IZABELA SEKŚCIŃSKA
ORCID: 0000-0002-9974-0513
Państwowa Wyższa Szkoła Informatyki i Przedsiębiorczości w Łomży, Poland

The role of inner speech in the speech production process

1. Introduction

Briefly defined as "silent verbal thinking" (Morin 2012:436), inner speech appears to be indispensable for language-related processes and mental operations. However, with no overt behavioural manifestation, it is elusive and nebulous, and hardly possible to measure empirically. As a cognitive process, inner speech is unique and strictly private. It operates within the confines of our minds and is typically investigated by means of self-report methods, such as questionnaires or experience sampling techniques (Alderson-Day/Ferynhough 2015). Although neuroimaging techniques have also been successfully used to research into inner speech production (Ren et al. 2016), it remains a highly understudied phenomenon. Relatively little is known about "the extent and nature of everyday inner speech use" (Alderson-Day/Ferynhough 2015:1), let alone about individual differences in this everyday inner experience. Yet, the existence of the inner voice cannot be denied, as well as its impact upon our mental lives and language abilities. The intent of this paper is to provide an overview of the study of this psychological process and to show its role in language production and slips of the tongue phenomena.

2. Definition of inner speech

The inner voice inside our heads accompanying our daily activities has attracted a great deal of scientific attention. However, as noticed by Alderson-Day and Feryhhough (2015:931), "[d]espite a growing body of knowledge on its phenomenology, development, and function, approaches to the scientific study of inner

speech have remained diffuse and largely unintegrated". Suffice it to say that a great many expressions have been devised to refer to this significant human experience. Morin (2012:436) enumerates the following terms associated with this phenomenon: 'subvocal speech', 'self-talk', 'self-directed speech', 'subvocal, covert or accommunicating speech', 'auditory imagery', 'speech-for-self', 'propositional thought', 'self-verbalisations', 'internal dialogue or monologue', 'subvocalisations', 'self-statements', and 'silent verbal thinking'. Alderson-Day and Ferynhough (2015:931) add to this list 'subjective experience of language in the absence of overt and audible articulation', 'inner speaking', and 'covert self-talk' and emphasise that this concept is often used interchangeably with 'thinking'. Connected with the term 'inner speech' is also Piaget's 'egocentric speech' and Vygotsky's 'private speech', regarded as a transitory stage between public or social speech and inner verbal thinking (Vygotsky 1986).

Besides great diversity in terminology, there is also no unified definition of inner speech and, what is more, this concept is completely disregarded in dictionaries of linguistics. The difference between 'langue' and 'parole', as well as the definitions of 'language' and 'speech', are provided in most dictionaries such as Peter Matthews's "Concise Oxford Dictionary of Linguistics" (2014), David Crystal's "Dictionary of Linguistics and Phonetics" (2008), Kirsten Malmkjær's "Linguistics Encyclopedia" (2002) and Vyvyan Evans's "Glossary of Cognitive Linguistics" (2007). However, none of the aforementioned dictionaries contain an entry for 'inner speech' which, as underlined by Alderson-Day and Ferynhough (2015:931), "appears to perform significant functions in human cognition". The only dictionary of linguistics that the author of the paper managed to find and which offers the definition of 'internal language' is Hadumod Bussman's "Routledge Dictionary of Language and Linguistics" (2006:585). According to the dictionary, 'internal language' is "[l]anguge form that serves not as a verbal communication but rather as a vehicle for thinking", which is contrary to the assumptions adopted by John R. Johnson, a professor in the Department of Communication at the University of Wisconsin. Johnson (1994) views inner speech as a form of intrapersonal spoken language which is directed at the individual and signifies communication with one's self and claims that alongside the interpersonal level of speech communication there coexist two types of intrapersonal communication: (1) inner speech which is covert and inaudible, and (2) self-talk which is a silent and audible self-directed monologue (Johnson 1994). Johnson's assumptions are in line with Baars and Gage (2010:50) and Morin (2012:436) who also underline that inner speech is 'the activity of talking to oneself in silence'. Similarly, Mazur et al. (2006:113, own translation) claim that "speech signifies the ability to communicate with the surrounding environment and one's self", adding that there are three main layers of communication: limbic, verbal, and inner language. All those layers interact with one another and form a feedback loop linking limbic (emo-

tional) language to verbal and inner language. Stanisław Grabias (1997:10), on the other hand, claims that speech encompasses two types of linguistic behaviours: (1) inner speech, which results in silent production of words in one's minds, and (2) external speech, usually referred to as linguistic or verbal communication or interaction. Bearing in mind that linguistics is defined as "[t]he scientific study of language" (Crystal 2008:283), inner speech, which is also regarded as "a streamlined version of outer speech" (Wiley 2006:320), should undeniably have its place in dictionaries of linguistics, and its connection with external language should be investigated by linguists as well.

3. Inner speech features

Although inner speech and external speech "exist in a symbiotic relationship" (Johnson 1984:214) and rely upon the same linguistic system, they are quite different and exhibit diverse functions. Apart from language-related operations, inner speech appears to be essential for working memory functioning, short-term memory enhancement, self-reflective abilities, and mental rehearsal (Geva et al. 2011:3071; de Guerro 2005). Its impairment may lead, for instance, to the slowing of mental faculties, tachyphrenia, logorrhea, and absent-mindedness (Michalik and Siudak 2010). There is no denying that inner speech is also conducive to language acquisition and second language learning (de Guerro 2005); in fact, "[a]ll aspects of normal language functions (e.g. reading, writing, speaking and calculating) require intact inner speech, and indeed, loss of inner speech following brain damage invariably leads to aphasia, agraphia, alexia, acalculia, and impaired verbal short-term memory" (Morin 2009:395). Significant for consciousness and self-awareness, inner speech is also reported to be vital to autobiographical memory and self-regulation and plays an important role in decision making, emotional release, and task switching performance (Morin 2009, 2012).

