance, Levelt has put forward the idea of the syllabary, a collection of memory representations that specify motor actions required to produce syllables, which is accessed at the moment of transition between phonological encoding and articulation [63, p. 230].

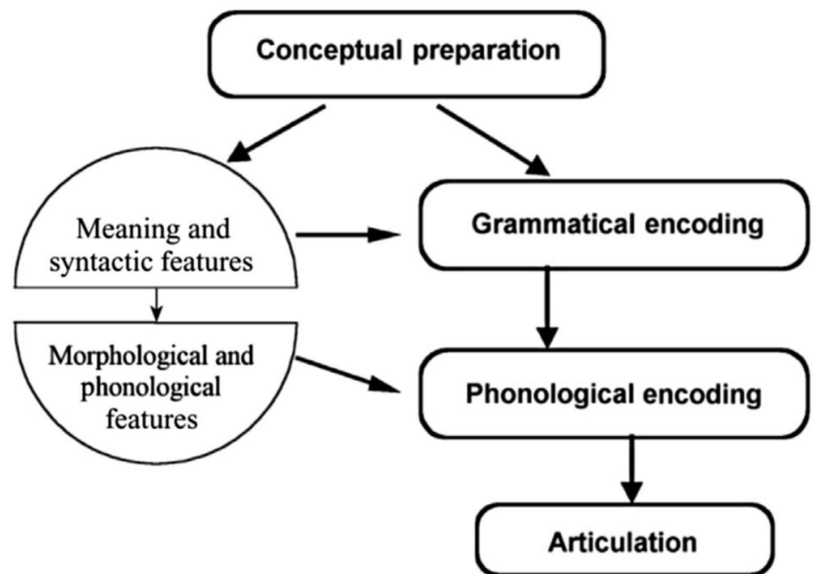

Figure 1.1. Levels of processing in speech production reproduced from Schriefers and Vigliocco (2001) [92, p. 255]

Despite an overwhelming consensus on the existence of two stages of lexical access, speech retrieval models can differ in the interpretation of ways in which this notion is implemented, especially whether the stages of processing are discrete or interactive [16, p. 310]. The differing views on the discreteness of the stages stem from two major historical roots of word production studies. The chronometric tradition of studying the naming of objects, drawings, and words provided an array of important insights into the mental lexicon and mechanisms of accessing it. For instance, P. Fraisse discovered that the response during the naming of a line drawing took longer than the response during the reading of a letter or a number. This led to the conclusion that the tasks of reading and naming involved different coding processes. Therefore, there is considered to exist a direct link between the word and its phonological content, while a line drawing first activates a concept of the object, which constitutes an additional step [33, p. 354-355]. Chronometric studies provided

the foundation for the discrete stage theory, according to which the second stage that involves accessing the phonological form can only begin when the first, semantic-to-lexical stage is completed. Discrete models presuppose the activation of a set of lemmas that includes both the target lemma and lemmas that share semantic features with it. However, since phonological encoding doesn't begin until the selection of the target lemma has been completed, phonological forms of words semantically similar to the target do not receive activation [92, p. 257].

Nevertheless, the discrete model was created to explain response delays rather than speech errors and it has shown difficulties in explaining mixed speech errors that display phonological and semantic motivation, thus suggesting the simultaneous activation of different levels [6, p. 279]. The theorists advocating for the discrete model of word production have proposed the existence of an internal self-monitor to account for the existence of mixed errors. The said self-monitor searches for and eradicates errors in internal speech, however, phonological errors may still pass undetected when they are semantically similar to a target and vice versa [5, p. 2]. However, the study by Hartsuiker et al., although admitting the possibility of the monitoring bias, concluded that the effect of feedback was preeminent [50]. Moreover, experiments conducted in a study by Severens et al. showed inconsistencies in the results with the predictions of the biased monitoring theory and led to the conclusion that certain phenomena, like homophone completions, could only be explained by an interactive model, rather than a discrete one [5, p. 15].

On the other hand, interactive stage models are derived from the tradition of speech error analysis and suggest that semantic, lexical, and phonological processes overlap in speech production [47, p. 292]. A two-step interactive activation model proposed by Dell has become the most influential, with theorists of speech production having largely arrived at a consensus that it can be successfully used to describe the process of speech retrieval, as well as account for speech errors. Essentially, it upholds the differentiation between the two stages but allows for lower levels to initiate processing prior to the completion of processing in the higher levels (a

process known as cascading) [25, p. 520]. Hence, each activated lemma lends activation to the corresponding word form, and, unlike in the discrete model, phonological forms of the words semantically related to the target also receive some activation [92, p. 257]. Moreover, as evident from the term “interactive”, all connections within the model are bidirectional and it permits processing at lower levels to impact that of higher ones through feedback. Therefore, there exists a bottom-up connection corresponding to each top-down connection that transmits positive feedback from later to earlier levels [24, p. 288]. In this way, a lemma provides activation to concepts with features matching its specifications, while phonological representation can activate lemmas with shared phoneme nodes. Therefore, items that are both semantically and formally linked to the target receive higher levels of activation [6, p. 279].

The CAT example is usually used to describe this model. Firstly, the lemma node CAT gets activated due to its characteristic features, and, subsequently, it spreads activation to phoneme nodes /k/, /æ/, and /t/. On the other hand, activation also spreads to its semantic neighbors (e.g., DOG) through shared semantic features. Moreover, words like MAT receive activation through feedback from phonemes shared with the target. Since words like RAT gain activation through both shared semantic (e.g., “animate”, “mammal”) and phonetic features (/æ/ and /t/), they are usually more activated than purely semantic or formal neighbors. Therefore, there exists a possibility that through activation noise a semantic, formal, or mixed neighbor will be selected instead of the target [25, p. 522]. Furthermore, increased activation of words that are both semantically and phonetically related to the target explains the statistically overwhelming number of mixed errors in speech error corpora. The notion of feedback is also used to explain the tendency of speech errors to result in real words, instead of nonwords (e.g., MAT, rather than GAT or LAT) [63, p. 225].

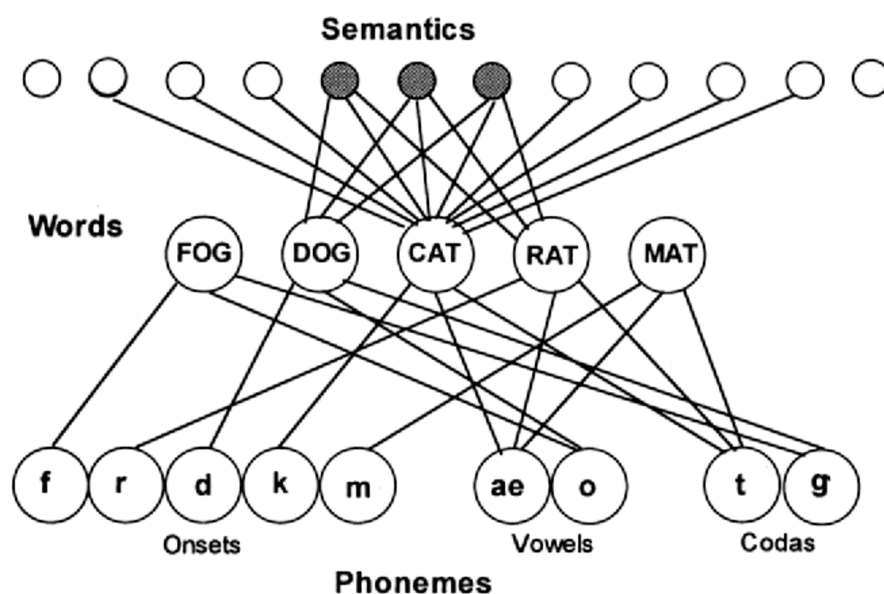

Figure 1.2. Lexical network of the interactive two-step model reproduced from Dell et al. (1997) [64, p. 805]

In their extensive review of language production literature, Vigliocco and Hartsuiker considered evidence for and against assumptions of interactive and discrete models about the nature of input between stages and the directionality of the information flow. Drawing on a large number of studies, they disproved the idea of unidirectional information flow, as well as the notion that activation spreads to the following level only after the completion of processing at the previous level. Instead, they discovered empirical evidence that language production incorporates cascading activation, which accounts for fluency and effectiveness, as well as feedback that serves as a monitoring system for improving accuracy [101, p. 468]. Similarly, Lee and Gibbons discovered evidence of feedback in speech production by observing the link between the use of the complementizer and the preference for rhythmic alternation. Interestingly, such empirical proof of the influence of phonology on grammatical encoding could not be fully explained by the internal self-monitor envisioned by the discrete models, since sentences that do not adhere to the alternating rhythm, although not preferred, can not be considered errors. Therefore, the study's findings reaffirm the existence of feedback from the phonological to the grammatical encoder that ultimately influences the decision of the latter [62, pp. 453-454]. Thus, overwhelming evidence in support of the two-stage interactive model, as

well as the fact that it was designed specifically to account for speech errors, make it especially relevant for the present study.

1.1.4. Word processing deficiencies

There has been an extensive amount of research on naming difficulties as pertains to individuals with aphasia and language impairments, and their underlying causes for naming have been explained by the structure and processes pertaining to the mental lexicon. For instance, Best summarizes them by distinguishing between deficits in storing lexical semantic and phonological information in the mental lexicon and deficits in accessing lexical semantic information and phonological form for word production [10, p. 281].

In the early stages of word acquisition, children have been shown to make naming errors that entail overextension of a familiar word to a new example that shares some prominent features with it (e.g., perceptual features such as shape, size, or color, as well as functional features). Typical examples of childhood overextensions include calling a cow “doggie”, a raccoon “bear” or a grapefruit “moon” [41, pp. 665-666; 81, p. 20]. Overextensions have been interpreted as either comprehension or performance deficiencies. The former view suggests that overextensions may reflect the child’s underlying semantic representations (e.g., a child assumes that both dogs and cows are meant by “doggy”). According to this interpretation, overextensions happen because children are not yet fully familiar with the criterion qualities of specific lexical items. On the other hand, overextensions might appear due to limited vocabulary breadth or inability to retrieve a known word. In the first instance, children are forced to extend to the inadequate number of words available to them with the aim of effective communication. In the second instance, children may be familiar with the correct word or be in the process of learning it (which corresponds to the stage of fast mapping) but ultimately experience difficulties with retrieving it spontaneously, which necessitates its substitution with a related, more well-known word. It is important to note that, although the phenomenon of overextension has been studied primarily in relation to children due to it becoming

less frequent with age, it has been observed that overextensions and mechanisms leading to them persist to some extent throughout the whole lifespan of a speaker [81, p. 20-21].

As for the instances of faulty processing and storing of information in the mental lexicon of adults, it is also advantageous to consider the good-enough language processing theory that highlights the propensity for superficial analysis of linguistic input by the comprehension system, which may lead to the erroneous interpretation [30, p. 218]. The idea of superficial processing of linguistic input is supported by the evidence from semantic illusion research, which demonstrates that people tend to overlook obvious errors or inconsistencies and normalize nonsensical utterances. These findings challenge the postulation that language comprehension unerringly constructs complete and correct representations of utterances but rather suggests the opposite: people tend to come up with partial and distorted representations (i.e., they are only "good enough" instead of authentically reflecting the actual information). Such distortions may occur due to noisy channels and end up being more plausible to the receiver than the actual message [29, p. 197].

1.1.5. Speech production effects and biases

According to the interactive activation model of word production, speech errors occur when incorrect entries get chosen instead of correct ones due to higher levels of activation. However, there is a number of constraints and biases that influence the selection. Dell distinguished four types of main influences on the probability of a speech error: output biases, similarity effects, speech rate effects, and distance effects.

Output biases reflect the propensity of speech errors to result in plausible or frequently observed sequences of units. Among them is lexical bias, which refers to the tendency of sound errors to produce real morphemes or words, and semantic bias, i.e., the propensity of sound errors to result in words that share semantic connections with other words in their environment. In the interactive activation model, output biases are explained by positive feedback altering the activation pattern at a certain

level to produce errors that are meaningful in relation to higher levels. Thus, sound errors tend to result in real words or morphemes, which tend to be associated with those nearby [24, pp. 292-293].

As has been demonstrated previously, frequency plays a significant role in the word storage and recognition processes. Likewise, its strong effect on word production has been well-documented in speech error research, with Dell classifying them as a subtype of output biases. At the phonological level, frequent sounds tend to substitute infrequent ones, since they share connections with more syllable nodes. The quantity of these connections (the sound's syllabic frequency) affects the sound's activation: as it spreads from sounds to syllables and back, frequent sounds receive greater activation from activated syllables. The same frequency bias occurs in relation to frequent syllables and morphemes [24, p. 301]. At the lexical level, low-frequency words were found to be more likely to lead to a semantically similar speech error [49]. According to Gollan and Brown's experiments, word difficulty associated with frequency affects the ability to retrieve words from pictures, and resulting errors reflect both difficulties with word form and word meaning access [43]. On the other hand, increased familiarity with a word facilitates the retrieval of its representations from the mental lexicon, increasing its likelihood of being produced [113, p. 483]. Furthermore, in speech errors, target words tend to be substituted by higher-frequency words, rather than by words of lower or equal frequency [106, p. 223].

Similarity effects present the second type of influence on errors and are manifested in several ways. Firstly, words interacting in an error tend to be phonologically and semantically related, while interacting morphemes and phonemes are also usually phonologically similar [24, p. 293]. Among the most often investigated similarity factors in psycholinguistic research are the number of syllables, stress patterns, and common initial segments [6, p. 275]. Secondly, the immediate surroundings of the interacting items frequently match. The repeated phoneme effect may serve as an illustration of this: phoneme exchange errors are

more frequent when the nearby sounds are the same (e.g., in the phoneme exchange 'left hemisphere – heft lemisphere', the vowel /e/ is directly next to the phonemes being exchanged). According to the interactive activation model, similarity effects are caused by close connections at different levels in the network, which act as routes through which similar items activate each other. Thus, the target possesses similar competitors, the activation level of which may surpass that of the intended item and lead to speech error [24, p. 293].

At the syntactic level, the categorical constraint presupposes that even when the incorrect word is selected, it should, nevertheless, adhere to the same syntactic category as the target word [6, p. 275]. The categorical constraint is also manifested at the phonological and morphological levels. However, the category in question is not syntactical, instead, vowels usually interact with other vowels, consonants with consonants, stems with stems, and affixes with affixes [24, p. 284].

The third factor that affects the likelihood of an error is the speech rate: the likelihood is higher if the speaking rate is fast. Theoretically, this is explained by the fact that at a fast rate, previously activated items lack time to decay, while new intended items have yet to gain enough activation from higher levels for efficient competition.

Finally, distance effects refer to the tendency of items involved in misordering errors to move over short distances. For instance, misordered sounds and morphemes usually move to nearby content words. The theory explains this tendency by the fact that the most activated competitors for a certain slot are usually those that have recently been or are about to be chosen for neighboring positions [24, p. 293].

Aside from the aforementioned effects, the error phenomenon known as the initialness effect has been widely observed: initial sounds of words and syllables have a bigger tendency to be involved in a speech error than other parts [24, p. 312]. This can be explained by the fact that in the case of non-initial positions the network can forecast upcoming segments due to the memory of previous ones [2, p. 7].

Another speech error pattern widely reported in corpus analyses and experiments is the mixed-error effect, which predicts that errors involving both phonologically and semantically related words are more likely that errors involving words that share only one of those connections. This effect can also be explained by the interactive spreading activation model which predicts that activation at the semantic level and feedback from the phonological level converge at the lexical level [6, p. 279].

It is equally important to consider the neighborhood effect in relation to speech production. Numerous studies suggest that words with large phonological neighborhoods have a tendency to be produced more accurately and efficiently than those with small neighborhoods [48; 95; 110]. This effect is most often explained by referencing the interacting spreading activation model which presupposes that the more phonological neighbors a word possesses, the more feedback activation it receives from them, which facilitates the selection process [44, pp. 119-120]. On the other hand, Vitevitch's research on speech errors provided additional findings regarding the neighborhood effect: namely, low-frequency words with small neighborhoods and high-frequency words with large neighborhoods were found to result in speech errors more often [106, p. 225]. Vitevitch explains these findings by the assumption that high-frequency words surrounded by highly activated competitors may be "drowned out", and another activated item may be erroneously selected. On the other hand, if a low-frequency word has a large number of neighbors, it may gain facilitatory activation from them. Conversely, if the number of neighbors is small, the activation that a low-frequency word receives may not be enough for the speech production system to select it [105].

Moreover, since high-frequency phonological neighbors have higher levels of activation, the target word receives more feedback activation from the phonemes they share. On the other hand, the target word receives less activation from low-frequency phonological neighbors, since they are less activated themselves. Therefore, words

with a low neighborhood frequency have a smaller chance to be retrieved successfully and might be involved in a speech error [3, p. 232].

1.2. Types of speech errors

1.2.1. Types of speech production errors

Speech production error is typically defined as an unintentional departure from the speaker's linguistic intention during speech production. It is widely believed that the tradition of speech error research was established with the publication of Meringer and Mayer's corpus in 1895 [63, p. 224]. In addition to the extensive collection of German speech errors, it included their theoretical analysis and established several influential distinctions. For instance, Freud's claim that every form-based error had an unconscious but meaningful rationale behind it was debunked, with Meringer and Mayer making a now universally recognized distinction between form-based and meaning-based errors. Moreover, they acknowledged the frequent appearance of a phonological connection in meaning-based errors. Therefore, the prevalence of mixed errors was already documented at the starting point of the speech error investigation. Importantly, Meringer and Mayer also introduced the standard categories of speech errors known as anticipations, perseverations, exchanges, and blends (or contaminations). More specifically, anticipations occur when a later item replaces an earlier one (e.g., "taddle tennis" for "paddle tennis"); conversely, perseverations occur when an earlier item takes the place of a later one (e.g., "been abay" for "been away"); exchanges are two units switching places (e.g., "mell wade" for "well made"); blends are two competing words combined together (i.e., both are selected only partially) [78; 63, p. 224].

The tradition established by Meringer and Mayer has been carried on by other linguistics and psychologists, with a substantial revival beginning in the late 1960s. In 1973, Fromkin collected significant studies on speech errors, supplementing them with her own compilation of errors. At the level of a phone (a consonant or a vowel), Fromkin differentiates between substitution, transposition (metathesis), omission and addition of segments, and movement. Furthermore, substitution errors may be

anticipatory or preservatory. Transposition (or metathesis) errors present a more complex error type that involves both anticipation and perseverance but can be described more simply as the "switch in the linear ordering of the sounds intended" (for example, "keep a tape" – "teep a cape") [38, p. 219]. When a metathesis involves the first sound of two separate words, it is known as a spoonerism, named after Dr. William Spooner who was prone to make this type of slip of the tongue [100, p. 321]. Research on spoonerisms shows the effect of similarity, i.e., consonants invariably exchange with consonants and vowels with vowels, and the phonemes involved in the exchange typically sound similar [55, p. 403]. Movement (or shift) occurs when one segment is transported from its appropriate location to another one (e.g., "ice cream" – "kise ream") [37, p. 115]. The individual segments can be further combined in clusters (for example, when "fish grotto" is pronounced as "frish grotto", the addition of /r/ in the first word creates a cluster instead of the target single segment). Consonant clusters can be split or moved as a single unit. Furthermore, although rare, speech errors involving a single feature can occur (e.g., in "clear blue sky" – "glear plue sky" voiceless velar /k/ becomes a voiced /g/ and the voiced labial /b/ becomes a voiceless /p/) [38, pp. 220-225].

According to the lexical bias effect, the large majority of speech errors are constrained by grammatical rules. Therefore, segmental errors conform to grammatically permitted sequences of sounds: for instance, "slips of the tongue" cannot be uttered as "tlip of the sung" because, although the sound "tl" can be physically pronounced by the speaker, it is not allowed to be used at the beginning of a word [37, pp. 113-114].

While errors at the level of a phone are the most common, morphemes are another type of speech performance unit that can be involved in the substitution, omission, addition, and blend, yielding speech errors. Derivation morpheme errors demonstrate a breakdown in the application of word formation rules, leading to the utterance of a nonexistent word (e.g., "grouping – groupment", "bloody" –

"bloodent"). Such speech errors reveal that morphemes are stored as separate items with an inherent meaning in the mental lexicon [37, p.115].

Another speech performance unit is an entire word, which can be involved in transposition, substitution, and blending. A portmanteau word is a subtype of blending and presents the combination of two words with similar meanings (e.g., "instantaneous" and "momentary" – "momentaneous"). This type of error serves as evidence of the fact that the semantic concept is generated independently of the word representing it. Consequently, the speaker unsure of the right word to use may blend two synonyms. In word substitutions, the target and the resulting word often belong to the same semantic class or, on the contrary, are antonyms, which suggests that semantic representations of words are stored as "a composite of hierarchically ordered semantic features" [37, p. 117].

Finally, syntactic structures have also been shown to be units of linguistic behavior that can be involved in a speech error. The fact that words are organized into larger phrases and stored in the so-called "buffer memory" before individual segments or words are disordered is demonstrated by such transposed phrases as "Nerve of a vergeous breakdown" instead of "Verge of a nervous breakdown", as well as by the fact that the intonation pattern of such erroneous utterances often remains identical to the one in the intended phrase [37, p. 116].

