Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, 63, 317-351. doi:10.1016/j.cortex.2014.08.014

Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words

Dror Dotan[1,2] and Naama Friedmann[1 ∞]

[1] Language and Brain Lab, School of Education and Sagol School of Neuroscience, Tel Aviv University
[2] INSERM, U992, Cognitive Neuroimaging Unit

The Stimulus Type Effect on Phonological and Semantic errors (STEPS) describes the phenomenon in which a person, following brain damage, produces words with phonological errors (fine→fige), but number words with semantic errors (five→eight). To track the origins of this phenomenon and find out whether it is limited to numbers, we assessed the speech production of six individuals with conduction aphasia following a damage in the left hemisphere, who made phonological errors in words. STEPS was found in all six participants, and was not limited to number words – several other word categories were also produced with semantic rather than phonological errors: function words, English letter names, and morphological affixes were substituted with other words within their category. This supports the building blocks hypothesis: when phonological sequences serve as building blocks in a productive process, they end up having pre-assembled phonological representations, ready for articulation. STEPS reflects a deficit that causes substitutions of one phonological unit with another. In the case of plain content words, this causes substitutions of one phoneme with another, but in the case of pre-assembled phonological unit, this causes substitutions of number words with other number words, function words with function words, and morphological affixes with other affixes. An analysis of the participants' functional locus of deficit revealed that they all had a deficit in the phonological output buffer, and this was their only common deficit. We therefore concluded that the pre-assembled phonological units are stored in dedicated mini-stores in the phonological output buffer, which processes not only phonemes but also whole number words, function words, and morphemes. We also found that STEPS depends on the word's role: number words were produced with semantic errors only when they appeared in numeric context, and function words triggered semantic errors only in grammatical context. This suggests that the phonological representation of a word can be obtained either from the phonological output lexicon or from a store of pre-assembled representations in the phonological output buffer, depending on the word's role.

∞ We are grateful to the participants for their patience and cooperation during the many sessions of testing, to Aviah Gvion, Maya Faran, Michal Biran, and Nurit Borer-Alafi for their help in finding participants for this study and for discussions of the findings and their implications, to Adi Lifshitz Ben Basat for her contribution to encoding the results, and to Claudio Luzzatti for reviewing the CT images. We are also grateful to Simon Fischer-Baum for his excellent comments on our manuscript. This research was supported by the Israel Science Foundation (grant no. 1066/14, Friedmann), by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021), and by a grant from the Bettencourt-Schueller Foundation. Dror Dotan is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

1 Introduction

SZ was admitted to hospital following a left fronto-parietal infarct that damaged his speech. A series of examinations revealed that he had conduction aphasia (a phonological buffer deficit, Gvion & Friedmann, 2012a). More than half of the words he said were produced with phonological errors (such as *kangaroo* → *kanbaroo*, or *bell* → *cell*, but SZ was tested in Hebrew, his native language). He showed a similar pattern in nonwords, which he produced with phonological errors (e.g., *kizuma* → *dizuma*). Quite surprisingly, when SZ said numbers, he managed to say them without phonological errors. However, he often did not produce the numbers he intended to say, but different numbers. His errors in numbers were semantic – substitutions of a number word with another number word (e.g., *forty-two* → *forty-five*), and syntactic errors – changing the syntactic structure of the number (e.g., *thirteen* → *one hundred and three*; note that the terms syntactic and semantic errors have slightly different meanings when talking about numbers than when talking about speech in general). In the current study we explore how systematic the difference in error pattern is in SZ's production, and how systematic it is for other individuals with aphasia who are impaired in the same functional locus as SZ. A systematic difference in error patterns between words and numbers would indicate that number words and non-number words are processed in different ways. We then explore what gives rise to this pattern, by identifying the participants' locus of functional deficit.

The pattern of SZ's performance is characterized by two phenomena: one phenomenon is the occurrence of phonological errors in non-number words and nonwords, but not in number words. We called this phenomenon *the stimulus type effect on phonological errors*. The other phenomenon is the occurrence of semantic errors in number words but not in non-number words (*the stimulus type effect on semantic errors*). When the two phenomena co-exist in the same patient, like in SZ's case, we call them *STEPS – the Stimulus Type Effect on Phonological and Semantic errors*.