As neuroscience reveals, inner and overt speech yield overlapping cortical language areas, but covert speech activations tend to be much weaker than those of overt speech (Oppenheim 2001). Additionally, inner and external speech may "produce separate activations in other regions of the brain, reflecting distinct non-motor cognitive processes" (Geva et al. 2011:3072). Neuroimagining research confirms that inner languaging involves the instigation of the left inferior frontal gyrus which houses Broca's area involved in language processing, as well as the activation of Wernicke's area, "supplementary motor areas, insula, and superior parietal lobe on the left side, [and] right posterior cerebellar cortex" (Morin 2012:439). Surprisingly, however, Huang et al. (2001) notice that Broca's area is more active during internal language processing than during overt production, which is also reported by Morin (2009, 2012), who adds that any destruction to the left inferior

frontal gyrus leads to the disruption of inner speech. Although this inner voice is regarded as 'subvocal speech' with no mouthing involved (Morin 2009), it appears to exhibit slight activation within the speech-motor cortex as well. Morin (2012) claims that the inner experience of verbal thinking is even accompanied by minor movements of the mouth and tongue muscles, which was also predicted by Sokolov (1972:65), according to whom "even if thoughts are communicated immediately as they occur before they become expressed through external speech they still are preceded by a discharge of motor speech impulses […] they are always antecedent to the utterance of words, be it even a matter of fractions of a second". What is more, as reported by Scott et al. (2013), inner speech is sonorous to the individual, thanks to corollary discharge[1] that enables us 'to hear' the internal voice and to differentiate our voice from other voices or sounds and allows us to talk with ourselves inside our minds.

Geva et al. (2011) underline that inner speech should not, however, be treated as the internal equivalent of overt speech or as external speech without sensorimotor activation (Geva et al. 2011). It is a unique mental operation that constitutes "around one-fourth of [our] conscious walking life" and therefore most certainly is "a significant human mental activity" (Morin 2012:436). Involved in linguistic production, inner speech is believed to constitute a mental draft for overt production (Sokolov 1972, Vygotsky 1986). Morin (2009:393) underlines that inner speech does not provide a detailed plan for external speech, but rather "a fragmentary series of verbal images" on which bases linguistic message is construed. Levelt (1993), on the other hand, regards it as an 'articulatory buffer' that temporarily retains the phonological representation of a linguistic message until it is executed by the articulators, or until it is decoded into meaningful units during the comprehension process.

Noteworthy as it is, the inner voice is believed to rely upon various modalities to a great extent. Not only does it draw upon personal experience but it also makes use of sensory engrams stored within modality-specific memory subsystems, be it visual, auditory, kinesthetic, and gustatory (Wiley 2006). As a result, inner speech evokes only some key linguistic expressions to be externalised as it receives a cognitive backup from regions related to perceptual and sensorimotor processing. The syntax of inner speech is thus not elaborative. On the contrary, it is extremely condensed, abbreviated and predicative (Johnson 1994), and sometimes even chaotic or fragmentary (Grodziński 1976). What is more, it lacks the main parts of speech, which do not essentially need to be internally voiced for the gist of the message to be construed. For instance, the subject of the sentence, which

---

[1] "Corollary discharge is a neural signal generated by the motor system that serves to prevent confusion between self-caused and externally-caused sensations" (Scott et al. 2013:EL286).

is obvious to the speaker, is typically omitted in inner speech, as well as information known to the speaker. Grodziński (1976) even claims that inner speech is composed of word images and lacks a formal syntax peculiar to external speech, whereas Wiley (2006) emphasises that this abbreviatedness of internal language is liable to change, as it is determined by the situation and the task that the speaker is involved in or reports about. Undeniably, some activities, such as mental rehearsal before public speaking or an important exam, will require the use of elaborated inner speech but, for instance, making a shopping list will not evoke a word-by-word version of inner speech, but rather an extremely condensed message which will draw extensively upon sensory modalities.

Supported by mental imagery, the semantics of inner speech are characterised by semantic embeddedness and agglutination (Johnson 1994). Internalised concepts tend to infuse into hybrid units with a much broader meaning, which is emphasised by Vygotsky (1986:247) who claims that "[a] single word is so saturated with sense that […] [t]o unfold it into overt speech one would need a multitude of words". De Guerro (2005) even assumes that the 'little voice in the head', as inner speech is often referred to, could also be incomprehensible to the interlocutor if 'translated' directly without all the necessary alternations or expansions. The internal language has a rather 'messy', non-linear structure, whereas overt speech is organised and subject to a number of linguistic constraints.

As a streamlined version of external speech, inner speech appears to be much faster than its overt counterpart (Johnson 1994; Wiley 2006). Korba (1990) estimates the rate of inner speech at 4000 words per minute and Coltheart (1999) claims that it may be even 15–25% faster than external speech. Wiley (2006), to the contrary, assumes that inner speech is ten times as fast as its overt counterpart but also notices that the speed of the internalised language is likely to change as it is influenced significantly by the mental state of the individual and contextual circumstances (Wiley 2006).

Silent verbal thinking is a common mental experience that is salient not only to an array of cognitive phenomena, but also influences the generation of external speech to a great extent. It is a never-ceasing process that takes place alongside and virtually simultaneous to linguistic processing that leads to the actual speech articulation. Operating within the confines of our minds, inner speech occasionally lays itself bare and slips into overt speech production, thus causing disruption in communication and the listener's confusion. Such speech errors or slips of the tongue phenomena are tangible evidence of the inner workings of our minds and need to be regarded as a vehicle for the analysis of the internal language. Before, however, the relation between inner speech and verbal blunders is discussed, the process of speech production will be elaborated on, as this knowledge is crucial for the acknowledgement of inner speech as playing a significant role in overt speech generation.