Dell provided the final significant theoretical instrument in this research tradition by producing his model of word production that accounts for the observed patterns in the production of speech errors and their types. According to Dell's exhaustive typology of speech errors, they can be divided according to various factors. Depending on the size of the unit involved, speech errors can be divided into sound, morpheme, and word errors. Among speech errors, those involving single phonemes, consonant clusters, and vowel-consonant combinations can be distinguished.

In terms of the type of disruption, errors are divided into contextual (syntagmatic) and non-contextual. Contextual errors involve a mistake in the ordering, while non-contextual errors do not have an obvious source within the actual utterance. Contextual errors include "exchanges, anticipations, perseverations, shifts, anticipatory and perseveratory additions, and some kinds of deletions". Shifts are either exchanges of a consonant and a null element (e.g., "murchin arriage" for "urchin marriage") or exchanges of a consonant and a cluster containing it (e.g., "intfan" for "infant"). On the other hand, deletions are substitutions of a null element for a consonant or a consonant for a cluster containing it [24, p. 298].

According to the theory, anticipations arise when a future item in the same category as the intended earlier one possesses more activation. The main cause for this is that later items get activated by the nodes in the higher level and ultimately receive more activation than the target item. For example, in the morphemic error "sim swimmers sink", the node of the vowel /i/ is activated through spreading activation due to "sink" and "swimmer" being processed by the syntactic and morphological systems while the syllable |sʌm| is being processed at the phonological level. If /i/ eventually receives more activation than /a/, it will take its place during the selection of the vowel for the first syllable. The same logic applies to perseverations, with the exception that the interference is caused by items that have already been encoded but still maintain activation.

On the other hand, exchanges are more complex: they occur when an anticipation error leads to a perseveration regarding the item replaced during the anticipation. When, during anticipation, a wrong item is chosen and tagged as part of the representation, it loses activation, while the intended item remains activation due to not having been selected and, consequently, can be selected for the following slot of the same category. For example, in the morphemic exchange "some sinkers swim", "sink" substitutes "swim" as an anticipation. Since after being tagged its activation level gets set to zero, it is unlikely to be selected for the next stem slot. On the other hand, "swim" is likely to possess a sufficient level of activation after being

overlooked in the present word and will be selected instead, leading to an exchange [24, p. 292].

Non-contextual errors, which are characterized by outside interference, include "noncontextual substitutions, blends, additions, and deletions" [24, p. 284]. In the case of non-contextual errors, nodes that are not meant to be part of the utterance are triggered by the spreading activation. Although typically their activation levels are not sufficient to allow them to be selected instead of the intended item, it is possible for such a node to gain activation from the background sources of verbal activation, such as "from concepts that were either presuppositions or inferences that were necessary in the semantic and pragmatic planning of the utterance" [24, p. 291]. For example, when planning the utterance "Could you close the door?" the presupposition that the door is open is bound to be processed. Consequently, the concept of "open" receives activation that spreads to other nodes associated with it, which can ultimately lead to the error where "close" is substituted for "open".

Table 1.1

Type	Examples	Unit involved[a]
Sound errors		
Misordering		
Substitution		
Exchange	*York library → lork yibrary*	Phoneme
	Spill beer → speer bill	Rime constituent
	Snow flurries → flow snurries	Consonant cluster
	Clear blue → glear plue	Feature
Anticipation	*Reading list → leading list*	Phoneme
	Couch is comfortable → comf is . . .	Syllable or rime
Perseveration	*Beef noodle → beef needle*	Phoneme
Addition		
Anticipatory addition	*Eerie stamp → steerie stamp*	Consonant cluster
Perseveratory addition	*Blue bug → blue blug*	Phoneme
Shift	*Black boxes → back bloxes*	Phoneme
Deletion[b]	*Same state → same sate*	Phoneme
Noncontextual errors (substitution, addition, deletion)	*Department → jepartment*	Phoneme
	Winning → winnding	Phoneme
	Tremendously → tremenly	Syllable
Morpheme errors		
Misordering		
Substitution		
Exchange	*Self-destruct instruction → self-instruct de . . .*	Prefix
	Thinly sliced → slicely thinned	Stem
Anticipation	*My car towed → my tow towed*	Stem
Perseveration	*Explain . . . rule insertion → . . . rule exsertion*	Prefix
Shift	*Gets it → get its*	Inflectional suffix
Addition	*Dollars deductible → dedollars deductible*	Prefix
	Some weeks → somes weeks	Inflectional suffix
Noncontextual errors (substitution, addition, deletion)	*Conclusion → concludement*	Derivational suffix
	To strain it → to strained it	Inflectional suffix
	He relaxes → he relax	Inflectional suffix
Word errors		
Misorderings		
Substitution		
Exchange	*Writing a letter to my mother → writing a mother to my letter*	Noun
Anticipation	*Sun is in the sky → sky is in the sky*	Noun
Perseveration	*Class will be about discussing the test → . . . discussing the class*	Noun
Addition	*These flowers are purple → these purple flowers are purple*	Adjective
Shift	*Something to tell you all → something all to tell you*	Quantifier
Noncontextual errors		
Substitution	*Pass the pepper → pass the salt*	Noun
	Liszt's second Hungarian rhapsody → second Hungarian restaurant	Noun
Blend	*Athlete/player → athler*	Noun
	Taxi/cab → tab	Noun
Addition	*The only thing I can do → the only one thing*	Quantifier
Deletion	*I just wanted to ask that → I just wanted to that*	Verb

Types of speech errors reproduced from Dell et al (1860) [24, p. 285]

According to the relation of the error to the target word, Dell distinguishes between semantic, formal, mixed, nonword, and unrelated errors. A semantic error is related to the target word through synonymic (e.g., toilet – commode), category coordinate (e.g., cat – dog), superordinate (e.g., apple – fruit), subordinate (e.g., flower – rose), associated (e.g., bench – park), or diminutive (e.g., dog – doggie) relations. Semantic errors might happen when concepts share common semantic nodes: for example, if the target word is "cat", during the lemma access stage the word node for "dog" is activated by their common semantic nodes and if its activation ends up being higher than that of other nouns, including "cat", it gets selected instead.

A formal error is phonologically similar to the target word. A formal error may occur at the level of either lemma or phonological access. During lemma access, feedback from the target phonemes to the word layer causes the activation of words that have common phonemes with the target. Therefore, if the target word is "cat", the words "sat" and "mat" would also become somewhat activated and could be selected if their activation surpasses that of the target. Form-related lexical errors labeled as malapropisms by Dell obey the constraint of the syntactic class while being phonological, which can also be explained by the fact that they represent errors of lemma access and, therefore, are syntactically governed.

A formal error can also occur during phonological access: for example, the target word "cat" may have been chosen at the lemma access stage, however, one or several of its phonemes may be substituted by other phonemes as a result of noise or influence of other activated words. The difference from the errors occurring during lemma access is that the phonological selection procedure is not influenced by the syntactic category constraint and is sensitive exclusively to the phonotactic constraints: the resulting sequence of phonemes may be a word that shares the syntactic category with the target (e.g., "mat"), as well as a word of a different category (e.g., "sat").

A mixed error satisfies the requirements for a formal error, while also displaying a semantic connection with the target. Such mixed semantic-formal errors (e.g., using “rat” for “cat”) indicate the combined impacts of semantic and phonological similarities and hence are particularly significant in demonstrating the assumption that semantic and phonological information are active concurrently. According to Dell, the word node for “rat” receives direct activation from shared semantic nodes, as well as feedback from shared phonemes. Moreover, compared to a solely semantic or formal neighbor, “rat” has a far greater probability of surfacing as an error because it combines top-down and bottom-up information, which is supported by the abundant evidence for the existence of the mixed-error effect.

Word substitutions that do not share semantic or formal connections with the target are defined as unrelated word errors. This type accounts for errors that display an observable distant relationship (for example, using “log” for “cat”, where “log” is connected to “cat” through a common neighbor “dog”), as well as errors where distant relations are less obvious. Although it is considered that unrelated errors arise during lemma access, they may potentially occur during phonological access due to a slim possibility of the target being correctly chosen during lemma access but phonologically encoded as an unrelated word. Conversely, this type of error may also indicate issues with both phonological and lemma access. For example, “dog” could be erroneously chosen instead of “cat” during lemma access and encoded as “log” during phonological access.

Finally, nonwords (e.g., “lat” or “cag” for “cat”) suggest issues with phonological access as they result from the erroneous replacement of one or several target phonemes. A target-related nonword (i.e., a nonword that resembles the target) indicates proper lemma selection but incorrect phonological access. On the other hand, an abstruse nonword (i.e., a nonword that does not resemble the target) may result from either a substantial disruption at the phonological level or issues with both lemma and phonological access [64, pp. 806-808].

Villers expanded the classification of speech errors depending on the size of the unit involved proposed by Dell, having described the category of phraseological errors. They are defined as "phraseological modifications that are the result of unintentional deviation and that do not match the attested form, usage, or meaning of a phraseme". As for the phraseme (or phraseological unit), this term is used by Villers to refer to preconstructed (or fixed) polylexical units [102, p. 109].

Phraseological errors are classified into substitutions, permutations, expansions, omissions, fusion ("blending"), inter-language calque, inappropriate contextual use, wrong interpretation, overuse, underuse, and mixed. Substitution is the most common type of phraseological error and can be exemplified by the idiom "social leper" being changed by the substitution of "leper" with the near homonym "leopard" ("social leopard"). Homonymy-based substitutions are dubbed "eggcorns" by Villers. Permutation can be viewed as a subtype of substitution and is evident when two terms are substituted with the reversal of order (e.g., "the pot calling the kettle black" – "look who's calling the pot black"). Expansion refers to words being modified into phrasemes and phrases elongated through the addition of letters or words (e.g., "scapegoat" – "escape goat"; "to exact revenge" – "to extract revenge"). Conversely, omission refers to the shortening of phrasemes (e.g., "for all intents and purposes" – "for all intensive purposes"). In the case of blending or fusion, two phrasemes are combined together (e.g., "kettle of fish" (a complicated situation) and "not my cup of tea" (something disliked by the speaker) – "not my kettle of tea"). Inter-language calque is a literal translation of a phrase from one language into another, where it does not occur (e.g., "to sell the bear's skin before you've killed it" from French "vendre la peau de l'ours avant de l'avoir tue").

Non-compositional phrasemes, the meaning of which can not be assumed from the sum of its components' meanings, are frequently misunderstood and misused, leading to acceptability errors. Inappropriate contextual usage can be exemplified by the idiom "off the top of the head" (without thinking) being used in the wrong context ("I know my lesson off the top of my head", where "by heart" would be more

appropriate). Wrong interpretation is the most common type of such acceptability error that can be explained by the absence of comprehension or erroneous comprehension and may involve semantic blending with a different phraseme or literal reading. Finally, hybrid errors present a combination of several types (e.g., "to step foot on" (from target "to set foot on") may be considered an addition, a blend of "set foot on" and "step on", as well as a substitution). Villers explains the occurrence of phraseological blunders by the notion of analogy or similarity of the target and the resulting blunder in terms of phonology, a lexical element, or meaning [102, p. 115].

1.2.2. Types of speech perception errors

Perceptual errors occur when the words are heard differently from what was intended by the speaker and are known as slips of the ear, mishearing, and misperceptions. Since most slips of the ear involve a word or a short phrase, they mostly entail a phonetic difference between the target utterance and the perception. This discrepancy offers an insight into the salience of certain features of phonetic structure and the phonology of casual speech [46, p. 269].

Some slips of the ear do not involve an incorrect perception of phonetic features. Instead, listeners correctly recover the phonetic sequence of the utterance but misinterpret it by selecting the incorrect homophone from the mental lexicon (for example, "courts" – "quarts"). This shows that listeners quickly make a connection between a phonetic form and the lexicon, with little to no interference from such factors as word frequency or context [104, pp. 423-424]. However, more widespread are misperceptions where the phonetic sequence is perceived accurately but a word boundary is added or omitted (e.g., Aleve → a leave, the ultimate in convenience → the ultimate inconvenience) [46, p. 270].

Vowel misperceptions may be associated with the phonetic structure or dialect variation, however, they rarely affect stressed vowels or prosodic patterns. These misperceptions usually involve vowel height and tenseness and are often associated with consonants that affect vowel quality, such as nasals and the liquids /l, r/. Typical examples are Elf → Alf, Jane → Jean, wool → . . . wall, snow pea → Snoopy, do

some laminating → lemon eating. However, the overall stress pattern and associated vowel quality are generally resistant to misperception, which suggests that they serve as a primary scaffolding for other elements of speech.

On the other hand, consonants are more likely to be involved in a misperception error in comparison to vowels. Listeners may either not perceive consonants, report false consonants or misperceive consonant quality. Although consonants may be omitted in any position within a word, the final consonant omission is the most widespread in American English, while speakers tend to use less force in the articulation of word endings (e.g., card – car, air assault – aerosol). Similarly, wrong consonants may also occur in any word position (e.g., slip of the ear → slip of the year, doggie → donkey, red spire pears → red spider pears). Consonant quality misperceptions are the most common, with any of such traditional features as manner, place, or voicing undergoing substitution. Obstruents tend to be replaced by obstruents, and resonants tend to be replaced by resonants. Some consonant misperceptions are dependent on phonetic factors, while others are related to misperceptions of word boundaries.

As opposed to homophones, which involve the correct perception of phonetic structure but result in incorrect lexical retrieval, there are misperceptions in which a significant amount of phonetic information is reported inaccurately (e.g., post doc → co-star, Q-Tip → toothpick). Although there can be significant discrepancies between spoken utterances and their perceived versions due to misperceptions, the perceived utterances tend to follow the phonological rules of the language, and the majority of misperceptions involve English vowels and consonants. Likewise, the phonotactic syllable structure in misperceptions aligns with the syllable structures of English. Moreover, when presented with sequences of consonants not acceptable in English, listeners often report the utterance in the form of a permissible sequence (e.g., tlumbering → klumbering). Comparing intended utterances with perceived ones can reveal stable phonetic properties, with vowels and prosodic patterns tending to be

perceived accurately due to their usefulness in identifying likely words in the lexicon [46, pp. 271-274].

At the lexical level, target words and their misperceptions have been shown to share many characteristics, such as the number of phonemes and syllables, word frequency, neighborhood frequency, density, etc. According to Vitevitch, this suggests that in case of a partial or inaccurate input, the processing system chooses the most suitable representation instead of coming to a halt [104, p. 416].

The lexicon does not restrict the possible perceptions, which is evidenced by listeners reporting nonwords (for example, “thone” for “phone”, and “the yarticle” for “the article”). This ability of listeners to quickly encode and report nonwords supports the theory that lexical representations comprise segments. Moreover, some instances of misheard speech suggest that the segment order from the target word to the perceived word is not fixed. Segment order errors (e.g., “throfing” for “frothing”, “pound of coffee” for “found a copy”) can be seen as a result of the erroneous use of overlapping articulatory information for consonants and vowels.

Due to function words being typically pronounced with reduced stress and showing considerable variation in pronunciation, they also exhibit little stability in slips of the ear, often resulting in word-boundary misperceptions and word substitutions. For instance, “I don't intend to stay in the picture” may be misheard as “...to stain the picture,” and “this friend of ours who visited” may be misheard as “...is an idiot.”

Slips of the ear that involve syntax are relatively uncommon since they usually seem to be well-formed. However, there are instances where listeners report ungrammatical sentences due to misinterpretation of the intended part of speech of a word (e.g., “we offered six” being misheard as “we Alfred six”). Slips related to the syntactic structure are often accompanied by modifications of function words to preserve well-formedness (e.g., “She wants to be a teacher” → “She wants me to teach her”) [46, pp. 278-279].

In most cases of misperception, the intonation and sentence stress patterns remain stable. However, this is not the case with the grammatical function of sentences, even though it is connected to intonation, and it may be involved in misperceptions that do not align with the intended meaning of the speaker's utterance. For instance, declarative sentences can be interpreted as interrogatives ("Islamabad" → "Is his Lama bad?"), or interrogatives can be reported as one word ("Where are my clothes? → germaphobe"). Listeners often report hearing incongruous and unexpected utterances (e.g., "immense ethereal gulf" → "immense empirical goat"), which suggests that semantics and pragmatics allow for almost any interpretation of an utterance. However, listeners also tend to attempt to make sense of what they hear by identifying the ambiguous input as the most semantically appropriate or reinterpreting it [39, p. 232]. Some slips of the ear, especially in the case of proper names, involve words from the same semantic domain as the target, even though they may demonstrate substantial phonological differences (e.g., Stockholm → Scotland, pathology → psychology). This may be explained by the fact that the general context of the conversation provides information for listeners to form expectations about what is being said [46, pp. 280-281].

A special type of perception error is presented by the mondegreen, a mishearing or misinterpretation of song lyrics. The term was coined by American writer Sylvia Wright who misheard ballad lyrics "And laid him on the green" as "lady Mondegreen". A mishearing of song lyrics occurs due to a variety of reasons, such as the noisy environment, lack of visual cues, and changes in usual stress and inflection patterns. According to Connor, when listeners are not able to decipher what they hear they strive to fill in the gaps by making assumptions, with mondegreens representing the "wrenchings of nonsense into sense" [21, pp. 3-4]. Mondegreens are often caused by oronyms, word strings that can be logically divided multiple ways, leading to incorrect parsing of sounds. Similar-sounding letters and letter combinations can also cause confusion, as shown by the McGurk effect described earlier, which is inhibited by the absence of contextual cues in songs and poems. The

frequency of words and the listener's expectations can also affect the processing of lyrics: listeners are more likely to select a more familiar or probable word or phrase [57].

As evidenced above, slips of the ear provide evidence for the same phenomena as speech errors. For instance, the fact that slips of the ear often involve a single segment or feature indicates that speech is divided into units during both speech perception and speech production. Moreover, since speech misperceptions are always represented by possible (if not existing) words, it can be assumed that listeners possess unconscious knowledge of phonotactic rules. Finally, the fact that many speech misperceptions involve the mix-up of word boundaries shows that, like speakers, listeners divide speech into morphemes.

1.3. Verbal blunders vs malapropism

Malapropism occupies a prominent but controversial place within the discussion and research on speech errors, with its definitions often drastically differing. In their influential work on malapropisms, Fay and Cutler define them as word substitutions errors that involve the intrusion of a real word unrelated to the target in meaning but closely related in pronunciation and resulting from the unintentional blunder instead of the "ignorance of the correct usage" [28, p. 505]. However, in most definitions, "classical malapropisms" differ from the so-called "Fay-Cutler malapropisms" in that they actually involve a competence error instead of the performance one, i.e., speakers have an incorrect idea about the meaning or accepted usage of a word and the error is intended by them. In fact, malapropism is named after a character from Richard B. Sheridan's play *The Rivals*, Mrs. Malaprop, who was prone to making this type of error [100, p. 323]. Moreover, contrary to Fay and Cutler's description of malapropisms, Zwicky found that malapropisms are not exclusively represented by existing words (e.g., "vanishment" instead of "banishment", "bastie" instead of "bastard") [115, p. 345].

Other characteristics attributed to malapropisms are their humorous effect (e.g, using "condoms" instead of "condiments") and occurrence as a result not only of

ignorance but also vanity, i.e., “a speaker’s attempt to rise above his lexical station” (e.g., indulging in legalisms without sufficient knowledge of them by saying “to revolt racing licenses” instead of “revoke”). The sample of malapropisms analyzed by Zwicky indeed shows the prevalence of humorous examples and frequent occurrences in contexts where speakers are prompted to reach for vivid or technical vocabulary. However, Zwicky concludes that although these two properties can be attributed to many examples of malapropisms, they can not be used to characterize classical malapropisms in general. For instance, he explains the bias towards humorous malapropisms by the fact that they are more likely to be noticed, remembered, and, consequently, recorded. Moreover, almost a quarter of malapropisms in the sample involved untechnical vocabulary of Anglo-Saxon origin, disproving the idea that malapropisms arise exclusively from the speaker’s attempt to surpass their abilities in using fine language [115, pp. 340-341].

Zwicky identifies three main sources of classical malapropisms: childhood slips of the ear that remained uncorrected, “reanalysis of the folk-etymological variety”, i.e., replacement of an unfamiliar form by a more familiar one (e.g., “Cynical Hotel” instead of “Seneca Hotel”), and faulty storage and retrieval within the mental lexicon. While the first two sources involve imperfect learning, the third type entails the speaker selecting the wrong item from the mental lexicon without realizing it and accepting it in place of the target, which leads to the inaccurate utterance becoming the stored item. The original retrieval error occurs because of the defective or incomplete storage of the target and it eventually secures the erroneous item in the speaker’s mental lexicon. Therefore, malapropisms reveal more about the creation and maintenance of the mental lexicon than about its organization [115, p. 347; 100, p. 323; 114, p. 129].

Chapter conclusion

This chapter of our study focuses on providing a comprehensive outline of available findings and theories concerning psycholinguistic phenomena that could shed light on the underlying causes and processes behind speech errors.