The present study investigated STEPS in detail. We aimed to identify the locus of the cognitive deficit in the lexical retrieval process that underlies STEPS, to discover whether the phenomenon is limited to number words or is a more general phenomenon that applies to other kinds of words, and eventually – to offer a theoretical framework that can account for STEPS and its properties as reflected in the results of the current study and of findings from previous studies of STEPS.

1.1 Previous cases of STEPS

SZ is not the first reported case of STEPS. The phenomenon was first investigated by Cohen, Verstichel, and Dehaene (1997). They reported a French teacher who had neologistic jargon following a left temporal infarct, with phonologically related errors in non-number words. This patient produced 98% of the target number words without phonological errors and with semantic errors. This is the only case we are aware of in which both phenomena – phonological errors that appear selectively in non-number words, and semantic errors that appear selectively in numbers – co-exist in a single person in a clear manner. Messina, Denes, and Basso (2009) analyzed an impressively large group of 57 aphasic patients and found the STEPS phenomenon on the group level, namely, the group had mostly semantic errors in number reading (20% semantic vs. 2% phonological errors) but mostly phonological errors in reading words (10% phonological vs. less than 1% verbal paraphasias) and nonwords (25% phonological errors). Essentially similar results were found in repetition tasks.

There are several single-case studies in which one of the two phenomena (phonological errors only in non-number words or semantic errors in number words) was found whereas the complementary phenomenon was not reported, or was less clear-cut. Table 1 lists these cases.

Girelli and Delazer (1999) described BP and GS, two patients who exhibited a STEPS phenomenon, although only in some tasks. In word production, GS made phonological and semantic errors, and BP made both neologisms and paraphasias the type of which was not specified. Both patients had semantic errors in numbers, and neither had phonological errors in most of the number production tasks, with the exception of neologisms when reading visually-presented number words.

Several studies reported phonological errors that appeared in non-number words but not in numbers. Bencini et al. (2011) and Semenza et al. (2007) investigated GBC, a man with Wernicke’s aphasia who produced numbers flawlessly but made many phonological errors in word production (only in vowels). Patient LT (Shallice, Rumiati, & Zadini, 2000) had an impaired phonological output buffer and made many phonological errors in word production. His phonological error rate in single word repetition was 30%-50%, yet in a digit span task his error rate in single digits was smaller than 20% (the error type was not reported)[1]. Regarding the semantic errors in number production, LT’s error rate when producing digit names in the digit span task was low, yet it is possible that the semantic error rate would have been higher had he been asked to produce multi-digit numbers. Such a difference between single digits and multi-digit numbers was found in other studies reported here, and, as we will see later, also in the present study (see Section 4.1.1.3). Another patient who showed the stimulus type effect on phonological errors is DPI (Bachoud-Lévi & Dupoux, 2003), who had phonological errors in word production, yet his number production was spared, with neither phonological nor semantic errors. We do not know, however, whether he was requested to produce multi-digit numbers or only single digits, so again it is possible that semantic errors would have appeared in multi-digit numbers. Another patient who exhibited this error pattern was FS (Delazer & Bartha, 2001), who produced content words with phonological errors but made only 7% errors in reading two-digit Arabic numbers. Finally, patient TM made phonological errors in reading and repeating words, but could read one- to four-digit numbers with almost no phonological or semantic errors (Lochy, Domahs, Bartha, & Delazer, 2004).

[1] In the digit span task, LT correctly repeated 80% of the 20 four-digit sequences presented to him. The authors said that only digits that were "accurately produced" were counted as correct. Thus, LT produced at least 64/80 digits with no phonological errors. It would be reasonable to assume that in several cases LT erred only in some of the digits in the four-digit sequence, and that some of his errors were not phonological but semantic or digit omissions, so his phonological error rate in single digits was probably lower than 20%.