4. Speech generation model

Levelt's (1993) model of language processing assumes that there are three main processing stages involved in speech generation: (1) conceptualization, (2) formulation, and (3) articulation. At the conceptualization level, a communicative intention arises and a preverbal message is formulated. Subsequently, this preverbal message is subject to linguistic encoding, and the formulation process which involves the conversion of the abstract conceptual structure into meaningful linguistic output begins. At this stage of speech production, appropriate lemmas are extracted from the mental lexicon and organised into a linear surface structure in accordance with grammatical and morphological constraints of a given language. Once appropriate inflectional endings are assigned to linguistic units and the surface structure of a message is partially construed, phonological encoding may proceed.

Significant as it is, in normal speech, two or three words per second are retrieved from the mental lexicon and subject to morphological, phonological, and phonetic encoding (Levelt et al. 1999). The process is, however, not that time-consuming, as lexical entries in the mental lexicon contain not only semantic and syntactic information about a given lemma, but also its morpho-phonological representation (Levelt et al. 1999, Levelt 2001). Hence, syntactic properties and diacritic parameters of a given lemma, as well as the articulatory gestures of a word, are readily available and the speech production may be carried out relatively smoothly. Inner speech appears to interact with working memory during the speech generation process as the phonetic or articulatory plan of the message is buffered as inner speech until articulation. The phonetic plan for a given lemma may, however, be stored only for a short period of time, owing to the limited capacity of working memory. Buchsbaum (2013:1) claims that only three or four items may be held in working memory at a time, and Marvel and Desmond (2012:42) emphasise that a subcomponent of working memory – the phonological loop – "is comprised of 1-2 seconds of passive storage of phonological content (i.e., sounds, words, and phrases), which is followed by a secondary active rehearsal process that retains this information beyond 1-2 seconds".

According to Levelt (1993), the speech generation process involves the activation of processing subsystems. The main component subsystems involved in the generation of overt speech are as follows:

(1) the Conceptualizer, which transforms the communicative intention into a preverbal message;
(2) the Formulator, which assigns a linguistic structure to the preverbal message;
(3) the Articulator, which executes the phonetic plan by triggering relevant articulators.

However, as stratificational linguistics proves, these three processing components are rather redundant notions. Stratificational linguistics views language as

"the code relating conceptual and phonic correlations" (Lockwood 1972:3) or, to use Sullivan's (2011) and Garcia et al.'s (2017) terms, the code relating a non-linear embodied conceptual and executive cognition with the oral-aural interface of the linguistic input/output. Hence, once a communicative intention arises, the conceptual thought undergoes linguistic structuring and, via the spreading activation theory (Dell 1986), the internal representation of a sentence 'goes through' all the linguistic levels and is translated into a verbal message. Undoubtedly, it is the stage when a linguistic message is formulated, but there seems to be no need to introduce component subsystems, as it is the linguistic system itself that triggers the activation of relevant linguistic units at appropriate linguistic strata. Significant as it is, structural layers interact with one another and operate by means of parallel processing (Lamb 2004). The encoding process at the upper level only slightly precedes the operations at the lower stratum, and the phonological representations evoked by lexical items are generated "in much the same time frame as that of articulation" (Dell 1986:289).

The stratificational model of the linguistic system provides a detailed account of how language works and shows that during the formulation stage interrelated linguistic elements are triggered by a spread of activation. Given that language is viewed as a network of relations, an activation in one part of the network evokes parallel activation of other interconnected nodes, which more often than not necessitates the spread of activation even beyond interstratal boundaries. According to the stratificational model, language is composed of at least four main strata: sememic, lexemic, morphemic and phonemic (Lamb 1966, Lockwood 1972). Each stratum has level-specific rules referred to as tactic patterns which determine the arrangement of intrastratal elements. Tactic patterns are also connected to a higher stratum by means of the alternation pattern which enables the interstratal transmission of neighbouring elements. As underlined by Sullivan (2011), such organisation and structuring of the linguistic system provides a smooth transition between the cognitive store and the oral-aural interface.

The speech production model proposed by the author of the paper coalesces Levelt's assumptions (1993) with the stratificational approach to language. Although Figure 1 also shows the direction of the decoding process, the stages of speech comprehension will not be discussed as it is the production process that the article focuses on. The speech production process starts with the intention to communicate. In consequence, a preverbal message is formulated and the linguistic encoding process ensues. As underlined by Sullivan and Tsiang (2011:359), activation "starts spreading gradually through the semotactics.[2] [...] Activated sememes are grouped and groupings of sememes are linearized. A short while after

---

[2] Semotactics is "the collection of ordering patterns specifying how sememes [structural units at the semological strata] may combine in a language" (Garcia et al. 2017:211).

the semotactic processing begins, activation starts spreading to the lexotactics[3] via semo-lexemic relations". Appropriate lexemes, together with their internal syntax, are retrieved from the mental lexicon. Each lexeme additionally activates a network of connections that incorporate linguistic elements reflecting not only semantic and syntactic information about a given lemma but also its morpho-phonological representation. Almost simultaneously, the tactics specifying the patterns of arrangement between intrastratal elements start operating and realizational relations between units of adjacent strata are established. Moreover, lemmas are reported to be associated with the appropriate articulatory gestures essential for its enunciation. Significant as it is, articulatory gestures may also be computed online (Levelt 1993) but, as Croot and Rustle (2004:376) state "pre-storage of articulatory commands for high-frequency syllables would efficiently reduce the online computational load in speech production".

Lamb (1966) and Garcia et al. (2017) notice that such languages as Polish and English exhibit six linguistic levels. Apart from the sememic and phonemic level, they distinguish two additional polar levels referred to as hypersememic and hypophonemic. The hypersememic stratum is part of semology. In contrast to the sememic level, which deals with the meaning of clauses and sentences, the hypersememic stratum (a.k.a. gnostemic) is connected with the organisation of knowledge and therefore regarded as part of cognition (Garcia et al. 2017). Regarded as "a sort of middle ground between language and non-language", the hypersememic strata "represents as nearly as possible the natural way of thinking about things" (Lockwood 1972:165). The hypophonological level of language represents the form of the linguistic structure just before articulation (Lamb 1966) as it specifies the actual realization of a phoneme in a given context.