Firstly, it was of utmost importance to consider the concept of the mental lexicon, since it is universally understood as a sort of mental dictionary that contains information about the word stock of a speaker and is represented by a complex network of interconnected entries (or nodes). Research suggests that the mental lexicon is organized according to both semantic and lexical properties and stores both complete forms and separate morphemes. The enrichment of the mental lexicon is conducted through fast and slow mapping of word forms to meaning. The size (also known as vocabulary breadth) and quality (vocabulary depth) of the mental lexicon depend on various factors and differ among the speakers.

Word recognition occurs through a combination of bottom-up processes, triggered by the auditory signal and top-down processes activated by linguistic context. Moreover, psycholinguistic experiments have revealed that spoken word recognition is incremental in that several nodes stored in the mental lexicon and compatible with the auditory information get activated simultaneously before the utterance is finished. The speed of lexical decisions is heavily dependent on the number of words phonologically related to the target, which is known as the neighborhood effect, as well as on the frequency of the target itself. Furthermore, the concreteness effect means that easily imageable words are processed with greater speed and accuracy than abstract words. Finally, due to real words being fixed in the mental lexicon, they are processed faster and more accurately than non-words (lexical status effect), while at the same time, non-words that conform to existing word patterns are processed better than those that do not (non-word legality effect).

It is universally recognized that breakdowns in the system reveal its inner mechanisms and, therefore, the study of speech errors is of paramount importance within the general investigation of speech processing and production. Word retrieval models depict the lexicon as a network of three levels: conceptual, lemma (word), and phonological. After the identification of the lexical concept, language production occurs in two stages: the selection of the lemma and the ordering of the word form. The tradition of speech error analysis has led to the emergence of speech production

models that argue for the interactivity of the aforementioned stages. According to the highly influential interactive activation model, all connections in the system are bidirectional, and processing at lower levels can impact that of higher levels by means of feedback.

The studies on individuals with aphasia and language impairments have lent a considerable number of findings regarding the mental lexicon and led researchers to distinguish between deficits in storing information in the mental lexicon and deficits in accessing information for word production. Moreover, the good-enough language processing theory postulates that even healthy adults have an intrinsic propensity for superficial analysis of linguistic input, which may result in its erroneous interpretation.

As for the probability of a speech error occurring, it depends on a number of influences, namely output biases (the propensity to result in plausible or frequent sequences), similarity effects (tendency to involve similar words), speech rate effects, and distance effects (the tendency of the involved items to move over short distances). Moreover, the neighborhood effect has been an object of numerous studies which suggest that words with large phonological neighborhoods have a tendency to be produced more accurately and efficiently than those with small neighborhoods.

Considering the different processes behind word production and word recognition, speech errors themselves should be considered from two perspectives. Speech production errors are defined as involuntary linguistic innovation deviating from the speaker’s intention. In the late 19th century, Meringer and Mayer laid the foundation for subsequent speech error research by introducing the now universal distinction between form-based and meaning-based errors, as well such standard categories of speech errors as anticipations, perseverations, exchanges, and blends (or contaminations). Since then, multiple new classifications were introduced by other linguists and psycholinguists, expanding on the initial scholarly contributions. For instance, Fromkin adds such categories as substitution, transposition (metathesis), omission, addition, and movement. On the other hand, Dell proposes a typology

based on various factors, such as the size of the unit involved, the type of disruption, and the relation of the error to the target word. Furthermore, Villers describes a separate category of phraseological errors in more detail by classifying them into substitutions, permutations, expansions, omissions, fusion ("blending"), inter-language calque, inappropriate contextual use, wrong interpretation, overuse, underuse, and mixed.

As for speech perception errors, they occur when the words are heard differently from what was intended by the speaker and usually involve a phonetic difference between the target utterance and the perception, offering an insight into the importance of certain phonetic features during speech processing. For instance, vowel misperceptions rarely affect stressed vowels or prosodic patterns. On the other hand, consonants are more likely to be involved in a misperception error, with listeners either not perceiving consonants, reporting false consonants, or misperceiving consonant quality. However, slips of the ear do not always involve the incorrect perception of phonetic features but rather the selection of the incorrect homophone from the mental lexicon or misperception of the word boundary. In most cases of misperceptions, the intonation and sentence stress patterns remain stable. A special type of perception error is a mondegreen, a mishearing or misinterpretation of a song lyric, which is often caused by oronyms, or word strings that can be logically divided in multiple ways, and is exacerbated by the absence of contextual cues. Overall, the psycholinguistic evidence obtained from the slips of the ear aligns with what is known from studying speech production errors: speech is divided into units during both speech perception and production and listeners possess unconscious knowledge of phonotactic rules.

When discussing speech errors, it is necessary to consider malapropisms, which present a unique case in that they represent a competence error instead of a performance error, i.e., they are intentional and are caused by the speaker's incorrect idea about the accepted usage of a word. Malapropisms are often accompanied by a humorous effect and in many cases represent a speaker's attempt to exceed their

abilities in exercising their fine language skills. The main sources of malapropisms are uncorrected childhood mishearings, reanalysis, and faulty storage and retrieval within the mental lexicon. Thus, the study of malapropisms may be highly beneficial in investigating the processes behind the creation and maintenance of the mental lexicon.

CHAPTER II. RESEARCH DESIGN

2.1. Quantitative approaches to the investigation of eggcorns

2.1.1. Computational and corpus-based approaches

Computational linguistics and corpus linguistics are two closely related fields that are concerned with the study of natural language. Both disciplines heavily rely on electronic corpora to extract linguistic information and employ similar methods to acquire this information. Computational linguistics focuses on formal modeling of natural language, which is used to make conclusions about the structure and functioning of language. In particular, computational phonology, the elements of which were employed in this study, uses computational methods to investigate the sound patterns of a language. This approach involves the analysis of large datasets of spoken and written language to identify patterns in sound distribution and organization, as well as to investigate the ways in which they are influenced by such factors as phonotactics, stress, and intonation.

The methods used in computational linguistics are diverse and are divided into two branches: the rationalist branch, which employs non-statistical and theory-driven methods, and the empiricist branch, which focuses on corpus-driven and statistical techniques. The corpus-driven branch of computational linguistics is naturally connected to corpus linguistics, as both fields are interested in investigating corpora. Consequently, numerous research topics can be attributed to both computational linguistics and corpus linguistics with equal relevance [26, p. 68].

Corpus linguistics is based on the notion that language development is realized through the process of communication and, therefore, should be considered in the context of usage. In relation to corpus linguistics, a corpus is defined as an electronic body of texts used to select and analyze linguistic elements. The advantage of corpus linguistics methodology is that the data obtained may be considered authentic since it is collected from natural sources that provide unmodified examples of language usage [68, p. 3].

In the present study, we have opted for a combination of computational and corpus-based methods, using Phonological CorpusTools and the Irvine Phonotactic Online Dictionary (IPhOD) to extract data on string similarity, phonotactic probability, neighborhood densities, and frequencies of elements involved in eggcorns, as well the English Lexicon Project to obtain data on words' neighborhood frequencies.

The analysis was conducted on a sample of 100 eggcorns selected from the Eggcorn Database, an online dictionary collecting language errors of this type [119]. The selection of the eggcorns was conducted by using a random sampling technique, which involved the arbitrary collection of items from a larger database of eggcorns, ensuring that each item had an equal chance of being selected. This sampling technique allowed us to obtain a representative sample of eggcorns from the database and, consequently, to draw meaningful conclusions about the characteristics and distribution of eggcorns in general. Furthermore, the use of the random sampling technique ensured the elimination of biases or patterns in the selection process, and, hence, the reliability of the sample for research purposes.

As for the tools that were used in the course of our analysis, Phonological CorpusTools (PCT) is an open-source software package that allows performing phonological analysis on transcribed language data and offers a range of capabilities, such as calculating the phonotactic probability of a word or determining the similarity between words using either orthographic or phonetic transcription [120].

Phonotactic probability refers to the likelihood of a particular sequence of sounds occurring in a given order within a corpus of transcriptions. This measure can be applied to both words and nonwords, and it has been utilized in various behavioral studies to explore the relationship between phonotactic probability and language processing. In particular, the phonotactic probability of words has been linked to their capacity to be acquired, processed, and produced [107]. The calculation of phonotactic probability can be done using various methods, however, the algorithm implemented in PCT employs average unigram or bigram positional probabilities across a word. The unigram average considers the likelihood of each segment occurring in each position within the word, while the bigram average looks at the probability of sequences of two segments and their positions. By using this algorithm, we were able to calculate and compare the phonotactic probability of target and error words (including non-words).

String similarity is a measure of the likeness between two sequences of characters, which can be composed of letters or phonemes. The Phonological CorpusTools (PCT) provides both methods for calculating string similarity, making it a fundamental tool for measuring form-based similarity. While string similarity is commonly used in various fields of linguistics, such as Natural Language Processing, where it can help determine possible alternative spellings for mistyped words, it is also useful for calculating the phonological proximity between two words. In this research, we opted for the phonological edit distance method for measuring string similarity. Phonological edit distance calculates the number of operations required to transform one word into the other, while also considering featural similarity. The result is a numerical value that represents the distance between the two words [45]. This feature of PCT was utilized during the research to identify patterns in the similarity between target and error words involved in eggcorns.

The software utilizes pre-existing publicly available corpora, as well as provides the option of uploading custom corpora or analyzing words not present in the corpus by inputting its transcription. For the purposes of our research, we opted

for the built-in Irvine Phonotactic Online Dictionary (IPhOD) corpus. IPhOD is a database of English words and pseudowords for speech perception and production research. It allows performing searches for word characteristics such as frequent or infrequent sound sequences and similar-sounding words [117].

The current IPhOD version contains phonotactic estimates, which are calculated on the basis of an extensive sample and can be used for speech research, as well as within the field of cognitive science, computational linguistics, and natural language processing. Moreover, it includes density estimates and word frequencies for over 54 000 words and 814 840 pseudowords. Each IPhOD entry includes an American English phonetic transcription and written word frequency.

The IPhOD database also provides the function of online search using value ranges or word lists, which was used as part of our research to determine neighborhood density values for target and error words. A word's neighborhood density, which has been found to impact phonological processing, is determined by measures of string similarity: a phonological neighbor of a given word is a word that closely resembles it, differing by one phoneme at most (for example, through addition, deletion or substitution). Phonological neighborhood density is determined by counting the number of words in the corpus that meet the criterion for being a neighbor. It is important to mention that neighborhood density was calculated using the CMU Pronouncing Dictionary, created by Carnegie Mellon University for speech recognition research and utilizing a transcription system different from the IPA format (for example, the word "car", which is transcribed as /K.AA.R/ has 38 neighbors, some of which are "card" /K.AA.R.D/, "cause" /K.AA.Z/ and "our /AA.R/). Moreover, the Irvine Phonotactic Online Dictionary features an online calculator that generates phonotactic and density values for user-entered phonemic transcriptions, which can also list the phonological neighbors of each input transcription. The online calculator was particularly useful for generating values for words and non-words that were not found in the IPhOD database.

Furthermore, the IPhOD database was used to obtain written word frequency counts, since it has been shown that cognitive processing heavily relies on word frequency, with high-frequency words being processed and produced more efficiently than low-frequency words. The frequency measures were extracted from the SUBTLEXus database, which is a corpus for American English based on subtitles for movies and TV shows and containing frequencies for 51 million words in total [14].

Finally, the English Lexicon Project (ELP) is a comprehensive database of over 40,000 English words and their psycholinguistic properties, developed for use in language and cognitive research [118]. The tools provided by the English Lexicon Project were used in the present research to determine the neighborhood frequency (i.e., the mean frequency of the word's neighbors) of target and error words in eggcorns.

2.2. Quantitative approaches to the investigation of eggcorns

2.2.1. Descriptive qualitative approach

Qualitative description involves systematically examining and categorizing data based on observable characteristics, and it is commonly used in linguistic research to identify patterns in language usage. It aims to ascertain the essence of the phenomena studied, as well as provide its objective description. The data analyzed with the help of this approach should be in its natural and original form and should not involve alterations or adjustments of any kind. The principle of organizing data is chosen by the researcher; however, it should be done in a logical manner [59, pp. 255-256].

The analysis of collected data is conducted primarily in a qualitative fashion by detecting key themes and patterns and consequently providing their description and interpretation. After examining data in a purely qualitative way, it may further undergo quantitative analysis to systematize the observations made and present them statistically for the purposes of convenient comparison and further assessment [82, pp. 129-130].

Within the framework of the qualitative descriptive approach, we provided the outline of the formal characteristics observed in eggcorns, and their further quantitative systematization and categorization according to the aforementioned characteristics. Additionally, the comparative analysis of the etymology of targets and errors added depth to the understanding of possible motivations behind the occurrence of eggcorns. Finally, we provided a description and categorization of eggcorns according to their defining semantic characteristics, providing further insight into the different types of eggcorns and how they relate to their intended meanings.

2.2.2. Functional semantic field approach

According to the semantic field theory, vocabulary can be classified into clusters of lexemes linked by common semantic properties. Despite the fact that the lexemes in a semantic field describe the same aspect of reality and are interrelated to a certain degree, synonymity is not a required criterion [69, p. 237].

The functional semantic field approach aims to analyze how language is used in its social context and involves the identification of patterns of language use in specific contexts, as well as the investigation of how those patterns relate to the social practices of the speech community. During the course of analysis, the lexemes describing particular phenomena of reality are assembled into semantic fields, after which they are examined in terms of their functions in communication. Thus, the semantic structure of vocabulary becomes apparent. Overall, this approach allows us to consider vocabulary as part of its natural context, as well as observe how it is used to express extralinguistic realities [54, p. 146].

We have utilized the functional semantic field approach to categorize eggcorns into semantic fields, which involved analyzing the specific meanings and semantic connotations of the words involved in the eggcorns. This analysis allowed us to identify common patterns and associations that contributed to the occurrence of eggcorns in certain semantic fields. Due to its focus on the functional use of language

and the relationship between language and the world, this approach enabled us to gain insight into the underlying cognitive processes involved in eggcorn formation.

Chapter conclusion

In this chapter, we have discussed several qualitative and quantitative approaches adopted to ensure the consistency and comprehensiveness of the research. A combination of computational and corpus-based approaches was used by us to analyze 100 eggcorns selected from the Eggcorn Database through random sampling. Computational linguistics and corpus linguistics are two closely associated fields that study natural language using electronic corpora to extract linguistic information. Computational linguistics uses formal modeling to investigate language structure and functioning. In this study, we used elements of computational phonology, which aims to identify patterns in sound distribution and organization by using computational methods). Regarding corpus linguistics, it focuses on language in the context of its use and utilizes authentic data collected from natural sources to obtain unmodified examples of language usage. It is important to note that the empiricist branch of computational linguistics is inextricably linked with corpus linguistics since both fields rely on the investigation of corpora. In this study, the Phonological CorpusTools software package and the Irvine Phonotactic Online Dictionary were used to extract data on string similarity, phonotactic probability, neighborhood densities, and frequencies of elements involved in eggcorns, while the English Lexicon Project was used to obtain data on words' neighborhood frequencies.

The descriptive qualitative approach was used to provide a thorough account of the formal and semantic characteristics, as well as other factors that might influence the occurrence of eggcorns, such as the etymological origins of target-error pairs. According to this approach, the collected data is analyzed qualitatively to identify key themes and patterns and then may undergo quantitative analysis for comparison and assessment.

Finally, the functional semantic field approach aims to categorize vocabulary into groups of lexical items linked by common semantic properties and identify

patterns of language use in specific contexts. Using this approach, eggcorns were categorized into semantic fields, providing further insight into the underlying cognitive mechanisms and motivations that lead to the creation of eggcorns.

CHAPTER III. LINGUISTIC PROPERTIES OF EGGCORN AS A SPEECH ERROR

3.1. Defining eggcorns

The term "eggcorn" is a relatively recent addition to the lexicon pertaining to the study of speech errors and, therefore, its definition and the description of features that a word must possess in order to be considered an eggcorn still exhibit a certain degree of variability, while its relation to other terms in the realm of linguistic errors continues to be debated by linguists. The term was coined in 2003 by professor of linguistics Geoffrey Pullum after an eponymous case of a speaker using the non-word "eggcorn" instead of "acorn" [66]. It refers to a type of linguistic error that involves the substitution of a word or a phrase with one that is phonetically similar and can be justified semantically.

Eggcorn differs from other speech production errors in that the speaker does not realize the error is being made. In this sense, eggcorns are similar to malapropisms. However, unlike malapropisms, eggcorns also display semantic logic apart from phonological resemblance. For instance, in the example that lent the term its name, an association between an egg and an acorn can be drawn due to their similar shape, while an acorn also resembles a corn kernel. The semantic link can be drawn either from the word's perceived meaning or origin, as in the above example, or from the context in which the word is used.

Moreover, while speakers whose utterances result in malapropisms are often striving to use sophisticated vocabulary without possessing sufficient fluency in it, eggcorns are usually coined by individuals who aspire to a high level of vocabulary or sophistication but instead are common in pop culture, cliches, and business jargon.

Those who use eggcorns are not usually trying to appear literate or sophisticated and often use them when writing to friends or on social media. Eggcorns often appear when the speakers attempt to use phrases, the written form of which they have never encountered, and they are often made by amateurs such as fan-fiction writers, gamers, and bloggers. Nevertheless, there is no consensus as to whether eggcorns are a completely separate type of linguistic error or a subtype of a malapropism (for instance, R.A. Rubin defines them as "quintessentially modern and democratic malapropisms") [90, p. 9].

The existence of a semantic justification allows us to draw a comparison between eggcorns and folk etymologies, errors due to incorrect interpretation of archaisms and borrowings. However, the main difference between these phenomena is that folk etymologies get adopted by entire speech communities, while eggcorns are made by individual speakers [87, p.17].

In view of the above, this section of the paper will focus on the analysis of eggcorns from the standpoint of its defining characteristics, including formal and semantic aspects.

3.2. Formal aspect of eggcorns

3.2.1. Neighborhood characteristics of eggcorns

Having analyzed a sample of 100 eggcorns taken from the Eggcorn Database according to their neighborhood characteristics, we have established that the average phonological neighborhood density of the target words was 15.18, while the average phonological neighborhood density of the resulting errors was 13.74 (see Appendix 1 for the complete research data). This suggests that, on average, the target phrases had a larger number of neighbors differing in one phoneme than the resulting errors. Moreover, in most cases, the density of the target was higher than that of the error. For example, in an eggcorn "take another tact", despite itself being a phonological neighbor of the error word "tact" /T.AE.K/, the target word "tack" /T.AE.K.T/ has a much higher neighborhood density (46 neighbors compared to 26 neighbors of "tack"). On the other hand, in 24 cases the opposite was true, i.e., the density of the

target was lower than that of the error. For instance, in such target-error pairs as "naught"-"knot" or "beck"-"back", the number of phonological neighbors of the error was considerably higher despite the significant phonological similarity with the target (error word "knot" /N.AA.T/ had 34 neighbors compared to 26 neighbors of the target "naught" /N.AO.T/ while "back" /B.AE.K/ had 47 compared to 33 of the target "beck" /B.EH.K/).

Overall, this data contradicts the phonological neighborhood effect observed in speech production errors, according to which words with larger phonological neighborhoods have a tendency to be produced more accurately and efficiently than those with small neighborhoods. On the other hand, it is consistent with the results of speech recognition studies, which demonstrate that words with large phonological neighborhoods are recognized slower and less accurately than words with small phonological neighborhoods. This suggests that eggcorns, similarly to malapropisms, are not performance errors but rather result from faulty storage in the mental lexicon, which may be a consequence of an uncorrected slip of the ear.

However, the density of both the target and error was equal in 26 eggcorns, and the density was equal to zero in both in 11 eggcorns (the remaining 10 instances were excluded from the analysis on the basis of them representing word collocations or phrases instead of words). For example, both words from the "ruckus"-"raucous" pair had 2 neighbors and were mutual neighbors of each other (i.e., phonological neighbors of "ruckus" /R.AH.K.AH.S/ are "raucous" and "rucks" /R.AH.K.S/, while phonological neighbors of "raucous" /R.AH.K.AH.S/ are "caucus" /K.AO.K.AH.S/ and "ruckus"). On the other hand, in pairs like "dander" /D.AE.N.D.ER/ -"dandruff" /D.AE.N.D.R.AH.F/, there were no attested neighbors for both words. These results, together with a relatively small disparity in average densities of targets and errors, indicate that the tendency to produce eggcorns is not significantly affected by the phonological neighborhood density, and the influence of other factors, such as frequency and semantics, is more pronounced.

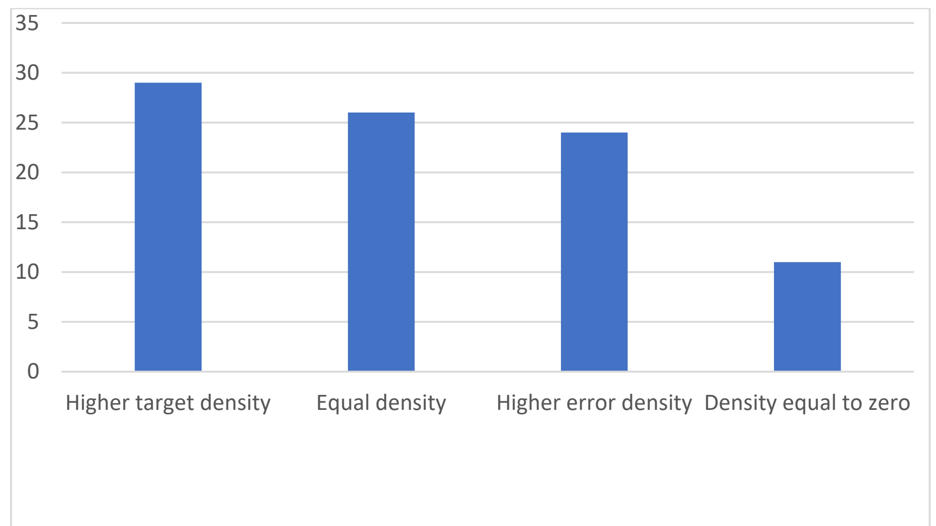

Chart 3.1. Comparison of target and error neighborhood densities

Furthermore, the average neighborhood frequency of targets, which is defined as the frequency of the word's phonological neighbors, was slightly higher (6.97) than that of errors (6.91) (see Appendix 1 for the detailed data). For example, in the eggcorn "in high dungeon", the target word "dudgeon" /D.AH.JH.AH.N/ has three phonological neighbors ("dungeon", "dozen" /D.AH.Z.AH.N/, and "gudgeon" /G.AH.JH.AH.N/), the mean frequency of which is equal to 4.043, while the error word "dungeon" /D.AH.N.JH.AH.N/ has two phonological neighbors ("dungeons" /D.AH.N.JH.AH.N.Z/ and "dudgeon"), the mean frequency of which is equal to 2.708. Since words with higher neighborhood frequency are likely to be recognized slower and less accurately than words with low neighborhood frequency, this also suggests the existence of a link between speech perception errors and eggcorns.

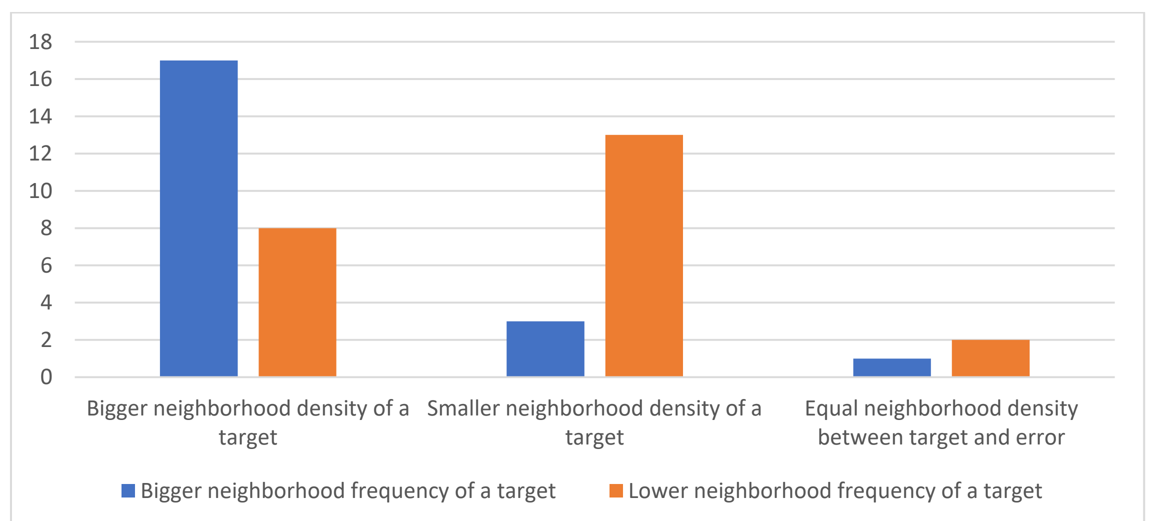

Chart 3.2. Correlation of neighborhood density and frequency in targets and errors

Furthermore, when considering neighborhood frequency in conjunction with neighborhood density, an interesting tendency can be observed, as presented in the chart above: in cases where the data on the neighborhood frequency of both the target and the error was available, there was typically a positive correlation between neighborhood frequency and density. Namely, targets that had higher neighborhood frequency than errors also had more phonological neighbors significantly in more instances (17) than less phonological neighbors (3) or the same number of neighbors as the corresponding error word (1). For example, in the target-error pair "wing" - "whim" from the eggcorn "on a whim and a prayer", the target "wing" has 42 neighbors, the mean frequency of which is equal to 8.006, while the error "whim" has 24 neighbors and the neighborhood frequency equal to 7.730. Likewise, targets that had lower neighborhood frequency also tended to have fewer phonological neighbors than the corresponding error (13 instances) rather than more (8) or an equal number (2). For example, in the target-error pair "cart" -"cat" from the eggcorn "put the cat before the horse", the target "cart" has 24 neighbors and neighborhood frequency equal to 7.794, while the error "cat" has 45 neighbors and neighborhood frequency equal to 8.452. This suggests that the more neighbors the word involved in an eggcorn possesses, the greater the likelihood that they will be of higher frequency. Nevertheless, taking into account a fairly negligible disparity between the averages, it can be assumed that the significance of the neighborhood frequency in the occurrence of eggcorns is not as tangible as that of the other factors.

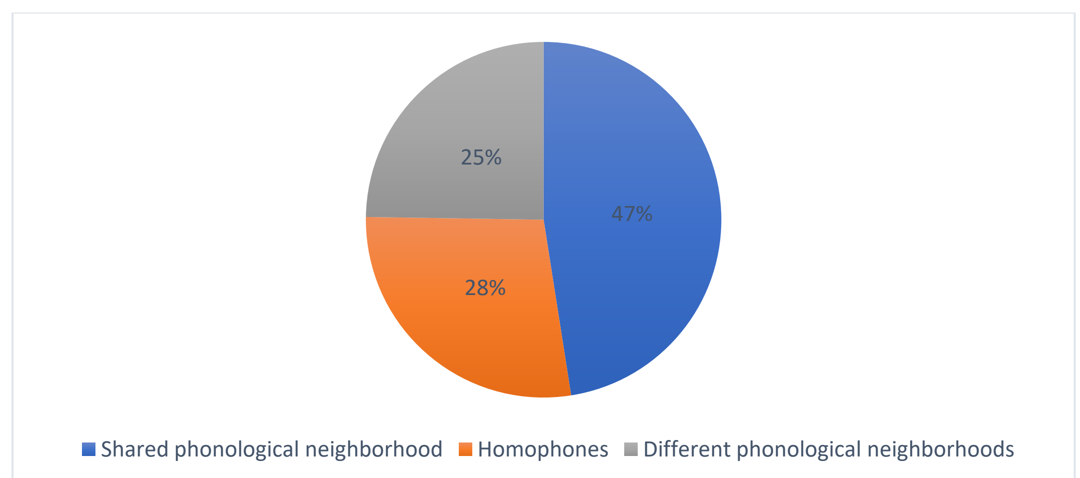

Chart 3.3. The phonological resemblance of targets and errors

On the other hand, the investigation into the influence of the phonological resemblance showed a strong tendency for the target word and the resulting error to share one phonological neighborhood. Moreover, there was also an observable tendency for the target and error to share identical pronunciation. As can be seen from the chart displayed above, 47% (48) of the analyzed pairs were phonological neighbors, while 28% (28) were homophones. For example, word pairs like “squib” /S.K.W.IH.B/ – “squid” /S.K.W.IH.D/, “defuse” /D.IH.F.Y.UW.Z/ – “diffuse” /D.IH.F.Y.UW.Z/, “locus” /L.OW.K.AH.S/ – “locust” /L.OW.K.AH.S.T/, and “windfall” /W.IH.N.D.F.AO.L/ – “winfall” /W.IH.N.F.AO.L/ are all examples of phonological neighbors, since they differ only by one phoneme, while target-error pairs like “doe” - “dough” /D.OW/, “hierarchy” – “higherarchy” /HH.AY.ER.AA.R.K.IY/, “star-craving” – “stark-raving” /S.T.AA.R.K.R.EY.V.IH.NG/, and “beck and call” – “beckoned call” /B.AE.K.AH.N.D/ are examples of homophones, since despite different spellings they share the same pronunciation. This suggests that the phonological similarity between words plays a role in the occurrence of eggcorns, as well as that phonological neighbor pairs are more likely to result in eggcorns than non-neighbor pairs.

Moreover, a large number of homophones among the target-error pairs can be interpreted in favor of the theory on the origin of malapropisms according to which they might appear from uncorrected slips of the ear. As has been stated in the first chapter of this research, some slips of the ear indeed involve the incorrect selection of a homophone from the mental lexicon rather than the erroneous perception of phonetic features. Therefore, the phonology of phrases like “beck and call”, “hair’s breadth”, “with bated breath”, as well as words like “shoo-in” and “pidgin”, could be correctly perceived by the speaker, who due to unfamiliarity with the words and opacity of the phrases’ meaning could nevertheless easily interpret them as more familiar homophones (“beckoned call”, “hare’s breadth”, “with baited breath”, “shoe-in” and “pigeon” respectively) and record as such in their mental lexicon. Moreover,

this theory is consistent with the fact that eggcorns often occur when the speaker is not familiar with the written form of the correct word or phrase.

Finally, the category of pairs that were not phonological neighbors (i.e., the ones that differed by more than a single phoneme) was the least represented, comprising 25% of the total number of pairs. Among them are word pairs such "crux" /K.R.AH.K.S/ – "crutch" /K.R.AH.CH/ ("crutch of the matter"), "exact" /IH.G.Z.AE.K.T/ – "extract" /IH.K.S.T.R.AE.K.T/ ("extract revenge/vengeance (on someone)"), and "dander" /D.AE.N.D.ER/ – "dandruff" /D.AE.N.D.R.AH.F/ ("get one's dandruff up"). As can be seen, despite representing the category of the most phonologically differing target-error pairs, these examples still share a certain number of phonemes. Hence, the effect of phonological similarity was explored in greater detail by considering the phonological distance between the pairs of targets and corresponding errors.

The average phonological distance between target and error words (including homophones) was 4.89. Excluding homophones, the phonological distance of which is equal to zero, the average phonological distance between target and error words was 6.76 (e.g., the distance between "foment" and "ferment" in "ferment trouble/fears/unrest" was equal to 6.75, between "rote" and "route" ("learn by route") – 6.5, and between "desist" and "decease" ("cease and decease") – 7). The smallest distance observed was 1 (e.g., in such phonological neighbors as "defuse" – "diffuse" ("diffuse the situation"), "ectopic" – "eggtopic" ("eggtopic pregnancy"), and "foolproof" – "fullproof"), while the biggest was 20.5 ("intents and" – "intensive" in "for all intensive purposes"). While not being particularly high, the average phonological distance suggests that the errors are not simply the result of minor phonological confusions and are influenced by other factors aside from the phonological resemblance. Nevertheless, the frequency distribution of the phonological distance data, which can be observed in Chart 3.3, shows that the majority of the target-error pairs have a phonological distance between 5 and 8.99 (31), which is followed by the group of homophones with a distance of 0 (28), and by

pairs having a distance between 1 and 4.99 (23). This suggests that in eggcorns the phonological distance between the target and error tends to be relatively small. Moreover, the occurrence of errors decreases with the increasing distance between the target and the error, which is demonstrated by the decreasing values in the higher ranges of phonological distance. For instance, there were 12 pairs with the phonological distance between 9 and 12.99 (such as “jaw” – “jar” in “jar-breaking” (9), “cahoots” – “cohorts” (10.75), “corroborate” – “collaborate” (11.5), and “augurs” – “all goes” in “all goes well” (12.5)), and 5 pairs with the phonological distance between 13 and 16,99 (“beg” – “beckon” in “beckon the question” (13.25), “dander” – “dandruff” in “get one’s dandruff up” (13.75), “arsed” – “asked” in “can’t be asked” (14.75), “exact” – “extract” in “extract revenge/vengeance” (14.75), and “told – “total” in “all total” (16)). Finally, there were only 2 pairs with the largest phonological distance between 17 and 20,99 (“wildfire” – “wildflowers” in “spread like wildflowers” (17.25) and the target-error pair of the eggcorn “for all intensive purposes”, which was already mentioned above (20.5).

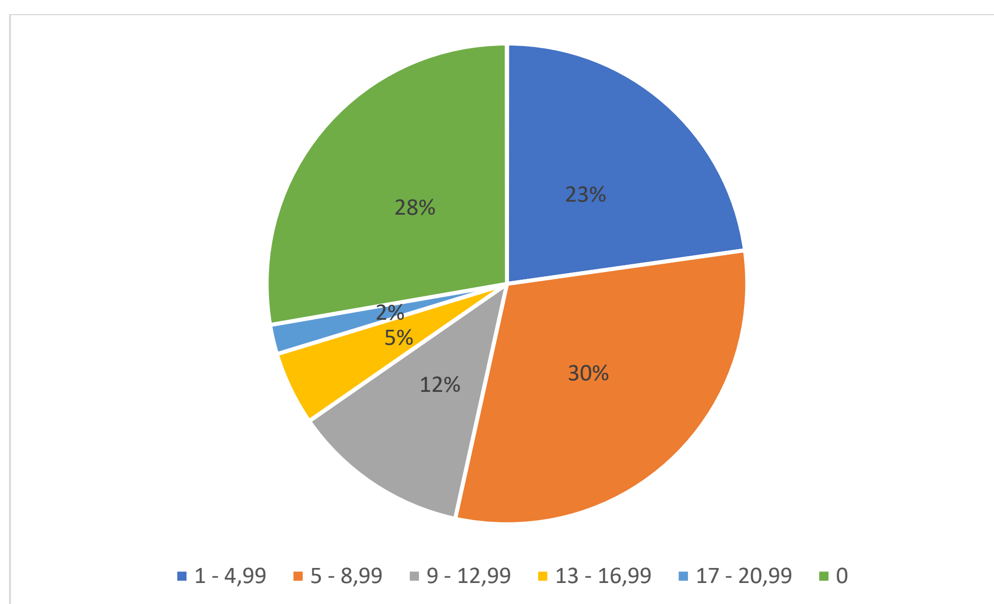

Chart 3.4. Frequency distribution of the phonological distances

Furthermore, the comparison of phonological distances in phonological neighbor pairs and pairs that are neither homophones nor phonological neighbors shows that although the average phonological distance between the non-neighbor

word pairs is larger than that of neighbor pairs (10.96 against 4.58), the difference is still relatively minor. This suggests that even the least phonologically similar target-error pairs in the sample share a considerable degree of resemblance, which could contribute to the occurrence of eggcorns.

3.2.2. Frequency effects. Correlation of frequency and neighborhood density

Since the effect of word frequency has been widely observed in both speech recognition and production errors, it is of paramount importance to also investigate its role in the occurrence of eggcorns. Word frequency estimates were based on the SUBTLEXus corpus which indicates the frequency per million words. It was determined that the average frequency of a target word is 19.59, while the average frequency of an error is 343.34. Evidently, the difference in frequency between the target and the erroneous word is quite significant, with the erroneous word being much more frequent on average. The minimum and maximum frequency of the target words in our dataset was 0.02 ("augurs" as in "augurs well") and 699.59 ("told" as in "all told") respectively. On the other hand, the minimum frequency of the error words was 0.04 ("decease" in "cease and decease" and "boggled" in "get boggled down"), which is twice the minimum frequency of the target words. The maximum frequency of the error words was 13387.8 ("and" in "and hoc/hominem/infinitum/nauseam"), which is significantly higher than the maximum frequency of the target words. Moreover, in the instances when both the frequency of the target and the corresponding error could be determined and compared (66), the target was less frequent than the corresponding error word in 38 cases (58%), while the opposite was true in 28 cases (42%) (see Chart 3.4). To exemplify the first category, the error word "dough" in the eggcorn "dough-eyed" is 3.67 times more frequent than the target word "doe" (15.88 vs. 4.33), the error word "fate" in the eggcorn "fate accompli" is 47.30 times more frequent than the target word "fait" (26.96 vs. 0.57), and the conjunction "and" erroneously used in phrases "and hoc/hominem/infinitum/nauseam" is 983.67 times more frequent than the target

preposition "ad" (13387.8 vs. 13.61). As for the second category, it can be exemplified by such instances as the target word "exact" being 7.70 times more frequent than the error word "extract" in "extract revenge/vengeance", the target word "fell" being 5.04 times more frequent than the error word "foul" and 105.80 times more frequent than the error word "fowl" in different variations on the idiom "one fell swoop", as well as the target "hair" being 40.20 times more frequent than the erroneous "hare" in the eggcorn "hare's breadth".

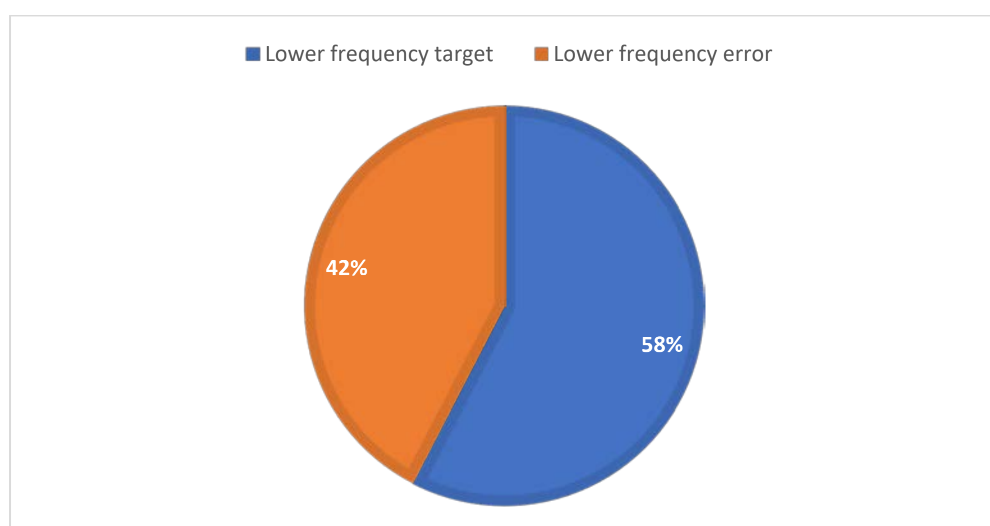

Chart 3.5. Target and error frequency comparison

Furthermore, the range of differences in the group where the target word was more frequent was determined to be 699.15, which suggests that there is a relatively small range of differences in frequency between the target and error words in this group. On the other hand, the range of differences in the group where the error word was more frequent than the target was much larger at 13373.82. This suggests that even in the instances where the target word was more frequent than the output, the difference in frequency was fairly small and, therefore, its influence was less prominent than in the more represented group where targets were less frequent. Overall, these results corroborate the strong presence of the frequency effect, according to which low-frequency words are retrieved with greater difficulty from the mental lexicon and have a tendency to be substituted by more frequent words due to their greater accessibility and familiarity to speakers.

Furthermore, we have tested out Vitevitch's findings regarding malapropisms, according to which low-frequency words with small neighborhoods and high-frequency words with large neighborhoods tend to result in errors more often, to determine if they would apply to eggcorns as well. To separate the words into dense and sparse neighborhood categories, as well as low and high-frequency categories, the median value was used (7 and 1.71 respectively). The words the data on which was absent from the SUBTLEXus corpus were assessed as low-frequency. Multi-word targets (7) were excluded from our analysis due to the impossibility of adequately measuring their neighborhood density since this concept is typically applied to single words, whereas in language processing multi-word phrases are treated as holistic units rather than individual words. In addition, repeated occurrences of targets (9) were disregarded in order to eliminate potential bias and ensure that each target was given equal weight in our analysis. Having completed these manipulations, we conducted the comparison of neighborhood densities and frequencies of 84 target words. The results were consistent with Vitevitch's findings: 75% (63) of targets displayed a positive correlation between neighborhood density and word frequency. Among them, the majority (40 out of 63) were of low frequency and low density, while a smaller proportion (22 out of 63) were of high frequency and high density. This suggests that the combination of low word frequency and low neighborhood density may be more strongly associated with the occurrence of eggcorns than that of high word frequency and high neighborhood density, although both factors are evidently influential. The correlation of low frequency and density measures was displayed by such target words as "cahoots" (density rate – 1, frequency rate – 0.63), "expatriate" (density – 2, frequency– 0.18), "foment" (density – 1, frequency–0.04), "dudgeon" (density – 3, frequency – 0.12), and "gobbledygook" (density – 0, frequency – 0.37). On the other hand, some examples of the target words displaying high measures in both frequency and density include "dawn" (density – 33, frequency – 25.51), "wing" (density – 42, frequency – 20.24), "fell" (density – 40, frequency – 73), "hair" (density – 33, frequency – 153.55), and "manner" (density – 17, frequency – 11.53). Among the target words that resulted in

errors, considerably fewer items had dense phonological neighborhoods and low word frequency (23% or 19), and examples include “pique” (density – 34, frequency – 0.25), “fait” (density – 35, frequency – 0.57), and “rote” (density – 43, frequency – 0.12). Finally, only a small minority of target words had sparse density and high frequency (2%), namely “exact” (density – 1, frequency – 22.63) and “foolproof” (density – 0, frequency – 1.84). These results also corroborate Vitevitch’s findings, which posit that low-frequency words with many phonological neighbors may gain additional activation from them, while high-frequency words with few neighbors are not at risk of being overpowered by them, which ultimately leads to a less pronounced tendency to result in errors.