Apart from the levels of speech processing, Figure 1 also presents speech monitoring loops that are to secure the proper realization of the communicative intention. Postma (2000) distinguishes three levels of speech monitoring: conceptual, inner, and auditory. Before being translated into a linguistic structure, the preverbal message is examined by the conceptual loop to determine whether the message fits the context and is relevant to the circumstances or situation. Subsequently, the articulatory plan is inspected via the internal speech monitoring system (a.k.a. inner loop) so as to prevent errors from floating into overt production. Postma (2000:109) estimates that "[t]he articulator will not have executed the speech plan until after 200–250 ms [which] leaves about 100 ms for detection and repair before overt realisation", and claims that speakers are also likely to be able to intercept errors while constructing phonemic representations. However, even when the error slips past the speaker's lips, it may be detected and corrected by the so-called external loop, or to use Postma's term (2000), the 'auditory loop'.

---

[3] Lexotactics is "the collection of ordering patterns specifying how lexemes [structural units at the lexical strata] may combine in a language" (Garcia et al. 2017:211).
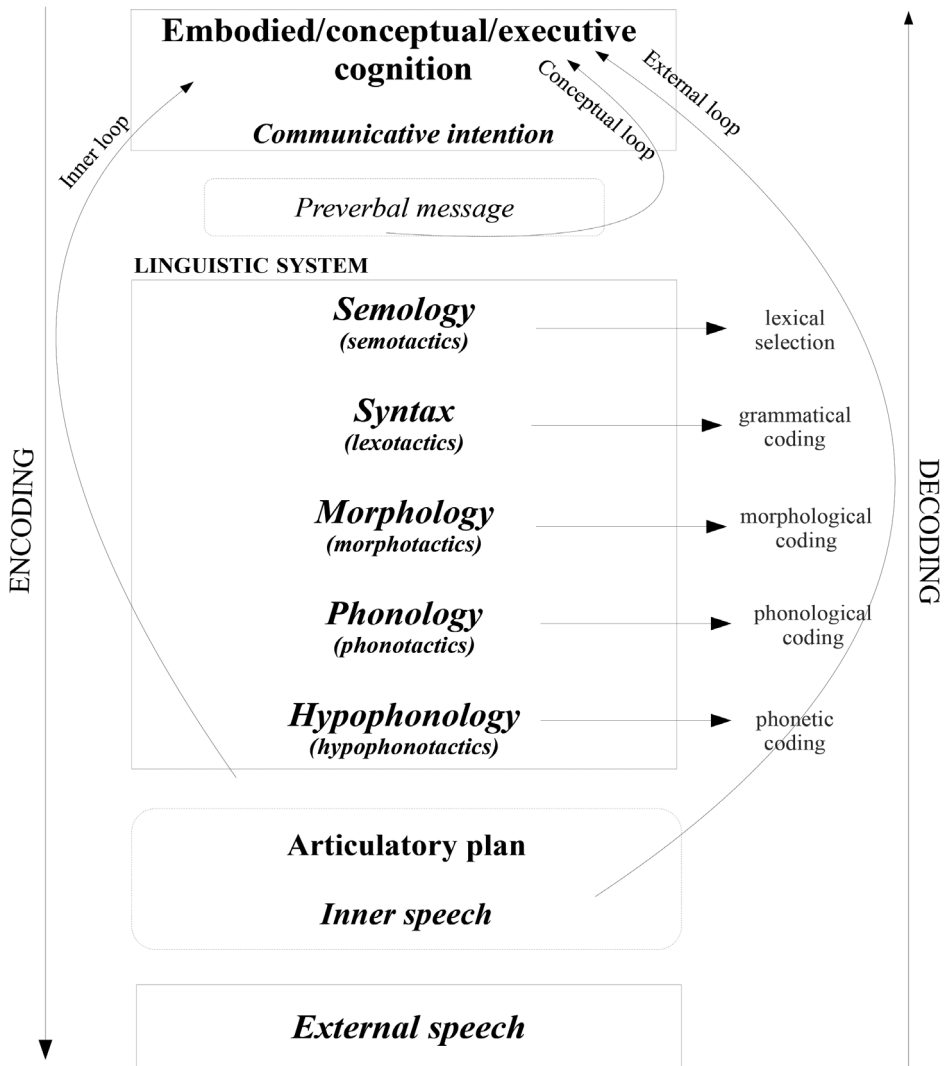
Figure 1. Linguistic processing scheme based on Levelt's speech generation model (1989), and Sullivan and Tsiang's description of the linguistic system (2011)

In addition to the speech monitoring systems, Postma (2000) distinguishes three types of monitors which are to detect errors during the speech planning process. Thus, he contends that there are lexicality and syntax monitors which are to form a proper lexico-syntactic structure, and node activation monitors which operate at all linguistic levels and are to prevent erroneous elements from being activated. The function of the aforementioned activation monitors is performed by stratificational tactic patterns which guide and constrain the arrangement of elements

at all linguistic levels. As they are considered obsolete, those additional monitors are not included in Figure 1.

5. The role of inner speech in the production of overt speech

As scientific research reveals, inner speech is extremely important for proper speech generation. First and foremost, it provides an articulatory plan for external speech and pre-stores linguistic information in verbal working memory (Baars/ Gage 2010). This assumption is in line with Sokolov (1972) and Vygotsky (1986) – the most prominent inner speech researchers – who state that covert speech constitutes a mental draft for overt production. The importance of the pre-planning stages in external speech generation has been highlighted by such researchers as Luria (1947:77 in Akhutina 2003:53) who views inner speech as "an abbreviated verbal schema that precedes the speech act and provides the basis for its subsequent expansion", and Leontiev (1969 in Akhutina 2003), who emphasises the importance of inner programming in the construction of the internal representation of an utterance. Similarly, Dell (1986) and Levelt (1983, 1991) stress that internal representations are built up in advance of overt production and that inner speech plays an extremely important role in monitoring the articulatory plan of an utterance. Levelt (1993) even calls inner speech an 'articulatory buffer', as it temporarily retains the phonological representation of a linguistic message until it is executed by the articulators or parsed into meaningful units during the decoding process. Inner speech thus appears to support working memory, as it enables maintaining linguistic information until the production or comprehension process takes place.