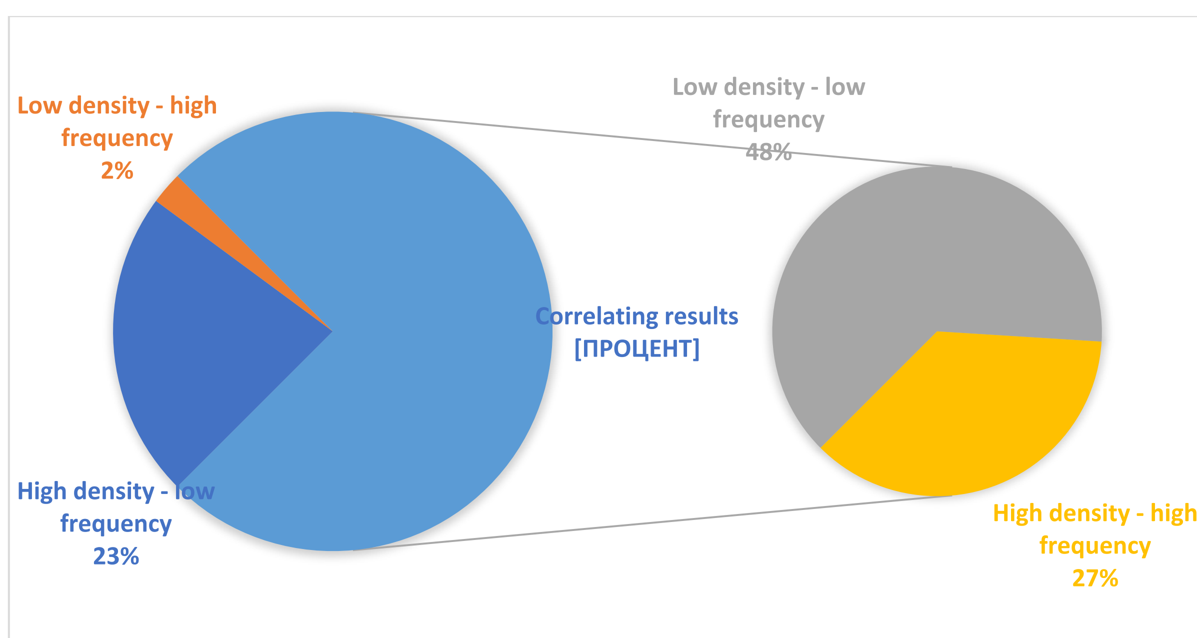

Chart 3.6. Correlation of neighborhood densities and frequencies of target words

3.2.3. Phonotactic probability effect

As has been established in the first chapter of this paper, conformity of the word to the phonological rules of the language plays an integral role in the processes of word recognition and word production. Both misperceptions of speech and word production errors tend to follow the phonological rules of the language, while the processing of real words is usually more efficient than that of non-words (the lexical status effect). Therefore, it was necessary to analyze eggcorns according to their

phonotactic probability, i.e., the likelihood of a word occurring in the English language.

In a framework proposed by Vitevitch and Luce, the word “cat” was chosen to represent high phonotactic probability words [108, pp. 377-378]. Since the phonotactic probability measure of the word “cat” is 0.007, it was chosen as a threshold to divide words into low and high phonotactic probability subgroups for the purposes of our research. The results depicted in Chart 3.6 showed that most of the target words (67 out of 93 or 71%) had low phonotactic probability, 26 (27%) target words were of high phonotactic probability, and 2 (2%) target words were phonotactically improbable, i.e., they contain sound sequences that were unlikely to appear according to the phonotactic rules of the English language. Examples of low probability targets include “foment” (0.006), “godsend” (0.005), “leper” (0.004), “squib” (0.003), “dawn” (0.002), and “foolproof” (0.001), while high phonotactic probability was displayed by such targets as “defuse” (0.007), “cart” (0.012), “pedestal (0.017), and “dander” (0.01). Finally, target words “jaw” and “augurs” were determined to be phonotactically improbable. Overall, the data suggest that eggcorns tend to occur when the target word has a low phonotactic probability, as well as that words with low phonotactic probability are more susceptible to being inaccurately processed and consequently incorrectly stored in the mental lexicon. Moreover, the target word was substituted by a more phonotactically probable word in 29 cases, compared to 21 cases where the opposite was true, which indicates that phonotactic probability also might play a role in the choice of the substitution, with more probable words likely to be chosen over the less probable ones.

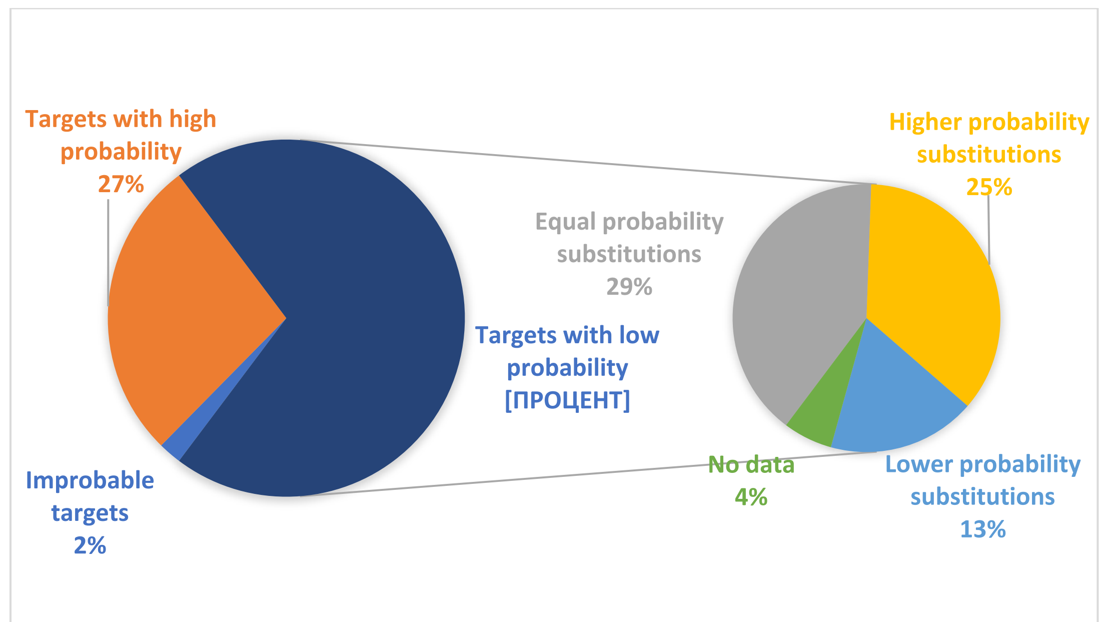

Chart 3.7. Phonotactic probability of target words (with data on low probability targets)

Furthermore, each of the two types of targets was analyzed according to the characteristics of the erroneous words they were substituted by. As can be seen in Chart 3.7, low probability targets were substituted by words of the same probability in 27 cases (e.g., "crux" – "crutch" (0.006), "dawn" – "dong" (0.002)), words of higher probability in 24 cases (e.g, "slings" (0.003) – "strings" (0.014), "seize" (0.004) – "cease" (0.005)), and words of lower probability in 12 cases (e.g., "rote" (0.003) – "route" (0.001), "wing" (0.006) – "whim" (0.004)). Therefore, speakers tend to substitute words that consist of unusual sound sequences either with equally unusual or more probable segments. Interestingly, even when a higher probability substitution was made for low probability target words, a significant portion of those substitutions were still of low probability (14 out of 24 cases), which indicates an absence of the propensity for the drastic change in phonotactic probability from the target to the output.

On the other hand, the results were slightly different for high-probability target words, which were substituted by words of the same probability in 10 cases (e.g., "defuse" – "diffuse" (0.007), "pray" – "prey" (0.014)), words of lower probability in 9 cases (e.g., "ruckus" (0.008) – "raucous" (0.007), "cart" (0.012) – "cat" (0.007)),

and words of higher probability only in 4 cases, such as "brim" (0.007) – "rim" (0.009) (see Chart 3.8). This data suggests that the role of the phonotactic probability of the substitution is less significant for target words that largely conform to the phonotactic rules of the language. Moreover, it is interesting to note that, similarly to the tendency observed in low probability targets, 8 out of the 9 substitutions of high probability targets with words of lower probability were still of relatively high probability (the only exception being "arsed" (0.011) – "asked" (0.03)). This suggests that even when high-probability words are substituted by less probable words, speakers still tend to select words with relatively common sound patterns. Overall, these findings point to the fact that phonotactic probability does play a significant role in the occurrence of eggcorns, with words of low phonotactic probability more likely to result in errors and be substituted by more probable words. As for the less common instances of high-probability words resulting in errors, they are highly unlikely to be substituted by a low-probability word (only 1 detected example).

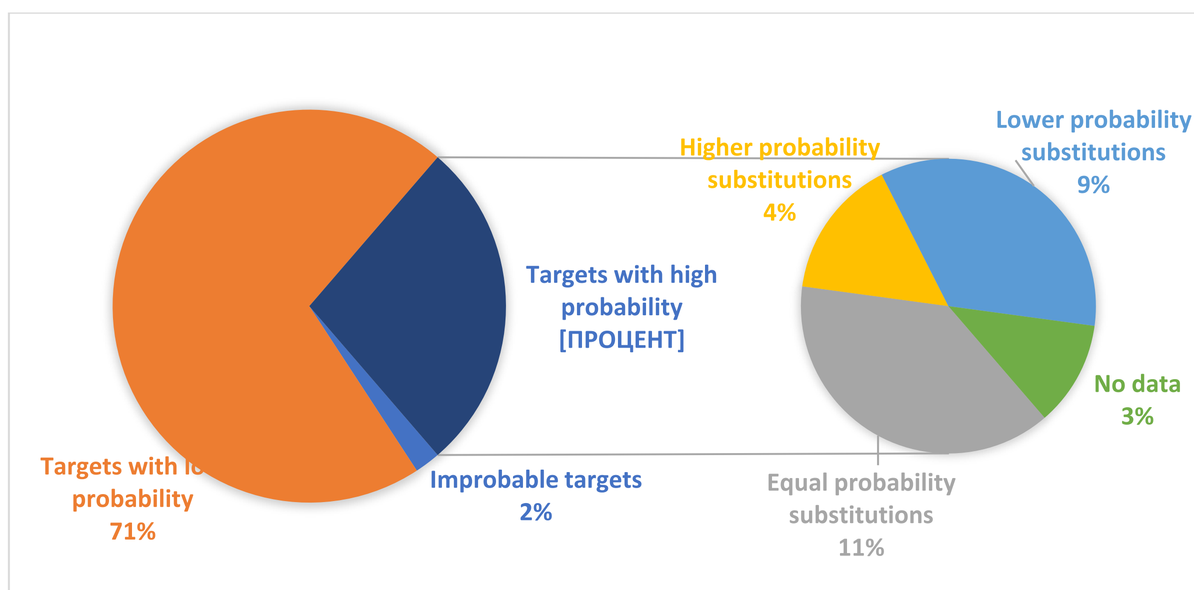

Chart 3.8. Phonotactic probability of target words (with data on high-probability targets)

When researching eggcorns, it is also important to consider the non-word legality effect, which refers to the better processing of non-words conforming to the legal language patterns, considering that a considerable portion of our sample of eggcorns is represented by non-words (17 out of the 100 units analyzed). As can be

seen in the chart below, which depicts the phonotactic probability of these non-words, most of them had low phonotactic probability (8), 6 had high phonotactic probability, and 2 were phonotactically improbable. These results point to the fact that, although the resulting non-words are slightly more likely to have low phonotactic probability, they still maintain a certain degree of conformity with the phonotactic rules of English, only rarely resulting in improbable sound sequences.

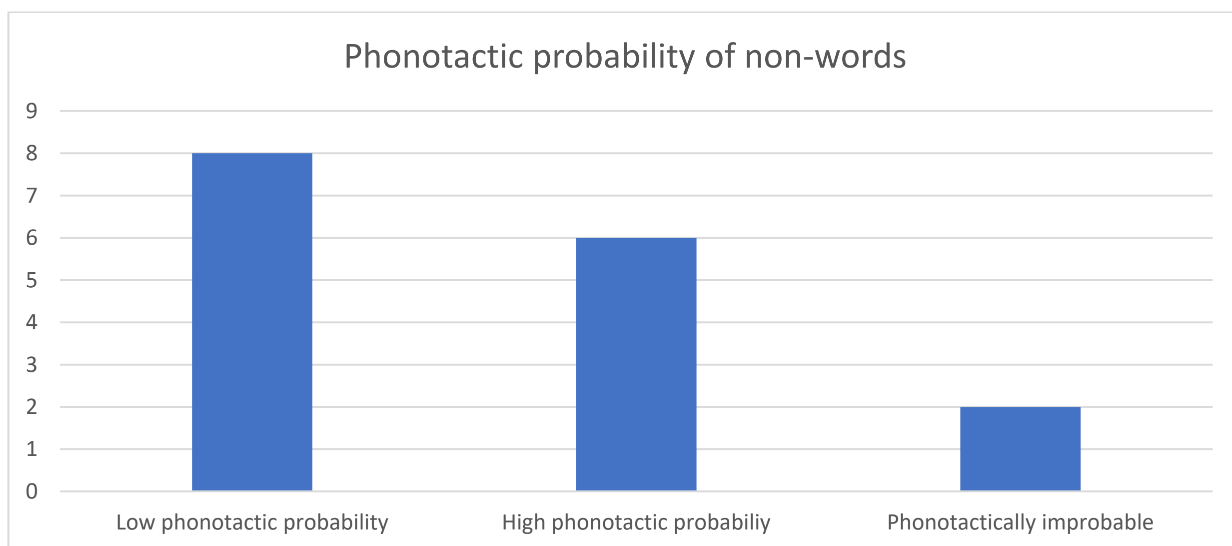

Chart 3.9. Phonotactic probability of non-words

Table 3.1

Non-word	Phonotactic probability of the non-word	Phonotactic probability of the target
1. alterior	0.004 (low)	0.003 (low)
2. bellweather	0.003 (low)	0.003 (low) (homophones)
3. bloodgeon	0.013 (high)	0.013 (high) (homophones)
4. buttkiss	0.004 (low)	0.004 (low)
5. cacoughany	0.011 (high)	0.011 (high) (homophones)
6. eggcorn	0.003 (low)	0.002 (low)

Table 3.1 continued

Non-word	Phonotactic probability of the non-word	Phonotactic probability of the target
7. eggtopic	Improbable	0.003 (low)
8. expatriot	0.008 (high)	0.008 (high)
9. fullproof	0.001 (low)	0.001(low)
10.garbledygook	Improbable	0.005 (low)
11.higherarchy	0.002 (low)	0.002 (low) (homophones)
12.in-term	0.011 (high)	0.008 (high)
13.ontray	0.004 (low)	0.004 (low) (homophones)
14.pedastool	0.007 (high)	0.017 (high)
15.pre-madonna	0.009 (high)	0.009 (high) (homophones)
16.segway	0.003 (low)	0.005 (low)
17.winfall	0.004 (low)	0.006 (high)

Correlation of phonotactic probability in targets and non-words

As for the correlation between the target and the non-word, it was observed in all but one instance (where the phonotactic probability of the target was high and that of the error was low: "windfall" (0.006) and "winfall" (0.004)), meaning that in most target-error pairs both members shared either low or high phonotactic probability. As can be observed in the table above, five of the low probability pairs had an equal count of probability (3 pairs being homophones), and in 3 cases the resulting eggcorn had lower phonotactic probability than the target. Interestingly, there were 2 cases where the resulting eggcorn had a higher phonotactic probability than the target, meaning that the non-word was actually more likely to occur in language than the correct word. Specifically, "eggcorn" (0.003) was more phonotactically probable than "acorn" (0.002), as well as "alterior" (0.004) more probable than "ulterior" (0.003).

Among the high probability pairs, six had an equal count of probability (3 pairs being homophones), one non-word had a lower probability than the target, and, conversely, one had a higher probability (namely, "in-term" (0.011) was more probable than "interim" (0.008), possibly due to the target being borrowed from the Latin language). Overall, this data suggests that eggcorns tend to be influenced by the non-word legality effect due to phonetic similarity with the real-word target, with several instances of target and error being homonyms. Moreover, although not very frequent, there exists a possibility of non-words emerging as more phonologically probable than the existing target word.

3.2.3. Morphosyntactic characteristics of target words and phrases

To understand what processes and phenomena give rise to the occurrence of eggcorns, it is necessary to investigate correct words and phrases that get substituted by them and attempt to find common characteristics and tendencies. Firstly, we conducted the analysis of the target utterances according to their part of speech.

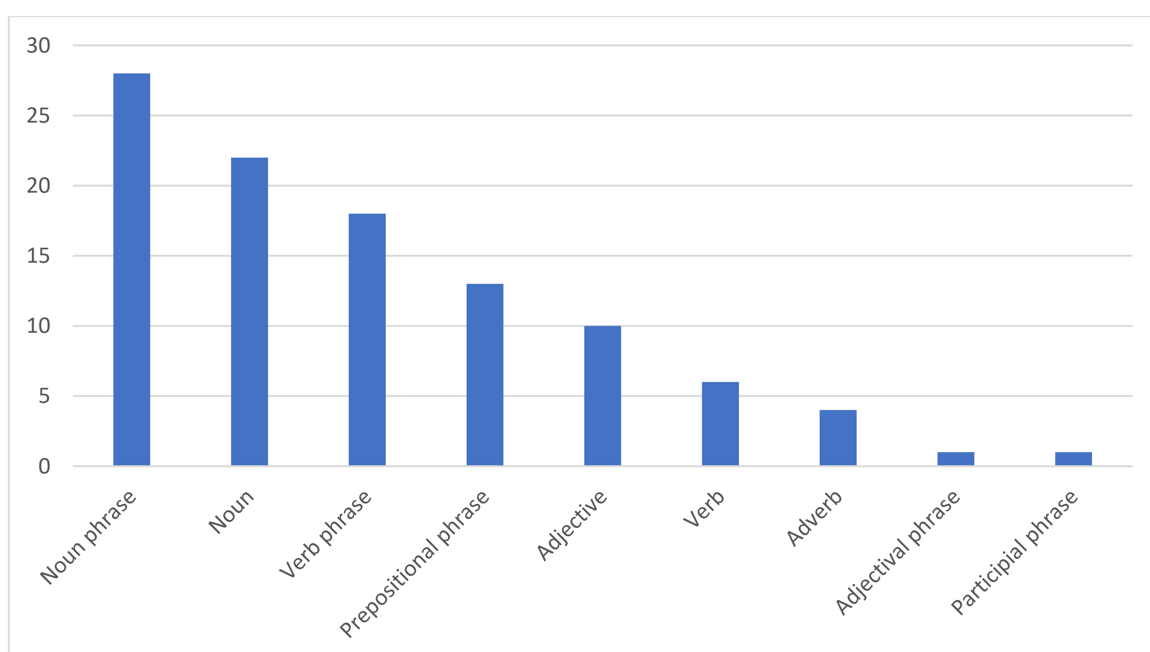

Chart 3.10. Targets according to their part of speech

As can be seen from the chart above, noun phrases comprised the most represented group (28), followed by nouns (22), verb phrases (18), and prepositional phrases (13). On the other hand, adjectives (10), verbs (6), and adverbs (4) were less frequently in involved in eggcorns, while adjectival phrases and participial phrases

were especially rare with only one representative per each group. It was shown that eggcorns conform to the categorical constraint evident in speech errors since they consistently adhere to the syntactic category of the target. Interestingly, when several groups of lexical units were looked into more closely, it was discovered that a substantial number of all nouns, adjectives, and adverbs were compound, while 3 out of 5 verbs were phrasal verbs (see Table 3.2).

Table 3.2

Part of speech	**Number of targets**
Noun phrase	28
Verb phrase	18
Prepositional phrase	13
Adjectival phrase	1
Noun	21 (14 compound nouns)
Verbs	5 (3 phrasal verbs)
Noun/verb	1
Adjectives	8 (5 compound adjectives)
Adjective/adverb	2
Compound adverb	2
Participial phrase	1

Targets according to their part of speech

Overall, the majority of target lexical units (85 out of 100) consisted of multiple words or free morphemes, which suggests that eggcorns are more likely to occur in more complex words and phrases. It is worth noting that compound words and phrasal verbs in particular lend themselves to be misinterpreted due to multiple possible interpretations of their constituent parts. The meaning of phrasal verbs is especially unpredictable due to being non-compositional, while 12 out of 21 compounds in our sample also had non-compositional meanings (e.g, the meanings of “doe-eyed” (innocent-looking), “jaw-dropping” (surprising), “bedside manner” (a doctor’s attitude to a patient), and “bellwether” (an indicator of a trend) cannot be

derived from the meanings of their constituent parts). Moreover, since eggcorns can be semantically justified, speakers might gain more context from multi-word lexical units to fill in the gap in their mental lexicon and select an erroneous but semantically fitting word. Additionally, the tendency of eggcorns to manifest in nouns and verbs in particular points to the fact that lexical units which carry more meaning and play a more central role in the message of the utterance are more susceptible to eggcornization.

Given such a substantial amount of multi-word lexical units in our sample, it was of paramount importance to also consider target phrases from the point of view of phraseology. As can be seen in the Chart 3.11 presented below, the majority of target utterances analyzed by us could be characterized as phrasemes (76 out of 100), i.e., multi-word or multi-morpheme lexical units, the meaning of which is more specific than the sum of meanings of their components or not predictable from them. Among them, there were 49 idioms, 14 fixed collocations, 9 non-compositional compounds with a figurative meaning, 1 simile, and 3 phrasal verbs.

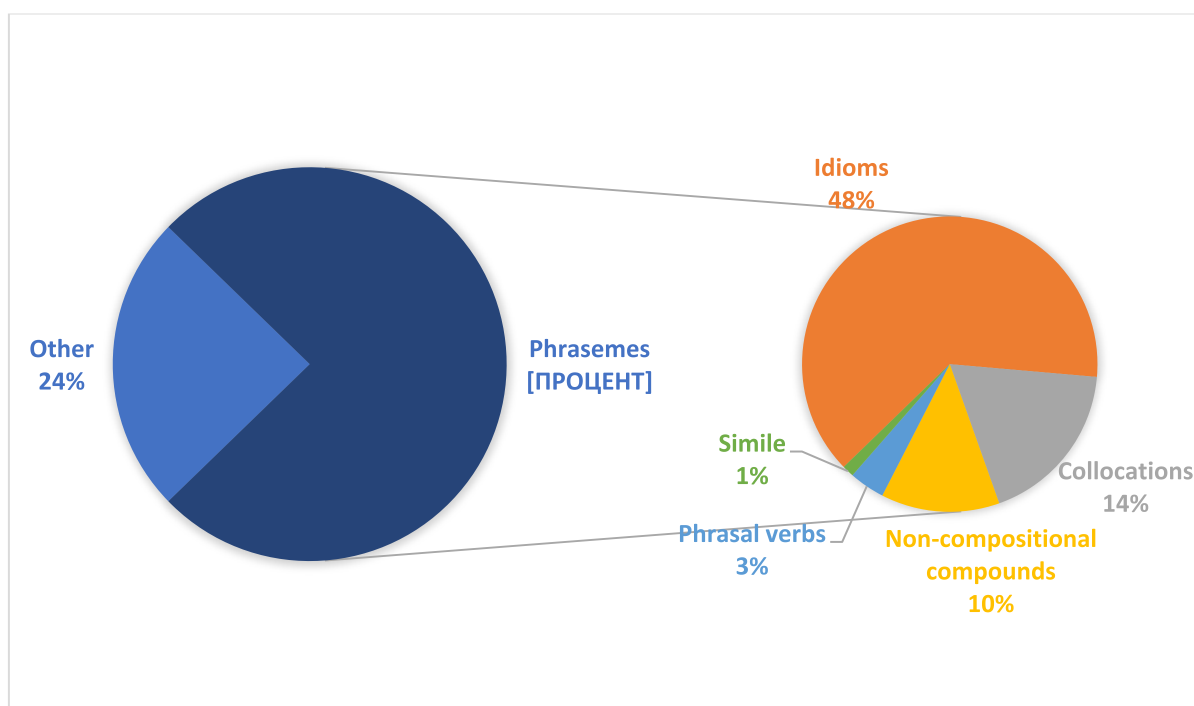

Chart 3.11. The phraseology of eggcorn targets

This data points to the fact that phrasemes are significantly more prone to being transformed into eggcorns. This might be explained by the fact that phrasemes

are acquired and retrieved as a single unit rather than through compositional analysis. Therefore, the speaker might try to match an unfamiliar phraseme to a similar-sounding and semantically justified word or phraseme that is already stored in their mental lexicon. Moreover, the susceptibility of phrasemes to eggcornization can be often explained by their opaque semantics. For instance, since idioms in particular were most likely to form the basis of an eggcorn, we can assume that their figurative and non-compositional meaning makes it easier to misinterpret them as a similar-sounding word or phrase. This also explains the presence of non-compositional compounds, or so-called morphological idioms, the meaning of which is likewise not readily predictable from the meanings of their constituent parts.

For example, in the case of the idiom "go at it hammer and tongs", which denotes energetic, forceful activity (especially arguing), the meaning of the phrase as a whole is non-compositional, and the constituent "tongs" may seem especially opaque to speakers, since blacksmithing is no longer a common practice (which is supported by its relative low-frequency (0.78). On the other hand, in the resulting eggcorn "hammer and thongs", the speaker instead selects the word "thongs", which is more frequent (2.1), phonetically similar to the target (both are phonological neighbors: /T.AO.NG.Z/ - /TH.AO.NG.Z/), and fits semantically: by using "thongs", referring to whips, the speaker mentions two instruments of punishment that can be used in aggressive confrontations. Conversely, "free phrases" are less prone to be the subject of an eggcorn since their constituent parts are more likely to be analyzed in isolation. For example, the only compositional compounds of English origin that were involved in an eggcorn are "martial law" ("marshal(l) law"), "cole slaw" ("cold slaw"), and "praying mantis" ("preying mantis"). Otherwise, while phrases like "clear blue sky" can be involved in speech production errors such as sound substitution (e.g., "glear plue sky"), they are less likely to result in an eggcorn, since their constituents are commonly used on their own, can be analyzed separately and, therefore, are more transparent.

It is interesting to note that despite the prevalence of multi-word and multi-morpheme lexical units in our sample, the number of elements constituting target words and phrases tended to be relatively small. As can be seen from the chart depicted below, multi-word targets consisted of 2 words in most instances (39), followed by 3-word and 4-word targets (in 25 and 10 instances respectively), while targets consisting of 5 and 6 words were considerably less frequent (5 and 6 occurrences respectively). Interestingly, the number of targets decreased in inverse proportion to the number of words they consisted of. The reason for this could be that shorter phrases are more easily and efficiently memorized and repeated and, therefore, may be more likely to be passed on verbally. Moreover, as has been noted by R.A. Rubin, eggcorns rarely involve heightened, literary language, and often stem from popular, cliched phrases, which would also explain the infrequency of overly complex phrases [90, p. 10].

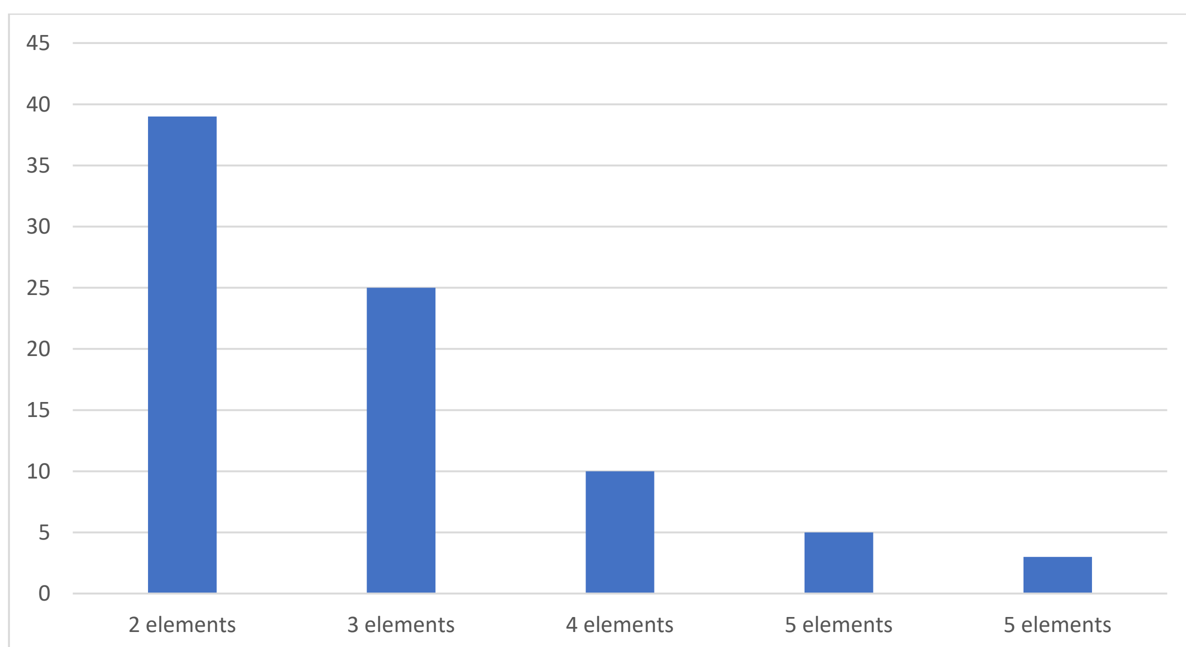

Chart 3.12. Number of elements in multi-word targets

As for single-word targets, most of them had 2 morphemes (11), words consisting of 1 morpheme were less frequent (5), and 3-morpheme words formed the least represented category with 2 cases of occurrence (see Chart 3.13). As we can see, although 3-word morphemes were fairly infrequent and more complex words were absent altogether, there was still a noticeable prevalence of 2-morpheme words over

single-morpheme ones, confirming the overall tendency of eggcorns to occur in multi-element lexical units.

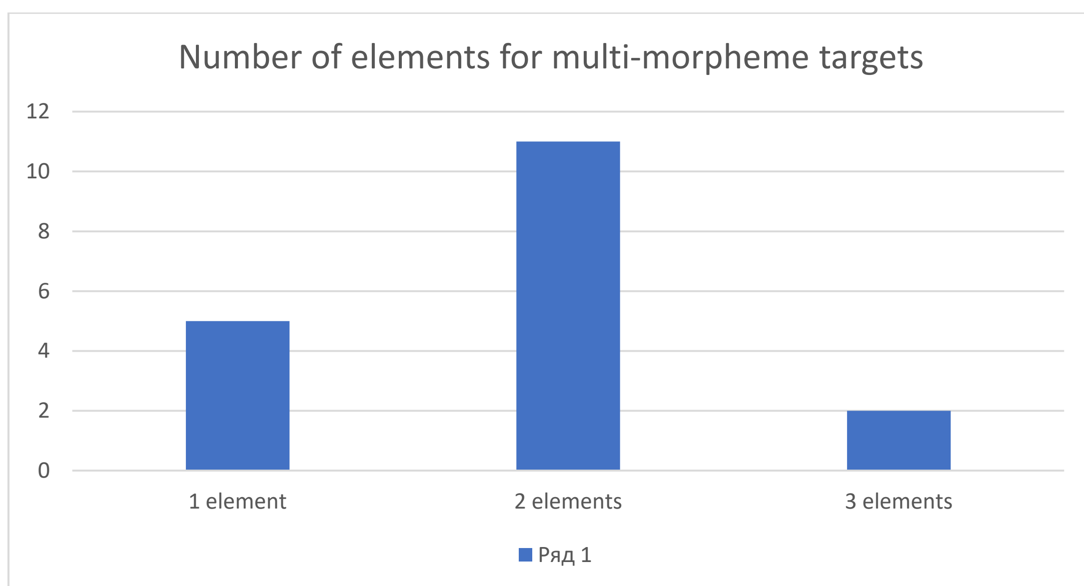

Chart 3.13. Number of elements for multi-morpheme targets

As for the correlation between the number of elements in the target utterance and eggcorn, the majority of the analyzed eggcorns had the same number of elements as their target expressions (83) (for example, both "one fell swoop" and "and one fowl swoop" consist of 3 elements). This suggests that speakers tend to preserve the overall structure of the original phrase or word, only modifying a certain part that they find challenging or unclear. However, it is interesting to note that there were still a few cases where the eggcorn had a different number of elements than the target phrase or word. Among them, there is a higher tendency to add elements (words or morphemes) to the original expression than to subtract them. Namely, there were 6 instances of a redundant morpheme (e.g., "acorn" – "eggcorn", "bludgeon" – "bloodgeon", "segue – "segway") and 6 instances of a redundant word appearing (e.g., "in layman's terms" – "in lame man's term", "augurs well" – "all goes well", "much ado about nothing" – "much to do about nothing"). On the other hand, the resulting eggcorn possessed a smaller number of words in 5 cases (e.g., "for all intents and purposes" – "for all intensive purposes", "bona fide – bonified"), and there were no instances of morpheme deletion. The means by which these variations

in the number of elements occur, such as morpheme addition and metanalysis, will be further explored in the following section of the chapter.

3.2.4. Formal classification of eggcorns

Having determined a noticeable influence of such factors as phonology and morphology on the creation of eggcorns, we endeavored to classify the eggcorns in our sample according to the size of the unit involved and the type of disruption by drawing on Dell's classification of production errors.

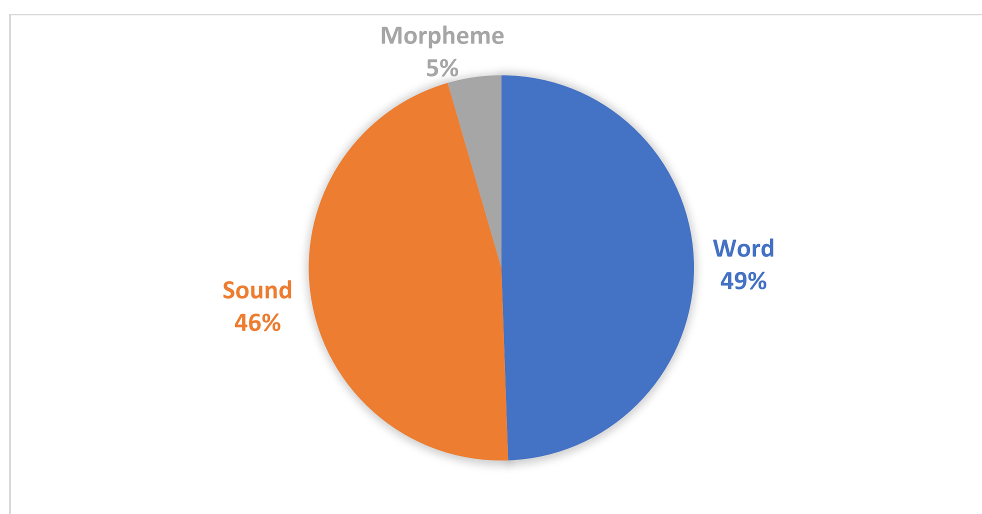

Chart 3.14. Type of the unit involved in eggcornization

The distribution of eggcorns according to the unit involved, which can be observed in the chart above, suggests that eggcorns most often involve erroneous word and sound selection, which occurred almost in equal measure (in 44 and 41 cases), rather than by morpheme selection (16 instances).

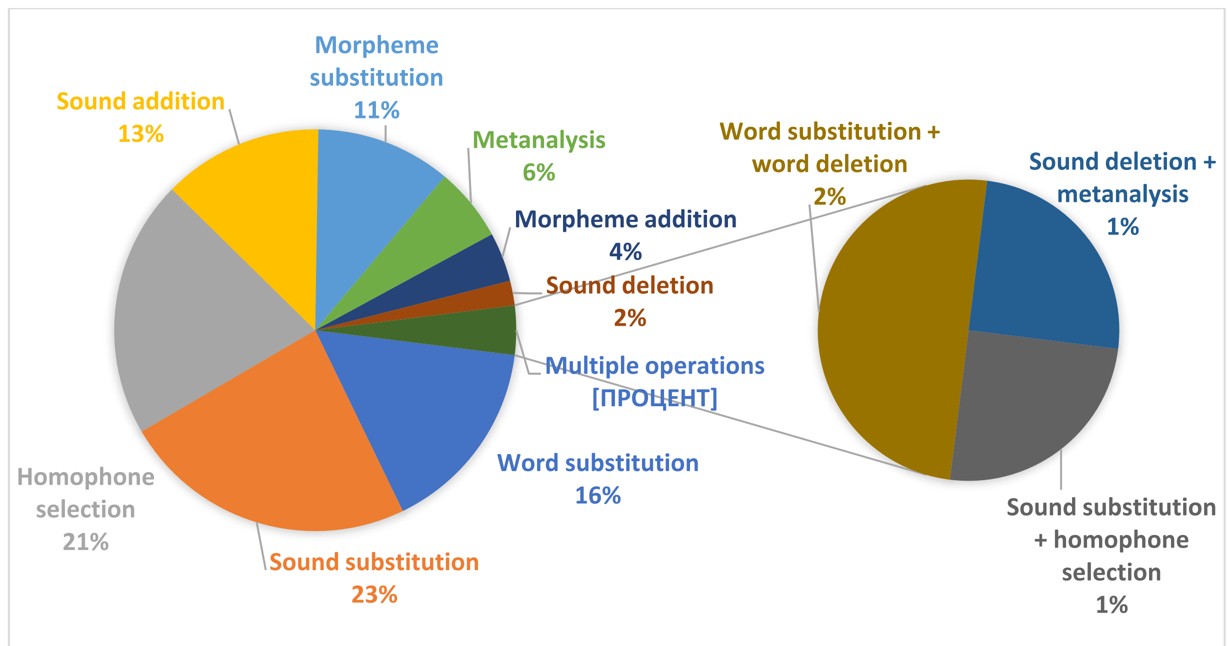

Chart 3.15. Formal classification of eggcorns

It is important to note that the overwhelming majority of eggcorn were non-contextual, i.e., they did not have an obvious source within the whole utterance. In fact, only 3 eggcorns could be defined as contextual (or syntagmatic), i.e., involving a mistake in the ordering. Namely, there were two anticipations ("pipe piper" (from "pied piper") and "bread and breakfast" (from "bed and breakfast"), and one perseveration ("cease and decease" (from "cease and desist")). The rarity of misordering errors in eggcorns may be explained by the fact that they are also motivated by semantic context aside from the syntagmatic one. As can be seen from the chart above, sound substitution was the most common mechanism of creating eggcorns, occurring 24 times and followed closely by homophone selection, which was not mentioned in Dell's original classification but was present in the classification of slips of the ear proposed by Pardo et al. and occurred in our sample 21 times. Word substitution and sound addition were also fairly frequent, with 16 and 13 occurrences respectively. On the other hand, morpheme substitution was less common, with 11 instances. It is important to note that the mechanism of metanalysis (i.e., the change in the division between words or morphemes resulting in new lexical units), which also was not mentioned in Dell's original classification, was recorded in

6 instances. Morpheme addition and sound deletion were most infrequent, having been recorded only 4 and 2 times respectively.

Interestingly, the nature of word substitution differed from what was observed in examples provided in Dell's classification for production errors. For instance, he exemplifies this type of error with the substitution of "pepper" with "salt" and "rhapsody" with "restaurant", with the phonological distance between the words being fairly large (24.5 and 30.25 respectively). On the other hand, target-error pairs that were involved in word substitution in our eggcorn sample displayed phonological distance that ranged from 6.25 to 16 and was significantly smaller on average (10.89). This suggests that, unlike in ordinary production errors, where word substitutions tend to be selected either on a solely semantic basis or due to slight phonological similarity, the similarity in eggcorn word substitution is considerably more pronounced.

Moreover, in most instances, morpheme substitution was achieved either through a change involving a single phoneme or selection of a homophonous morpheme. For instance, in three instances morpheme substitution resulted from a sound substitution (e.g., "foolproof" /ˈfuːl.pruːf/ » "fullproof" /fʊl.pruːf/), in one – from sound addition ("gobbledygook" /ˈɡɑbəldiˌɡʊk /» "garbledygook" /ˈɡɑːr.bəldiˌɡʊk/), and in one – from sound deletion ("windfall" /ˈwɪnd.fɑːl/» "winfall" /ˈwɪn.fɑːl/). In three instances, morpheme was substituted with an identically sounding morpheme (e.g., "bellwether" » "bellweather" /ˈbelˌweð.ɚ/). Moreover, in 3 out of 4 occurrences, morpheme addition resulted in a word that was still homophonous with the target. This occurred due to part of a morpheme being reinterpreted as a separate morpheme (e.g., the initial vowel in "acorn" was reanalyzed as a root morpheme "egg"). There were also several instances where a change in one sound led to a change in the number of elements contained in an eggcorn, as well as in the part of speech of the word. For example, the noun "ruckus" consisting of one morpheme was transformed into the two-morpheme adjective "raucous" through the substitution of the sound / ɑ/ with / ʌ/, while in the phrase

“much ado about nothing”, the noun “ado” was transformed into the infinitive verb “to do” through sound addition.

Considering the fact that the comparison between eggcorns and folk etymology is often drawn by linguists, with Reddy calling the latter “the closest cousin to eggcorns”, it is essential to note some similarities in the processes behind these two phenomena. Folk etymologies result from the reinterpretation of unfamiliar words, especially foreign borrowings and archaisms, through the substitution with more familiar words and morphemes. A typical example of folk etymology in the English language is the creation of the word “sparrowgrass” from the Greek “asparagus” by analogy. Similar logic seems to be at work in such eggcorn transformations as “entrée” – “ontray”, “bona fide” – “bonified”, and “augurs well” – “all goes well”, where borrowings from French and Latin are substituted with homophonous English morphemes and words. Additionally, in the eggcorn “old-timer’s disease” (from “Alzheimer's disease”) the same process happens with an eponymous medical term that might be relatively unfamiliar to an average speaker.