According to Baddeley (2000), working memory has a four-component structure; that is, it consists of the central executive, the phonological loop, the visual sketchpad, and the episodic buffer. The most important component is the 'central executive' which is "a superordinate cognitive control mechanism" (Buchsbaum 2013:2). There are also two slave subsystems: the 'phonological loop' and the 'visual sketchpad', which are responsible for holding verbal/acoustic and visuospatial information, respectively (Henry 2011). The phonological loop is further subdivided into two subcomponents: the phonological store and the articulatory rehearsal mechanism. Considered to be a passive component, the phonological store merely maintains information for a limited period of time, whereas the articulatory rehearsal mechanism recites the information stored within the phonological store to prevent its rapid decay (Henry 2011). Buchsbaum (2013:3) points out that inner speech performs the function of the phonological loop, as it involves both an 'inner voice' and an 'inner ear', adding that "it is not difficult to identify a resemblance between these two phenomena and the functional components of the phonological loop". The last and the most recently-added component of working memory is called the 'episodic buffer', which is also controlled by the central executive (Bad-

deley 2000). As a multimodal temporary store (Henry 2011:31), the episodic buffer integrates information from various modalities, working memory components, and long-term memory into 'a unitary episodic representation' (Baddeley 2000:417).

Significant for subvocal rehearsal, inner speech appears also to be crucial for internal self-monitoring. Levelt (1991, 1993) claims that there are two planes of speech monitoring: (1) the internal one, supported by inner speech which is used to recite the verbal information within the phonological loop, and (2) the external one, which checks the speech output for errors which go past the internal monitoring system and have been voiced by the speaker. Quite often, however, external speech output turns out to be inconsistent with the speaker's intention. It is then that we become aware of the inner workings of our brain and that we realize that alongside speech articulation, there is some verbal thinking going on which due to some attentional lapse may slip into the overt speech production, thus leading to the production of slips of the tongue or, more specifically, parapraxes.

6. Inner speech in slips of the tongue phenomena

Considered to be a window into backstage cognition, language, and especially speech errors, betray salient aspects of the speech generation process. As tangible outcomes of internal operations, verbal blunders enable us to achieve an increased understanding of the workings of the human mind. Slips of the tongue (hereafter SOTs) are particularly interesting phenomena, as they are treated as "unintended, nonhabitual *deviation[s] from a speech plan* [emphasis added]" (Dell 1986:284). Similarly, Jeri Jaeger (2005:2) notices that a tongue slip is "*a one-time error in speech production planning* [emphasis added]; that is, the speaker intends to utter a particular word, phrase or sentence, and during the planning process something goes wrong, so that the production is at odds with the plan". SOTs should thus not be regarded as mere mispronunciations, or indicators of the speaker's lack of knowledge (Söderpalm Talo 1980, Jaeger 2005), but rather as "slips of the mind, *betraying aspects of the mental control of speaking* [emphasis added]" (Nooteboom 1980:87). Dell and Oppenheim (2015) even notice that inner speech is not only generated in a similar way to external speech but is also prone to linguistic slip-ups. However, owing to methodological difficulties in measuring internal operations, those slips of the mind, as Dell (1978) calls them, are impossible to investigate. That is why only audible errors produced in external speech are taken into consideration and the role of inner speech in their generation is discussed.

The reasons for the occurrence of verbal stumbles are internally-driven but may also be externally determined. They are individual-dependant, as the speech production process is highly affected by the cognitive and psychological state of the speaker or his/her level of intoxication. Thus, mood, emotional involvement, communication barriers, anxiety, exhaustion, stress, or fast rate of speech, as well

as deficiencies in divided attention or attentional lapses may influence and distort significantly the speech generation process. As underlined by Noteboom (2010), the speaker's attention fluctuates and is directed more at inner or overt speech and when inner operations receive more attention, the speaker 'loses' control of the speech generation process, and a speech error may occur. From a linguistic point of view, Sullivan and Tsiang (2011:360) differentiate so-called 'timing errors' and claim that they arise as a result of the lack of coordination between rest periods of different linguistic strata, adding that "[…] each tactic pattern takes random rest periods during processing, while other tactic patterns continue processing or take overlapping rests. These rest periods, being random, are not coordinated. It may happen that uncoordinated rest periods in two adjacent strata cause processing to get out of synch. In such a case, we get a timing error".

No matter which stance we assume, the SOT must arise at some stage of the speech generation process. Even if the reason for the occurrence of verbal slip-ups is the lack of coordination between interstratal rest periods, it must be the state of the speaker that influences the speech production process. What is more, it is inner speech that is partially responsible for the generation of linguistic stumbles. As mentioned earlier, covert speech plays a significant role in the speech monitoring system. Apart from buffering the articulatory plan until it is checked by the internal speech monitoring system, it also checks the internal representation of the message for errors. Hence, when the error slips past the speaker's lips, it is due to the malfunctioning of the inner speech monitoring system. Interestingly, Noteboom (2010) notices that approximately 50% of the errors that surface in overt speech are corrected by speakers themselves, and all early self-interruptions and self-repairs of SOTs occur as a result of self-monitoring of covert, rather than external, speech (Noteboom 2005). It is highly likely that quite often a belated command is sent by the inner speech monitoring system, and the speaker stops the articulation of a word just after the initial sounds are voiced. In turn, late interruptions result from the operations of the overt speech monitoring system, which detects the error once it is completed.