It is also important to discuss the phenomenon of metanalysis, which previously has been largely overlooked by various classifications of speech perception and production errors. Also known as rebracketing, it is a process of breaking down the composition of a word into a different set, often leading to a change of semantics. Metanalysis plays a prominent role in historical linguistics since it is closely connected with folk etymology and leads to the formation of new words. Among the eggcorns analyzed by us, there were several instances of false splitting between morphemes (e.g., “stark-raving mad” – “star-craving mad”), which sometimes led to the creation of two words from one (“godsend” – “god’s end”, “layman – lame man”) or vice versa (“prima donna” – “pre-madonna”). Metanalysis often involves the confusion of a boundary between an article and a noun (for example, the word “newt” was actually resegmented from “an ewt”), and examples of this process were also present in the analyzed list of eggcorns (“straight as an arrow” – “straight as a narrow”, “opposable thumb” – “a posable thumb”). Such rebracketing

may stem from slips of the ear provoked by oronyms, which were already discussed in the first chapter of this study and refer to word strings that can be logically divided in multiple ways.

It is worth mentioning that there were also instances where several mechanisms were involved. For instance, there were two instances where one component of a phrase was affected by word substitution and another by word deletion ("for all intents and purposes" » "for all intensive purposes"; "at someone's beck and call" » "at someone's beacon call"), one instance where an eggcorn was formed through sound substitution in one constituent of a compound noun and homophone selection in another ("laissez-faire" » lazy-fare"), as well as one instance where the original word underwent a combination of sound deletion and metanalysis ("interim" » "in-term").

In view of the results presented above, namely tendencies for short phonological distances in word substitutions, selection of homophones, resegmentation of morphemes, as well as morpheme additions and substitutions involving single sound change, it can be concluded that eggcorns exemplify speakers' ability to make sense of unfamiliar linguistic information and take part in the logically justifiable linguistic innovation with minimal interference with the sound structures of the original lexical items.

3.3. Etymology

In section 3.2.4. of this chapter, a tendency to substitute unfamiliar forms, especially of foreign origin, was observed, resembling the process of folk etymology. Therefore, it was important to further consider the etymology of lexical units involved in the creation of eggcorns to determine whether it plays a significant role in the process.

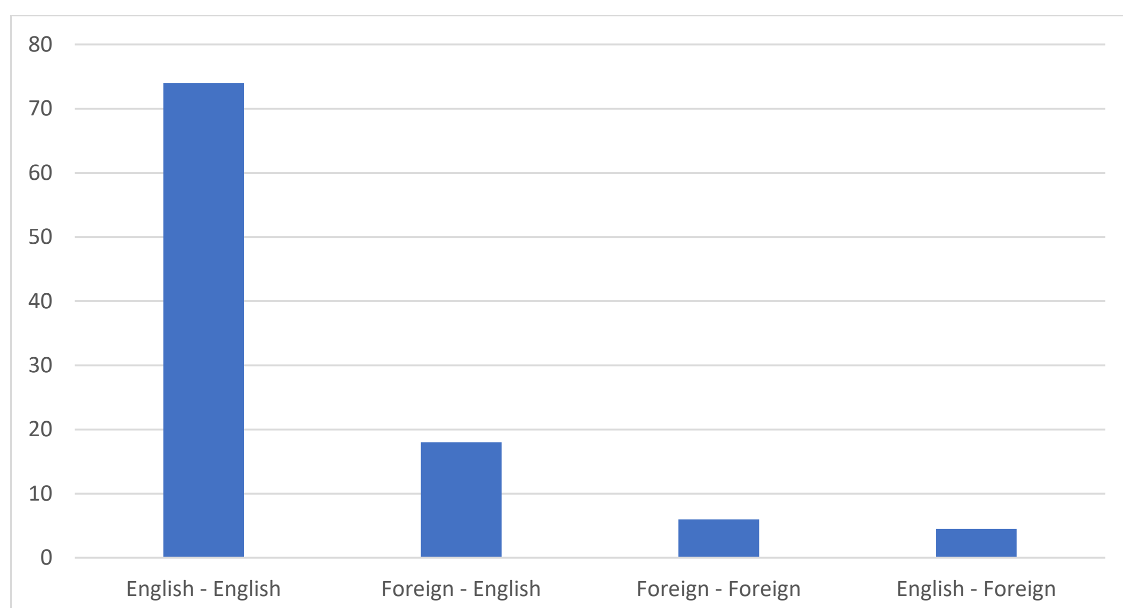

Chart 3.16. Correlation between the etymologies of targets and errors

As can be seen in the chart presented above, in 74 out of 100 eggcorns, both target and error were of English origin, the target was of foreign origin and the error word was of English origin in 18 cases, both target and error were of foreign origin in 6 cases, and, finally, only in 2 cases, the target was of English origin, while the resulting error was of foreign origin. Overall, the prevalence of English-origin words supports prior observations made by linguistics, namely that eggcorns do not tend to involve complicated vocabulary and are not made by speakers in an attempt to transcend their linguistic abilities. Nevertheless, there was also a sufficient number of instances where a word of foreign origin was substituted by a word of English origin to suggest that speakers tent to substitute unfamiliar words with similar but more English-sounding lexemes. Moreover, among targets were 14 non-integrated borrowings from Latin, French, Italian, and Yiddish languages, 12 out of which were substituted by English linguistic material, as can be seen in the table depicted below.

Table 3.3

Eggcorn	**Items involved**	**Etymology**
all goes well (for)	augurs » all goes	Latin – Middle English
alterior motive	ulterior » alterior	Latin –Middle English
and	ad » and (at)	Latin – Middle English

hoc/hominem/infinitum/nauseam		
bonified	bona fide » bonified	Latin – From Latin bonus + -fy
crutch of the matter	crux » crutch	Latin – Middle English
don't know buttkiss	bupkis » buttkiss	Yiddish – Middle English
eggtopic pregnancy	ectopic » eggtopic	Modern Latin – Middle English + From Latin *topica*
fate accompli	fait » fate	French – Middle English
in-term	interim » in-term	Latin – Middle English
lazy-fare	laissez-faire » lazy-fare	French – English
locust of control	locus » locust	Latin – Middle English
ontray	entree » ontray	French – Middle English
pre-madonna	prima donna » pre-madonna	Italian

Table 3.3 continued

Eggcorn	Items involved	Etymology
segway	segue » segway	Italian – Middle English

Eggcorns derived from borrowed words

Since non-integrated loanwords preserve their original form, they may appear less recognizable and understandable to speakers, who attempt to substitute them with more familiar lexical units to make sense of the utterance.

It is important to mention that 11 target-error pairs exhibited shared etymological roots. For instance, the word "expatriate" is transformed by analogy by adding the prefix "ex" to the word "patriot", with the etymology of both words tracing back to Latin *patris* ("fatherland"). Similarly, in an eggcorn "for all intensive purposes", where "intents" is substituted by "intensive", both words stem from Latin *intendere* ("to stretch out"), while the substitution of "naught" with "not" in "all for not" is also characterized by the shared origin (from the Middle English *naht* ("nothing")). Notably, in an eggcorn "above/beyond reproach", both the original word 'reproach" and the error word can be traced to the Latin "prope" (near"). However, both words acquire almost antonymic meanings due to different prefixes, "ad-" ("to") and "re- ("opposite of").

Interestingly, although in an eggcorn "get one's dandruff up" the original word "dander" is transformed into "dandruff" through morpheme addition, historically, the reverse process led to the formation of "dander", since it originally appeared as a result of backformation from "dandruff", both words referring to dead skin flakes (of human origin in the case of "dandruff" and of mammal in the case of "dander"). Nevertheless, the semantic implications of the interplay between both words in the eggcorn are fairly complicated and shall be discussed in greater detail further in this chapter. Likewise, the eggcorn "much to do about nothing" mirrors the historical language change that led to the creation of the target word "ado" from a contraction of Middle English "at do" (meaning "to do").

Moreover, two of the eggcorns derived from borrowed words also shared etymological roots with the target. Namely, both constituents of the "ad"-"at" pair in the eggcorn "at hominem/infinitum/nauseam" stem from the Proto-Indo-European root *ad- ("to, near, at"). Additionally, both "bona fide" and its eggcorn version "bonified" stem from the Latin word "bonus ("good").

Another interesting example of shared etymology is the eponymous eggcorn, which gave the term its name: the confusion that led to the creation of "eggcorn" from "acorn" is evident in the etymology of the target itself, which stems from Old English *æcern* ("nut") since its modern spelling itself appeared by erroneous, folk etymological association with the word "corn". Finally, the idiom "at someone's beck and call" is especially prone to eggcornization, with several variations attested. The case of "beckon call" is similar to the eggcorn "get one's dandruff up" in that, historically, the target actually stems from the substitution, i.e., "beck" emerged as a shortening of "beckon". On the other hand, words involved in the change from "beck and call" to "beacon call" have longer etymological paths but both ultimately stem from Proto-Germanic **bauknq* ("sign, symbol"). Overall, shared etymology in these target-error pairs could indicate that the historical linguistic reasons for the similarity between two words may contribute to their confusion and the occurrence of eggcorns.

3.3. Semantic aspect of eggcorns

3.3.1. Semantic fields

Since semantics play a defining role in the phenomenon of eggcorns, it was of paramount importance to explore the semantic tendencies of eggcorns. Determining the semantic fields in which they occur and finding common patterns can help identify areas of language that are more prone to the appearance of eggcorns, as well as understand the factors that are conducive to them.

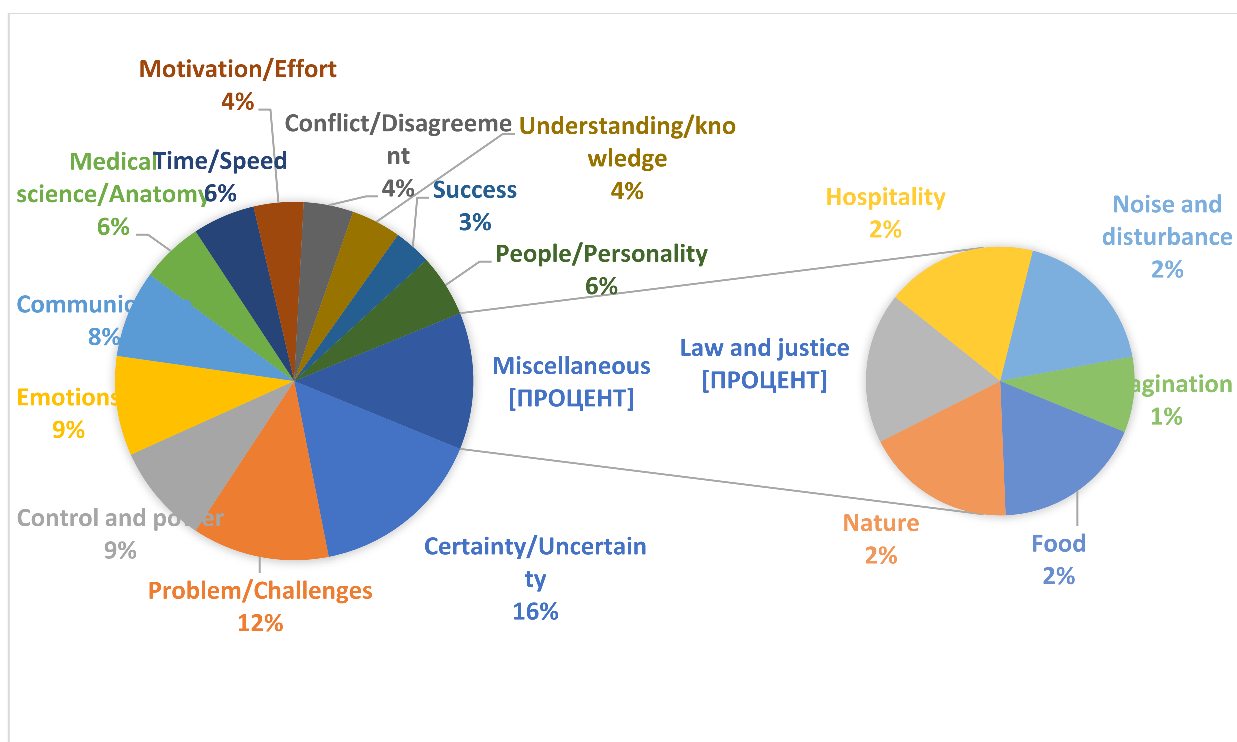

Chart 3.17. Semantic fields

As can be seen from the distribution of semantic fields presented above, the range of semantic fields, to which eggcorns could be attributed, is fairly wide and diverse, with 18 recognized categories in total, which suggests that eggcorns are not exclusive for a particular field or domain of language use and have a chance to appear in a variety of contexts. Overall, the distribution demonstrates that eggcorns arise in semantic fields that are relevant to the everyday lives of speakers and are common and familiar to them. For instance, such semantic fields as "Certainty and uncertainty", "Problems and challenges", "Communication", "Emotions", and "Control and power" are fairly versatile and are central to many aspects of human experience, including relationships, decision-making, and work, which may make them more salient and hence more likely to come up in a conversation, thus becoming more prone to the appearance of eggcorns. On the other hand, more specialized semantic fields, such as "Law and justice", and "Hospitality", were among the least represented (2% each), however, the fact that they may include less frequently used vocabulary and special terms could be another factor that led to the appearance of an eggcorn. Likewise, frequency and restricted field of use may also play a role in the

occurrence of eggcorns belonging to the semantic field of “Medical science and anatomy”.

In terms of imageability, i.e., the degree to which a word may elicit a clear mental image, the vast majority of semantic fields documented in our research tend to be conceptual and include abstract words that are low in imageability. On the other hand, fields that tend to include more imageable, concrete words are less well represented, with “Medical science and anatomy” constituting 6%, and “Hospitality”, “Nature”, and “Food” constituting 2% each out of all eggcorns analyzed. This finding confirms that eggcorns conform to the concreteness effect, i.e., the slower and less accurate processing of abstract, low-imageability words.

Moreover, a large number of semantic fields, including those with higher numbers of eggcorns, tend to carry a negative connotation (“Certainty/Uncertainty” – 16%, “Problems/Challenges” – 12%, “Control and power” – 9%, “Conflict/Disagreement” – 4%, “Noise and disturbance” – 2%). The reason for such a strong presence could be that words represented by these fields are often highly emotionally charged and may be used in stressful situations leading the speaker to experience cognitive load. It is important to mention that the field of “Emotions” itself is one of the most well-represented (9%) in the sample. Furthermore, negative connotations may be associated with abstract and complicated ideas that are more challenging to express linguistically. All of the above may prompt the speaker to rely on more familiar or instinctive language patterns, increasing the likelihood of using incorrect words, especially if they seem acceptable contextually.

3.3.2. Semantic classification

According to S. Reddy, eggcorns can be categorized into two types in terms of their semantic justifications. The first category consists of eggcorns where the changed form or one of its constituents is semantically related either to the target word or to the context represented by the whole lexical unit in which the reanalysis takes place. On the other hand, eggcorns where the changed form is related to an image evoked by the original phrase or word belong to the second type [87, p. 18].

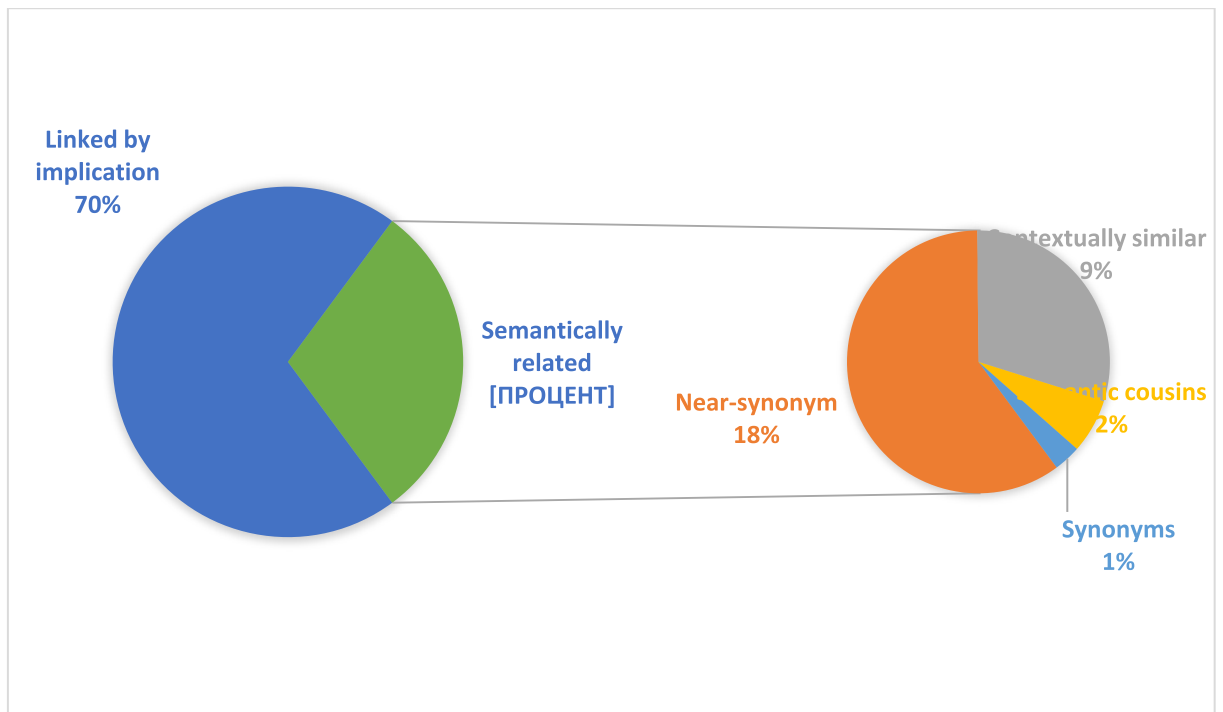

Chart 3.18. Semantic classification of eggcorns

As can be seen on the chart above, the majority of eggcorns (70) included target-error pairs that belonged to the second type, i.e., the error was related to the image or concept implied by the original (also designated as “linked by implication” by Reddy). [87, p. 22]. In 29 instances, eggcorns belonged to the first type, i.e., they were in some way semantically related to the target phrase or word.

Notable, the remaining eggcorn “get one’s dandruff up” could be attributed to both types and displayed a fairly complex semantic justification. The target word “dander” has a meaning of “passion, temper, or anger” when used as part of the idiom “get one’s dandruff up”. However, when used in isolation, it possesses the meaning of “dead skin flakes shed from mammals” and is directly connected to the word “dandruff” (their shared etymology was explained in more detail in section 3.3). Therefore, it seems that the semantic connection follows the progression from the opaque element of the idiom to the individual word describing dead skin flakes, and finally to a more common and semantically similar word “dandruff”. However, the additional activation of the word “dandruff” may be prompted by the idiom itself, meaning “to become angry and agitated”, since it may conjure up an image of a

person becoming so infuriated that they shake their heads, tear their hair out or that their hair stands up on the scalp, thus bringing forward an association with dandruff.

As for the most represented category of eggcorns, which are linked with the original phrase or word by implication, it can be exemplified by the following lexical units: "eggtopic pregnancy" derives from "ectopic pregnancy", which refers to a fertilized egg implanting outside of a person's uterus, hence the choice of the "egg" element; "expatriot" derives from "expatriate" (i.e., a citizen living abroad) and is based on the assumption that a true patriot would not leave their country; "garbledygook" derives from "gobbledygook", coined in imitation of the sound produced by a turkey and used to denote nonsensical or unintelligible use of language, with the component "garbled" (i.e., difficult to understand, unclear) mirroring its meaning; "preying mantis" stems from "praying mantis", with the confusion being understandable, since the target compound refers to an insect that preys on other insects, as well each other (in fact, this is such a distinctive feature of the species that it has come to be used figuratively in relation to predatory women taking advantage of men); "alterior motive" makes use of the more transparent Latin root "alter" in the sense of "other" (due to the familiarity of such words as "alternative" and "alter ego") to fit the meaning of the phrase ("an alternative or extrinsic reason for doing something"); "cacoughany" reinterprets "cacophony" by associating the name for an unpleasant, discordant noise with the sound of coughing. Additionally, an interesting case is represented by the eggcorn "bloodgeon", since the probability of error is heightened by the fact that the target "bludgeon" is of unknown etymological origin, however, the association with violence allows the speaker to select a homophonous morpheme "blood".

It is also worth mentioning that the occurrence of a certain number of eggcorns (11) linked with the target by implication could also be facilitated by the association with other existing phrases. For example, "dawn on someone" could be reinterpreted as "dong on someone" due to the sound of the bell being figurately used to signify understanding and recognition in phrases like "ring a bell", "fullproof" seems to draw

upon phrases like “fully proof against” and “fully secure”, “all total” essentially blends the target “all told” and the phrase “in total”, both of which are synonymous, and “at someone’s back and call” could also be influenced by the idiom “have someone’s back”. In two instances, the idiom “on a wing and a prayer” (meaning “with hope despite little chances”) is reinterpreted through idiom blending: “on a whim and a prayer” combines the original phrase with “on a whim”, connoting spontaneity, while “on a wink and a prayer” is influenced by “a nod and a wink”, suggesting trusting luck. Another instance of an idiom blend is the eggcorn “straight as a narrow”, which seems to be influenced by the idiom “straight (strait) and narrow”, which has a similar meaning to that of the target “straight as an arrow”. Another interesting example of an eggcorn appearing through the association with similar phrases is “don’t know buttkiss”: the original “bupkis” is fairly opaque, since it is borrowed from the Yiddish language, which is a less common source in English than languages like Latin and French, and the eggcorn is made to fit into the group of negative and vulgar idioms such as “don’t know shit” and “don’t know squat”.