Noteworthy as it is, errors hardly ever violate the tactic patterns of a given language. They are "practically always a phonetically possible noise" (Wells 1951:86) or semantically- or phonetically-related words, and hardly ever pseudo-words or non-words (Noteboom 2004). Typically spoken with perfect fluency (Erard 2008), verbal blunders are presumed to arise at some point of the speech generation process. Due to the influence of internal or external factors affecting the speaker, an erroneous linguistic element is activated and therefore an inappropriate part of the linguistic network is triggered. Subsequent operations abide, however, by the rules peculiar to a given linguistic layer, and therefore verbal blunders „fit the regularities of language and usage as much as possible" (Dell 1986, Baars 1992:10), and hardly ever violate phonotactic or syntactic constraints of the language.

The influence of inner speech upon the speech production process is very noticeable in Freudian slips, which arise as a result of the integration of intruding background thoughts into the overt speech production plan. Inner verbal thinking appears to be tagging along the generation of external speech and shards of inner languaging may seep into the overt speech plan.

As an example, consider the following slips of the tongue:

(1) A waiter in a restaurant wanted to go for a smoke, but first, she had to serve a table. She went up to the table, and instead of asking *What would you like to order?*, she uttered *What would you like to smoke?*[4]

(2) A person, extremely terrified of sharks, after a lot of teasing, was finally coaxed into having a swim in the ocean. One of her friends, who knew about her fears played some rude joke on her, which made her pretend to be offended. In lieu of *I'm shocked*, she responded *I'm sharked*.[5]

(3) A mother and a child were building a Lego house. The mother was looking for some windows to fit into the house when the child hid something behind his backs and screamed *Zamknij okna, tzn. oczy (Close your windows, I mean eyes)*.

Undeniably, the examples above disclose the speaker's inner thoughts. However, it is essential to know the context of the utterance to be able to discover the reasons for the occurrence of the verbal blunders. The analysis of slips of the tongue phenomena reveals that the speech generation process is often accompanying and influenced by the thinking process. When the attention of the speaker is mitigated and directed at the background line of thought, an inappropriate linguistic element is activated and erroneously slips into the overt speech plan. What is, however, worth noticing is that the error does not violate the tactics of the stratum at which it occurs. The aforementioned examples are lexical substitutions in which the source of the error belongs to the same word class category as the target element. The error must have arisen during the process of lemma retrieval, which was disrupted by internal mental operations. Noteworthy as it is, the erroneous lexeme accommodates to the syntactic environment of the utterance, which is especially noticeable in the second example. The error results from the fusion of the lemma *shark* with the syntax of the target word *shock*, and in consequence, a new pseudoword *sharked* was created. The target inflectional ending must have been added at the morphological strata because, as Lamb (2004:377) emphasises, "the morphotactics of a language provides a mechanism for forming new words and phrases, [whereas] the lexotactics guides the speakers of a language in constructing clauses and sentences which they have never used before".

---

[4] Cf. http://hotword.dictionary.com/freudian-slip/.
[5] Cf. http://hotword.dictionary.com/freudian-slip/.

The analysis of slips of the tongue phenomena shows that the speech generation process revolves around a frame-and-slot mechanism. Levelt (1991:321) notices that "[s]peech errors provide ample evidence for the independent availability of word sketches or frames and of the elements that are to fill them. […]. There are frames with positions for morphemes, phonons, or other elements; during speech the frames are filled with candidate elements". Linguistic processing necessitates the activation of appropriate frames within a generalised tactic pattern. The intrastratal frames contain slots that are structured in accordance with the inherent tactics and filled in with elements intrinsic to a given stratum. More often than not, however, competing fillers may be rendered active and the one on which the speaker's attention is focused wins its way to overt production.

Figure 2 shows the main stages of sentence generation which reflect the slot-binding process. Generally speaking, there seem to be two internal phases of overt speech plan creation. Once the communicative intention arises, appropriate lemmas that best reflect the meaning of the concepts to be expressed are retrieved, together with their syntactic and morpho-phonological properties. Additionally, via spreading activation, interrelated linguistic elements are triggered at each linguistic level. Subsequently, relevant frame structures are evoked and filled in with appropriate linguistic units. It is at the second stage, which involves the extraction of background frames and the creation of the surface structure that slips of the tongue phenomena typically occur as a result of the activation of erroneous elements.

Inner speech appears to play a threefold role in overt speech production. Apart from providing an internal draft for overt speech production, it supports the internal monitoring system and working memory by storing temporarily information to be processed. Therefore, any impairment in its functioning may lead to the divergence from the original speech plan and the occurrence of various types of speech errors. Substitutions or Freudian slips prove that speech errors arise at some point of linguistic encoding. Additionally, such errors as preserverations, anticipations, or transpositions show that part of the internal speech plan is buffered within working memory before articulation. Styles (2005:154) claims that approximately 7±2 elements can be held in short-term memory, whereas Buchsbaum (2013:1) notices that only 3 or 4 items may be retained in working memory at a time. What is more, an item is stored for 1-2s in the phonological store and then it is refreshed for another 1-2s by a secondary rehearsal process before it finally is articulated or fades away. The overt speech plan may thus be violated when phonologically-, semantically- or syntagmatically-related items are prestored in the phonological loop at once because, as noticed by Lamb (1999), an element which is retained within the inner speech loop may slip into overt production as a result of some mistiming. It should also be emphasised that the inner speech loop – an internal monitoring system supported by inner speech – performs a twofold function in linguistic processing. Not only does it facilitate the encoding process, but it is also crucial for decoding linguistic input. Therefore, when overburdened with linguistic and

COMMUNICATIVE INTENTION
Preverbal message

LINGUISTIC ENCODING

1st stage
WORD RETRIEVAL

2nd stage
SURFACE STRUCTURE CREATION
Extraction of background frames

Conceptual-semantic processing – appropriate concepts are evoked and their semantic representation is established. Additionally, relevant lexemes reflecting the meanings of a concept are activated.

Syntactic encoding – appropriate lexemes are retrieved from the mental lexicon and fitted into a frame which imposes a linear ordering upon the evoked lexemes. The structure of the frame is constrained by the stratal lexotactics.

Morphemic encoding – lexemes are divided into morphemes which in turn are structured in accordance with morphotactics and subservient morphological frames.