As for the second category of semantically related eggcorns, they can be further divided into ones where the changed constituent shares a semantic connection with the original word (near-synonyms, synonyms, or semantic cousins) (21 instances) and ones where the changed component is semantically connected to the context of the whole phrase (9 instances). Examples of near-synonyms include such target-error pairs as “crux” (cross) – “crutch” (forked stick) in “crutch of the matter”, “fell” (fierce, cruel) – “foul” (offensive, unfair) in “one foul swoop”, “brim” (flow over the edge) – “rim” (form an outer edge) in “rimming with tears”, “ruckus” (noisy commotion) – “raucous” (loud, harsh), and “beseech” (ask urgently and fervently) – “besiege” (one of the meanings being “to harass, as with requests or complaints). Moreover, in the case of the eggcorn “much to do about nothing”, aside from sharing common etymology, both “to-do” and the target “ado” are synonyms and designate commotion or fuss. Two of the eggcorns could also be connected to a common general concept (hypernym) and, therefore, characterized as semantic cousins,

according to Reddy [87, p. 22]. Namely, both “marshal” and “martial” (from the eggcorn “marshal law”) have military origins, while “entrée” and “ontray” share the semantic field of food, despite having drastically different etymological origins.

As for the cases where the change was influenced by the context of the original phrase, it is represented by such examples as “damp squid” influenced by the association of a sea mollusk with wetness, “put the cat before the horse”, where the semantic field of animals is evoked by the word “horse”, and “strings and arrows”, where the changed form “strings” references bowstrings due to the mention of arrows. It is important to note that the cases of contextual anticipation mentioned in section 3.2.4, are not exclusively formal but also semantic: in “Pipe Piper”, the original and relatively uncommon adjective “pied” (multicolored) is reanalyzed as “pipe” by drawing on the following word “piper”, and the resulting eggcorn might be understood as “someone who plays on a pipe”; on the other hand, “bread” in “bread and breakfast” is clearly influenced by the context, which evokes the semantic field of meals and food.

Furthermore, during the investigation of the semantic aspect of eggcorns, several other intriguing observations were made. For instance, in eggcorns like “free reign”, “shoe-in”, “take another tack”, “with hammer and thongs”, and “bellweather”, the original phrases became opaque in the course of time since the fields of human activity they refer to have lost their relevance (namely, horse-drawn carriages as means of transportation, horseraces, sailing, and farming) and, therefore, are substituted by words more familiar and universal in relation to the modern member of the speech community. Another interesting observation was that some eggcorns resulted in a change of connotation or even a semantic shift. For example, the metaphor of removing lice from hair is exchanged for a less unpleasant metaphor of picking pills of fabric from knitted clothes or picking the knitting apart in the eggcorn “pick knits”. Likewise, in the “benighted” – “beknighted” pair, despite the quite vague semantic link between the target and the error, perhaps influenced by the association of knights with the so-called Dark Ages, the resulting eggcorn lends a

more positive connotation to a word that is used to describe the state of intellectual or moral ignorance. The appearance of the eggcorn in these cases may be motivated by the desire of the speaker to avoid using negatively charged language (the same may apply to eggcorns like “social leopard” (instead of “social leper”) and “can’t be asked” (instead of “can’t be arsed”).

As for the change of meaning, the eggcorn “damp squid” derived from the idiom “damp squib”, which denotes a disappointing situation, can be used in the sense close to that of “wet blanket” due to analogy, as can be seen from this example of usage taken from the KVR Audio forum and provided by on the Eggcorn Database website: “Hate to be the damp squid but I hope that the low cpu usage doesnt [sic] come at the expense of sound quality which the reputation of vas [sic] was built on?” [119]. Another example of a semantic shift is represented by the eggcorn “beckon the question” based on the saying “beg the question”. Interestingly, the original phrase is often misused by the speaker in the sense of “raising a question”, while it actually refers to the logical fallacy in which one assumes the truth of an argument that has to be proved (i.e., a circular argument). Therefore, the eggcorn “beckon the question” seems to be synonymous with the mistaken usage of the phrase, since the verb “beckon” lends the meaning of “inviting the question” to the eggcorn. Overall, the ability of eggcorns to shift the connotation or meaning of the original phrase proves that they represent one of the multiple processes behind the language as a dynamic and constantly evolving system shaped by how speakers use it.

Chapter conclusion

In this chapter, we have intended to explore the phenomenon of eggcorns and attempt to determine its role within the larger framework of other language errors. Eggcorns are typically defined as a type of linguistic error where a word or a phrase is substituted with one that is phonetically similar and can be semantically justified. Similarly to malapropisms, eggcorns can be classified as competence errors rather than processing or production errors (i.e., the speaker is not aware an error is being made). However, unlike malapropisms, this type of linguistic error does not stem

from speakers unsuccessfully attempting to use heightened and sophisticated language and often makes use of slang, cliches, and jargon. Nevertheless, due to being a fairly new linguistic concept, the precise definition of an eggcorn, its defining characteristics, and its place within the classification of linguistic errors (particularly, the issue of whether eggcorns are a subtype of malapropism or a completely separate type of linguistic error) is still being debated by linguists.

As for the neighborhood characteristics of eggcorns, the analysis showed that in this aspect eggcorns more closely resemble speech perception errors rather than speech production errors, since neighborhood density and neighborhood frequency counts tended to be higher in target words than in the erroneous ones. This is consistent with the reports from speech recognition experiments, which posit that words with large phonological neighborhoods and high-frequency neighbors are recognized slower and less precisely than words with few neighbors. Overall, this finding lends credibility to the theory that competence errors result from faulty storage in the mental lexicon, which in some cases can be the consequence of uncorrected slips of the ear, i.e., speech community members incorrectly perceive the unfamiliar lexical unit and record it in their mental lexicon as such, without realizing that the corruption occurs. When considering both neighborhood frequency and density, there was typically a positive correlation between the two, suggesting that the more neighbors a word involved in an eggcorn possesses, the greater the likelihood that they will be of higher frequency. Nevertheless, considering a relatively small disparity in the average figures, the influence of neighborhood density and frequency of target words on the occurrence of eggcorns is not as significant as in the case of errors that are solely connected with processing or production, which suggests that other factors may play a more decisive role.

On the other hand, the present study found that phonological similarity between words plays a major role in the occurrence of eggcorns, and phonological neighbor pairs are more likely to result in eggcorns than non-neighbor pairs. The investigation also showed a tendency for the target and error to share identical

pronunciation, with a large number of homophones among the target-error pairs, which provides further credibility to the theory of the slip-of-the-ear origin of eggcorns. The effect of phonological similarity was explored in greater detail by considering the phonological distance between the pairs of targets and corresponding errors, with results suggesting that in eggcorns, the phonological distance between the target and error tends to be relatively small. The results demonstrate that even the least phonologically similar target-error pairs in the sample share a considerable degree of resemblance, which could contribute to the occurrence of eggcorns.

Subsequently, we investigated the role of word frequency in the occurrence of eggcorns, determining that the erroneous word tends to possess a higher frequency than the target word. Moreover, in the instances when the target was more frequent, the range of differences in frequency between the target and error was fairly small. This suggests that the difficulty of retrieving low-frequency words from the mental lexicon leads to a higher likelihood of substitution with more familiar and accessible words that are more frequent in usage. Furthermore, it was found that target words with a positive correlation of word frequency and neighborhood density (i.e., low-frequency words with small neighborhoods and high-frequency words with large neighborhoods) have a considerably bigger chance of being involved in an eggcorn, which is consistent with previous findings on malapropisms.

Since the conformity of the word to the phonological rules of a language plays a decisive factor in how effectively this word is processed, it was of paramount importance to investigate the tendencies of eggcorn formation in terms of phonotactic probability. It was established that eggcorns tend to occur more often with target words that have low phonotactic probability, most likely due to the fact that such words are more difficult to process and store in the mental lexicon. Other interesting observations include the fact that speakers tend to substitute low-probability words with words that have the same or higher degree of probability, as well as the fact that even in cases when high-probability words are substituted with less probable ones, they still tend to have relatively common sound patterns.

When studying eggcorns, it is also important to consider the presence of the non-word legality effect, which refers to the processing advantage of non-words conforming to legal language patterns. Although was determined that most of the non-word eggcorns in our sample had low phonotactic probability, they still conformed to the phonotactic rules of English, and only rarely resulted in improbable sound sequences. Intriguingly, non-words also displayed the capability of emerging as more phonologically probable than the existing target word, although such instances were rare.

As for the morphosyntactic characteristics of target lexical units, the majority of them consisted of multiple words or free morphemes, suggesting that eggcorns are more likely to occur in complex words and phrases. Moreover, it was found that phrasemes are significantly more prone to be transformed into eggcorns, largely due to their opaque semantics. Interestingly, despite the clear tendency of target words to contain multiple elements (words or morphemes), the number of the said elements tended to be relatively small. This is likely because shorter phrases are easier to remember and repeat, while eggcorns tend to stem from popular, cliched phrases rather than complex language. The majority of eggcorns had the same number of elements as their target expressions, indicating that speakers tend to preserve the overall structure of the original phrase or word.

Subsequently, eggcorns were categorized according to their formal characteristics. For instance, the distribution of eggcorns according to the unit involved in the transformation showed that eggcorns tend to result from erroneous word or sound selection. Sound substitution was the most common mechanism of creating eggcorns, followed closely by homophone selection, while morpheme addition and sound deletion were rare. It is important to note that the nature of word substitution in eggcorns differed from what was observed in ordinary production errors, with the similarity between the target and error in word substitutions being significantly more pronounced. A significant discovery was the resemblance between some of the processes involved in eggcornization and folk etymology (i.e., the

historical reinterpretation of unfamiliar lexical items through analogy with more familiar ones). For instance, the process of metanalysis (resegmentation of morphemes), which often plays a prominent role in folk etymologies was also recorded as one of the means of forming eggcorns. Overall, such findings as the prevalence of short phonological distances in word substitutions, homophone selections, resegmentation of morphemes, as well as morpheme additions and substitutions involving a single sound change suggest that eggcorns serve as a testament to speakers' capacity to plausibly reinterpret confusing linguistic information with minimal alteration of the original sound structure.

Since throughout the entirety of our research eggcorns consistently exhibited the tendency to substitute unfamiliar or opaque elements with more usual ones, the part of the etymology in this process was also examined. Although most of the target-error pairs were of English origin, there were also multiple instances of an English-origin word substituting a word of foreign origin or a non-integrated borrowing. Moreover, a significant number of target-error pairs shared common etymological roots, which could signify that this may also be one of the factors leading to the confusion of two words and the subsequent occurrence of eggcorns.

Having conducted a comprehensive analysis of phonological, morphosyntactic, and etymological factors that could collectively lead to the emergence of eggcorns, we turned the focus of our research to the semantic aspect. The range of semantic fields, in which eggcorns occurred, proved to be quite diverse, with 18 categories in total, indicating that they can occur in various contexts. Eggcorns often appear in semantic fields that are essential to human experience and are highly relevant to speakers' everyday lives, such as "Certainty and uncertainty", "Problems and challenges", "Communication", and "Emotions". Imageability was another important factor, with eggcorns tending to occur in conceptual, abstract fields that have low imageability. Moreover, semantic fields, the words in which usually carry negative connotations, were also prevalent, which may be explained by the fact that emotionally charged words are often used in stressful situations, prompting the

speaker to experience cognitive overload and rely on more familiar language patterns, particularly when they seem semantically justifiable.

Finally, the eggcorns were categorized according to their semantic justification. Most of the analyzed eggcorns belonged to the group where the changed form was related to an image or concept evoked by the target by association (e.g., "praying mantis" – "preying mantis"). Interestingly, a significant number of eggcorns belonging to this category were likely produced by way of analogy with other existing phrases. The second, less frequent group consisted of the eggcorns where the changed form was semantically related to the target word (e.g., "ruckus" – "raucous" or the context of the whole phrase (e.g., "Pipe Piper"). In studying the semantic aspect of eggcorns, several other interesting observations were made. For instance, certain eggcorns replaced phrases, the meaning of which became opaque with time due to the outdated character of the phenomena they were originally referencing, while others led to the change of connotation or a semantic shift. These observations demonstrate that eggcornization may be one of the stepping stones in the process of linguistic innovation and ultimately contribute to the constant evolution of language and its meaning.

CONCLUSION

In this paper, we have investigated the phenomenon of eggcorns within the broader notion of speech errors, determining common patterns and characteristics in their occurrence that allow us to arrive at a better understanding of this relatively new concept.

To ensure the scientifically grounded theoretical foundation for our research, we first focused on providing an extensive review of previous research and theories on psycholinguistic phenomena and concepts essential for understanding the issue of speech errors. In particular, the concept of the mental lexicon as a speaker's mental dictionary was closely examined. Consequently, we discussed the various processes

in word recognition, such as the activation of nodes in the mental lexicon based on auditory information and contextual cues, as well as outlined the role of various factors, such as neighborhood effect, lexical status effect, and non-word legality effect. Similarly, we considered the processes behind the production of words, examining speech production models, paying special attention to the theory based on speech error research, which argued for the interactivity of different stages of word production and levels in the mental lexicon. Factors influencing word production such as output biases, neighborhood, similarity, speech rate, distance, and mixed-error effects were described in detail. Despite the prevalence of research on faulty storage in the mental lexicon specifically in relation to individuals with aphasia and language impairments, the good-enough language processing theory argued that similar deficiencies are also observable in healthy individuals due to superficial analysis of linguistic input.

Subsequently, the dominant classifications of word production errors according to different aspects were detailed, including such conventionally accepted categories as form-based and meaning-based errors, as well as categories such as substitution, transposition, omission, addition, and movement. Moreover, it was discovered that in many aspects evidence from speech perception errors aligns with the results of the research on speech production errors, for instance, confirming the division of speech into the separate units during processing and the existence of the unconscious knowledge of phonotactic rules of a language. It is important to mention that special attention was paid to the notion of malapropism due to its unique nature as a competence error, as well as its close relation to eggcorns.

The consistency and validity of the research were guaranteed by the implementation of a wide array of qualitative and quantitative approaches. Firstly, the random sampling technique was used to extract a sample of 100 eggcorns from the Eggcorn Database to avoid potential biases and ensure the representativeness and generalizability of the findings to the larger population of eggcorns. Subsequently, a combination of computational and corpus-based approaches was used to analyze

phonological patterns characteristic of eggcorns, as well as to consider such factors as neighborhood density and word frequency. In addition, a descriptive qualitative approach was used to analyze the formal and semantic characteristics of the collected data and identify key themes and patterns. Lastly, the functional semantic field approach was used to categorize eggcorns into semantic fields and identify features of their use in specific contexts.

Having established a sufficient theoretical foundation and the most efficient methods of research, we turned to the practical analysis of our eggcorn sample, aiming to determine their common features and factors contributing to their occurrence. The analysis showed that phonological similarity between target and corresponding error words plays a significant role in eggcorn formation, with phonological neighbor pairs, as well as homophonous pairs, more likely to result in eggcorns than non-neighbor pairs. Additionally, the study found that word frequency is also a crucial factor, with the erroneous word typically having a higher frequency than the target word. The positive correlation between word frequency and neighborhood density was found to be a contributing factor to the occurrence of eggcorns, which is consistent with previous findings on word production errors. The study also explored the role of phonotactic probability in eggcorn formation, with the results demonstrating that low-probability words have a strong probability of being involved in an eggcorn. Moreover, the reinterpreted segment of the eggcorn tends to be either more phonotactically probable than the target or at least display a relatively common sound pattern for the English language. The investigation also confirmed the existence of the non-word legality effect in the process of eggcorn formation, since non-word eggcorns largely followed the phonotactic rules of English.

Furthermore, the study also highlights the importance of considering the morphosyntactic characteristics of eggcorns, with eggcorns showing a strong tendency to occur in complex words and phrases, particularly phrasemes with opaque semantics. The investigation of etymology further confirmed the propensity of

eggcorns to result from the reinterpretation of unfamiliar elements, as well as highlighted a trend of target-error pairs to share a common etymological origin.

The formal categorization of eggcorns revealed the capacity of speakers to come up with a plausible reinterpretation of opaque items through minor alterations of the overall phonological and syntactic structure of the original phrase. The use of the functional semantic field approach helped determine that eggcorns can occur in a wide variety of contexts, with the prevalence of semantic fields relating to basic human experience, displaying low imageability and carrying negative connotations, which confirmed the universal and emotionally charged character of eggcorns. In general, the exploration of the semantic characteristics highlighted the complex and varied ways of selecting semantically justifiable lexical items, which lead to the creation of an eggcorn. Importantly, the evidence that eggcorns can lead to changes in connotation and meaning, as well as be produced through analogy with other common phrases emphasizes their potential role in the process of linguistic innovation.

Overall, the aforementioned research results allow us to make several generalizations and assumptions about the nature of eggcorns and their place in the general framework of language errors. Firstly, eggcorns displayed characteristics of both word production and word perception errors. For instance, they largely adhered to such word production effects, as the mixed-error effect and similarity effect, including the categorical constraint. On the other hand, the neighborhood density and neighborhood frequency effects observed in eggcorns were more in line with the data on word perception errors, since previous research shows that words with large phonological neighborhoods and high-frequency neighbors are recognized less effectively. Moreover, eggcorns also showed the influence of the non-word legality effect typical for word processing. As for the formal mechanisms leading to the creation of eggcorns, the combination of processes typical for word production errors (such as word substitution and morpheme addition) and word perception errors (e.g., homophone selection and metanalysis) could be observed. This suggests that

eggcorns represent a hybrid type of language error, which appear as a result of reinterpretation and incorrect storage of incorrectly processed lexical items.

Furthermore, it can be concluded that the combination of such factors as low frequency and phonotactic probability, opaque semantics, and foreign etymology of a lexical item leads to a high probability of it being involved in an eggcorn. Moreover, it was shown that eggcorns usually involve cliched and emotionally colored phrasemes that relate to basic fields of human activity and experience.

Finally, the close relatedness of processes behind eggcorns and folk etymology, semantic plausibility, the tendency to replace outdated lexical items, as well to change the connotation and meaning of the original utterances lead us to conclude that, aside from being ordinary language errors, eggcorns may represent a stage in a productive process such as folk etymology, offering insights into the dynamics of language change as well as the underlying mechanisms and motivations driving it. In this context, the study of eggcorns is particularly important because it can demonstrate how speakers adapt to changing linguistic and cultural environments by reinterpreting the form, meaning, and usage of language through common errors.

SUMMARY

The phenomenon of speech errors offers an invaluable source of insights into the processes behind human processing and production of language since breakdowns in a complex system are usually revealing of its inner structure and functioning. The research on speech errors has provided numerous important contributions to a variety of different fields, such as psycholinguistics, cognitive linguistics, and speech therapy. For this reason, speech errors have been extensively studied since the late 19th century. However, this field of research continues to beget competing theories, and at the present moment, linguists are still attempting to find consensus on its many issues.

In particular, a relatively new concept of eggcorns coined in 2003 has recently gained considerable attention as a unique type of language error that involves both phonological similarity and semantic plausibility. This research aims to closely examine eggcorns, determine their defining characteristics and patterns, and understand their place in the broader framework of speech errors. проведение исследования Studying language errors is crucial for understanding language processing, and eggcorns present a novel opportunity to explore how speakers use and interpret language. Despite the recognition of eggcorns as a valuable resource for the investigation of language storage and processing intricacies, the number of studies tackling this subject remains sparse. Therefore, this paper aims to thoroughly examine the notion of eggcorns in detail, determine the patterns and effects that lead to their occurrence, as well as their place in the broader spectrum of speech errors.

In this research, a randomly selected sample of 100 eggcorns extracted from the Eggcorn Database was analyzed according to a variety of formal and semantic characteristics, such as neighborhood density, phonotactic probability, frequency, etymology, and semantic field distribution, using a combination of qualitative and quantitative approaches. The research was assisted by the use of various linguistic tools, such as special software for phonological analysis, as well as phonological and psycholinguistic corpora.

The research findings suggest that eggcorns represent a mixed type of language error that display characteristics typical for both word production and word perception errors, providing support for the theory that competence errors like eggcorns are manifested as a reinterpretation of linguistic input as a result of faulty processing and storage. Moreover, a number of factors increasing the likelihood of eggcorn occurrence were identified, including low frequency and phonotactic probability, non-compositional meaning, and foreign etymology of the target lexical item. In the process of our research, eggcornization was shown to affect mostly fixed and emotionally charged phrases that can be attributed to a wide range of semantic fields, often relating to basic human experience. Lastly, the fact that eggcorns

mirrored many of the processes and motivations behind folk etymology, had high semantic plausibility and the possibility of replacing obsolete terms, and be involved in the change of connotation and semantic shift points to a conclusion that eggcorns may not only reveal important information about language processing and production but also shed light on the ongoing changing of language and the processes involved in it.