Phonemic encoding – the output of morphological encoding is subseqently subject to on-line syllabification process, which is context-sensitive and transcend lexical and morphemic word boundaries (Levelt 2001). Subseqently, articulatory gestures are activated and speech production follows.

STRATUM **Tactic patterns**	**Elements of** **the stratum**
SEMEMIC **Semotactics**	**sememes**
LEXEMIC **Lexotactics**	**lexemes**
MORPHEMIC **Morphotactics**	**morphemes**
PHONEMIC **Phonotactics**	**phonemes**
HYPOPHONEMIC **Hypophonotactics**	**hypophonemes**

Figure 2. The main stages of speech production

7. Conclusion

Undeniably, inner speech is an extremely important mental operation that significantly affects our cognitive and linguistic performance and plays an important role in linguistic processing. Though its importance has never been undermined, the true nature of inner speech cannot be fully disclosed. Impossible to be studied directly or verified objectively, the inner voice is investigated via theoretical or indirect methods and raises more questions than answers. It goes without saying that advances in brain research have provided substantial insights into the neuroscience of language and event-related brain potential technology has increased the understanding of language processing and cognitive architecture. However, it is still impossible to track the speech generation process and probe into the neural circuits of the brain. Inner speech and its influence upon overt production remain shrouded in mystery, but given the rapid advancement of neuroimaging methodology, the unanswered questions concerning the inner voice will hopefully soon be attended to and educated guesses confronted with empirical findings from neuroscience studies.

References

AKHUTINA Tatiana V., 2003, The role of inner speech in the construction of an utterance, in: Journal of Russian and East European Psychology 41 (3/4), pp. 49–74.

ALDERSON-DAY Ben / FERNYHOUGH Charles, 2015, Inner speech: Development, cognitive functions, phenomenology, and neurobiology, in: Psychological Bulletin 141, no. 5, pp. 931–965, http://dx.doi.org/10.1037/bul0000021.

BAARS Bernard J. (ed.), 1992, Experimental Slips and Human Error: Exploring the Architecture of Volition, New York.

BAARS Bernard J. / GAGE Nicole M., 2010, Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience, Boston.

BADDELEY Alan, 2000, The episodic buffer: a new component of working memory?, in: Trends in Cognitive Sciences 5 (11), pp. 417–423.

BUCHSBAUM Bradley R., 2013, The role of consciousness in the phonological loop: hidden in plain sight, in: Frontiers in Psychology 4, article 496, doi: 10.3389/fpsyg.2013.00496.

BUSSMAN Hadumon, 2006, Routledge Dictionary of Language and Linguistics, London.

COLTHEART Veronika, 1999, Phonological codes in reading comprehension, short-term memory and memory rapid visual sequences, in: Coltheart V. (ed.), Fleeting Memories: Cognition of Brief Visual Stimuli, Cambridge, MA, pp. 181–225.

CROTT Karen / RASTLE Kathleen, 2004, Is there a syllabary containing stored articulatory plans for speech production in English?, in: Cassidy S./Cox F./Mannell R./Palethorpe S. (eds.), Proceedings of the 10th Australian International Conference on Speech Science and Technology, Australian Speech Science and Technology Association, Canberra, pp. 376–381.

CRYSTAL David (ed.), 2008, A Dictionary of Linguistics and Phonetics, Oxford/Malden.

DE GUERRO Maria C.M., 2005, Inner speech – L2: Thinking in a second language, New York.

Dell Gary S., 1978, Slips of the mind, in: Paradis M. (ed.), The Fourth Lacus Forum 1977, Columbia, SC, pp. 69–74.

Dell Gary S., 1986, A spreading-activation theory of retrieval in sentence production, in: Psychological Review 86 (93), pp. 283–321.

Dell Gary S. / Oppenheim Gary M., 2015, Insights for speech production planning from errors in inner speech, in: Redford M. (ed.), Handbook of Speech Production, Oxford, pp. 404–418.

Erard Michael, 2008, Slips, Stumbles, and Verbal Blunders, and What They Mean, New York.

Evans Vyvyan (ed.), 2007, A Glossary of Cognitive Linguistics, Edinburgh.

Garcia Adolfo M. / Sullivan William J. / Tsiang Sarah, 2017, Relational Network Theory, Bristol.

Geva Sharon / Jones P. Simon / Crinion Jenny T. / Price Cathy J. / Baron Jean-Claude / Warburton Elizabeth A., 2011, The neural correlates of inner speech defined by voxel-based lesion-symptom mapping, in: Brain 134, pp. 3071–3082, doi: 10.1093/brain/awr232.

Grabias Stanisław, 1997, Mowa i jej zaburzenia, in: Audiofonologia X, pp. 9–36.

Grodziński Eugeniusz, 1976, Mowa wewnętrzna – Szkic filozoficzno-psychologiczny, Wrocław.

Henry Lucy, 2011, The Development of Working Memory in Children, London.

Huang Jie / Thomas H. Carr / Yue Cao, 2001, Comparing cortical activations for silent and overt speech using event-related fMRI, in: Human Brain Mapping 15, pp. 39–53.

Jaeger Jeri J., 2005, Kids' Slips: What Young Children's Slips of the Tongue Reveal About Language Development, New Jersey.

Johnson John R., 1984, The role of inner speech in human communication, in: Communication education 33, pp. 211–222.

Johnson John R., 1994, Intrapersonal spoken language: An attribute of extrapersonal competency, in: Vocate D.R. (ed.), Intrapersonal Communication: Different Voices, Different Minds, Hillsdale, NJ, pp. 169–192.

Korba Rodney J., 1990, The rate of inner speech, in: Perceptual and Motor Skills 71, pp. 1043–1052.

Lamb Sydney, 1966, Outline of Stratificational Grammar, Washington, D.C.

Lamb Sydney, 1999, Pathways of the Brain: The Neurocognitive Basis of Language, Philadelphia.

Lamb Sydney, 2004, Selected Writings of Sydney Lamb: Language and Reality, London.

Leontiev Alexei A., 1969, Psycholinguistic Units and the Production of Verbal Utterances, Moscow.

Levelt Willem J.M., 1983, Monitoring and self-repair in speech, in: Cognition 14, pp. 41–104.

Levelt Willem J.M., 1989, Speaking: From Intention to Articulation, Cambridge, MA/London.

Levelt Willem J.M., 1991, Speaking: From Intention to Articulation, Cambridge.

Levelt Willem J.M., 1993, The architecture of normal spoken language use, in: Blanken G./Dittman J./Grimm H./Marshall J.C./Wallesh C.W. (eds.), Linguistic Disorders and Pathologies, pp. 1–15.

Levelt Willem J.M., 2001, Spoken word production: A theory of lexical access, in: PNAS 98 (23), pp. 13464–13471.

Levelt Willem J.M. / Roelofs Ardi / Meyer Antje S., 1999, A theory of lexical access in speech production, in: Behavioral and Brain Sciences 22, pp. 1–75.

Lockwood David G., 1972, Introduction to Stratificational Linguistics, New York.

Luria Alexander R., 1947, Traumatic aphasia: Its syndromes, psychology and treatment, Moscow.

Malmkjær Kirsten (ed.), 2002, The Linguistics Encyclopedia, London.

Marvel Cherie L. / Desmond John E., 2012, From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech, in: Brain & Langauge 120, pp. 42–51.

Matthews Peter H. (ed.), 2014, The Concise Oxford Dictionary of Linguistics, 2nd ed., Oxford.

Mazur Roman / Klimarczyk Maciej / Rudy Jacek / Nyka Walenty, 2006, Wielopiętrowość zaburzeń mowy w praktyce lekarskiej, in: Psychiatria 3 (3), pp. 112–117.

Michalik Mirosław / Siudak Anna (eds.), 2010, Zagadnienia mowy i myślenia, Kraków.

MORIN Alain, 2009, Inner speech and consciousness, in: Banks W.P. (ed.), Encyclopedia of Consciousness, vol. 1, Oxford, pp. 389–402.

MORIN Alain, 2012, Inner speech, in: Hirstein W. (ed.), Encyclopedia of Human Behaviour, San Diego, CA, pp. 436–443.

NOOTEBOOM Sieb G., 1980, Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech, in: Fromkin V. A. (ed.), Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand, London, pp. 87–96.

NOOTEBOOM Sieb G., 2004, Self-monitoring of inner speech and overt speech, in: Conference Proceedings. From sound to sense: 50+ Years of Discoveries in Speech Communication, Massachusetts Institute of Technology, pp. 187–190.

NOOTEBOOM Sieb G., 2005, Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech, in: Speech Communication 47, pp. 43–58.

NOTEBOOM Sieb G., 2010, Monitoring for speech errors has different functions in inner and overt speech, in: Everaert M./Lentz T./De Mulder H./Nilsen Ø./Zondervan A. (eds.), The Linguistic Enterprise: From Knowledge of Language to Knowledge in Linguistics, Amsterdam, pp. 213–234.

OPPENHEIM Gary M., 2001, The little voice in your head: Error-based investigations of abstracted and articulated inner speech, master's thesis, https://oppenheim-lab.bangor.ac.uk/pubs/OppenheimGM_2009_mastersThesis_innerSpeech.pdf.

POSTMA Albert, 2000, Detection of errors during speech production: a review of speech monitoring models, in: Cognition 77, pp. 97–131.

REN Xuezhu / WANG Tengfei / JARROLD Christopher, 2016, Individual differences in frequency of inner speech: Differential relations with cognitive and non-cognitive factors, in: Frontiers in Psychology 7, article 1675, pp. 1–12. doi: 10.3389/fpsyg.2016.01675.

SCOTT Mark / YEUNG Henny H. / GICK Brryan / WERKER Janet F., 2013, Inner speech captures the perception of external speech, in: Journal of Acoustical Society of America 133(4), pp. EL 286–EL 292.

SÖDERPALM TALO Ewa, 1980, Slips of the tongue in normal and pathological speech, in: Fromkin V.A. (ed.), Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand, pp. 81–86.

SOKOLOV Aleksander N., 1972, Inner Speech and Thought, New York.

STYLES Elizabeth A., 2005, Attention, Perception and Memory: An Integrated Introduction, East Sussex.

SULLIVAN William J., 2011, Input, output, and (de)linearization: What we owe to Sydney Lamb, in: LACUS Forum 36, pp. 279–289.

SULLIVAN William J. / TSIANG Sarah, 2011, Speech errors and the ontological status of the morpheme, in: Kwartalnik Neofilologiczny LVIII (3), pp. 357–373.

VYGOTSKY Lev S., 1986, Thought and Language, Cambridge.

WELLS Rulon, 1951, Predicting slips of the tongue, in: Fromkin V.A. (ed.), Speech Errors as Linguistic Evidence, pp. 82–87.

WILEY Norbert, 2006, Inner speech as a language: A Saussuerean inquiry, in: Journal for the Theory of Social Behaviour 36.3, pp. 319–341.

The role of inner speech in the speech production process

The article summarizes the current state of understanding of the concept of inner speech and evaluates the role of the internal language in the speech generation process. First, the available definitions of inner speech are presented and its features are briefly characterised. Subsequently, the in-

ner voice is compared to overt speech and the main differences between those two planes of speech: the internal and the external one are outlined. Since the aim of the paper is to show the role of inner speech in overt speech production, a speech generation model which coalesces Levelt's (1993) assumptions with the stratificational approach to language is presented. Different stages of linguistic processing are described and the impact of internal languaging on linguistic output is discussed. It is claimed that inner speech plays a threefold role in overt speech production: (1) provides an internal draft for external speech, (2) is vital for the self-monitoring system, and (3) supports working memory. Any impairment in the functioning of inner speech may thus lead to speech errors and slips of the tongue phenomena.

Keywords: inner speech, external speech, speech generation, slips of the tongue, stratificational model of language.