Table 1. Previous reports of the Stimulus Type Effect on Phonological or Semantic errors

Patient	Source	Language	Locus of Deficit	Errors in words	Errors in single digits	Errors in multi-digit numbers	Comments	STEP*	STES*
DPI	Bachoud-Lévi & Dupoux, 2003	French	Phonological output lexicon	Phonological	No errors	?	Color naming: no phonological errors	Yes	?
GBC	Bencini et al., 2011	Italian		Phonological, only in vowels	?	No errors		Yes	No
	Cohen et al., 1997	French		Phonological (neologisms)	Semantic	Semantic (lexical and syntactic)	Good naming of French letters	Yes	Yes
FS	Delazer &Bartha, 2001			Phonological, approximations, and “word finding difficulties”	?	Few errors of unspecified type in two-digit numbers	Conduction aphasia	Yes	?
BP	Girelli & Delazer, 1999; Delazer & Bartha, 2001			Neologisms, paraphasias of unspecified type	Semantic	*Reading and calculation:* semantic *Reading number words:* neologisms		Almost	Yes
GS	Girelli & Delazer, 1999; Delazer & Bartha, 2001			Phonological & semantic	Semantic	*Reading and calculation:* Semantic *Reading number words:* neologisms		Almost	No
TM	Lochy et al., 2004	German		Phonological	?	*Repetition:* errors of unspecified type *Reading:* OK	Good color naming	Yes	?
FA	Marangolo et al., 2004	Italian		*Most words:* OK, few phonological *Nonword /long words:* phonological	No errors	Semantic (lexical)		?	Yes
RA	Marangolo et al., 2005	Italian		OK, with unspecified errors in verbs, func. words, and nonwords	Semantic	Semantic (lexical)	Agrammatism	?	Yes
HK	Rodriguez & Laganaro, 2008			Phonological	?	?	Morphological (no phonological) errors when saying country names	For country names	No
LT	Shallice et al., 2000	Italian	Phonological output buffer	Phonological	Very few errors, type not clear	?		Yes	?

* STEP = Stimulus Type Effect on Phonological errors, namely, phonological errors in words but not in numbers.
STES = Stimulus Type Effect on Semantic errors, namely, semantic errors in numbers (or another category of words) but not in non-number words.

Several studies reported cases with phonological errors that spared word categories other than numbers: letter names (Cohen et al., 1997), color names (Bachoud-Lévi & Dupoux, 2003; Lochy et al., 2004), and country names (Rodriguez & Laganaro, 2008). Bachoud-Lévi and Dupoux's patient also had semantic errors in color names. Although the researchers suggested that the semantic errors may have originated in perceptual difficulties, it could also be possible that he had category-specific semantic errors in colors.

The complementary phenomenon – individuals who made semantic errors in number words but not in non-number words – was reported in two studies. Marangolo, Piras, and Fias (2005) reported patient RA, who made lexical errors (digit substitutions) and omissions in number production, including single digits. His non-number word production was intact, with the exception of verbs, function words, and nonwords, in which he erred (we do not know whether these errors were phonological or not). The second individual who exhibited such error pattern is FA (Marangolo, Nasti, & Zorzi, 2004), who produced semantic errors in multi-digit numbers. His word production was mostly spared, but he did have phonological errors in nonwords, in extremely long words, and in compound words.

In the present study we explore in detail the pattern of production of words, numbers, and other specific categories of words in six individuals with aphasia, aiming at describing in detail their error types and rates in the various word types and the locus of deficit in speech production that gives rise to this pattern of errors.

1.2 Cognitive models of speech production

1.2.1 Naming

Speech production models usually describe several stages in the naming process (e.g., Butterworth, 1979, 1989; Caramazza & Hillis, 1990; Caramazza, 1997; Dell, 1986, 1988; Ellis & Young, 1996; Friedmann, Biran, & Dotan, 2013; Garrett, 1992, 1976; Hillis, 2001; Kempen & Huijbers, 1983; Levelt, 1989, 1992; Nickels, 1997; Patterson & Shewell, 1987). These researchers generally describe a mechanism in which the conceptual representation of the word first activates its lexical semantic representation. The lexical semantic representation (and possibly the lexical syntactic representation, Biran & Friedmann, 2005, 2012; Bock & Levelt, 1994; Caramazza, 1997; Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Roelofs, 1992) activates the phonological form of the word in the phonological output lexicon. This phonological information is then transferred to the phonological output buffer – a working memory component that assembles the phonological representation of a word and stores it until the articulatory system, which produces the sounds of the word, has finished uttering it. The middle column in Figure 1 illustrates this process.

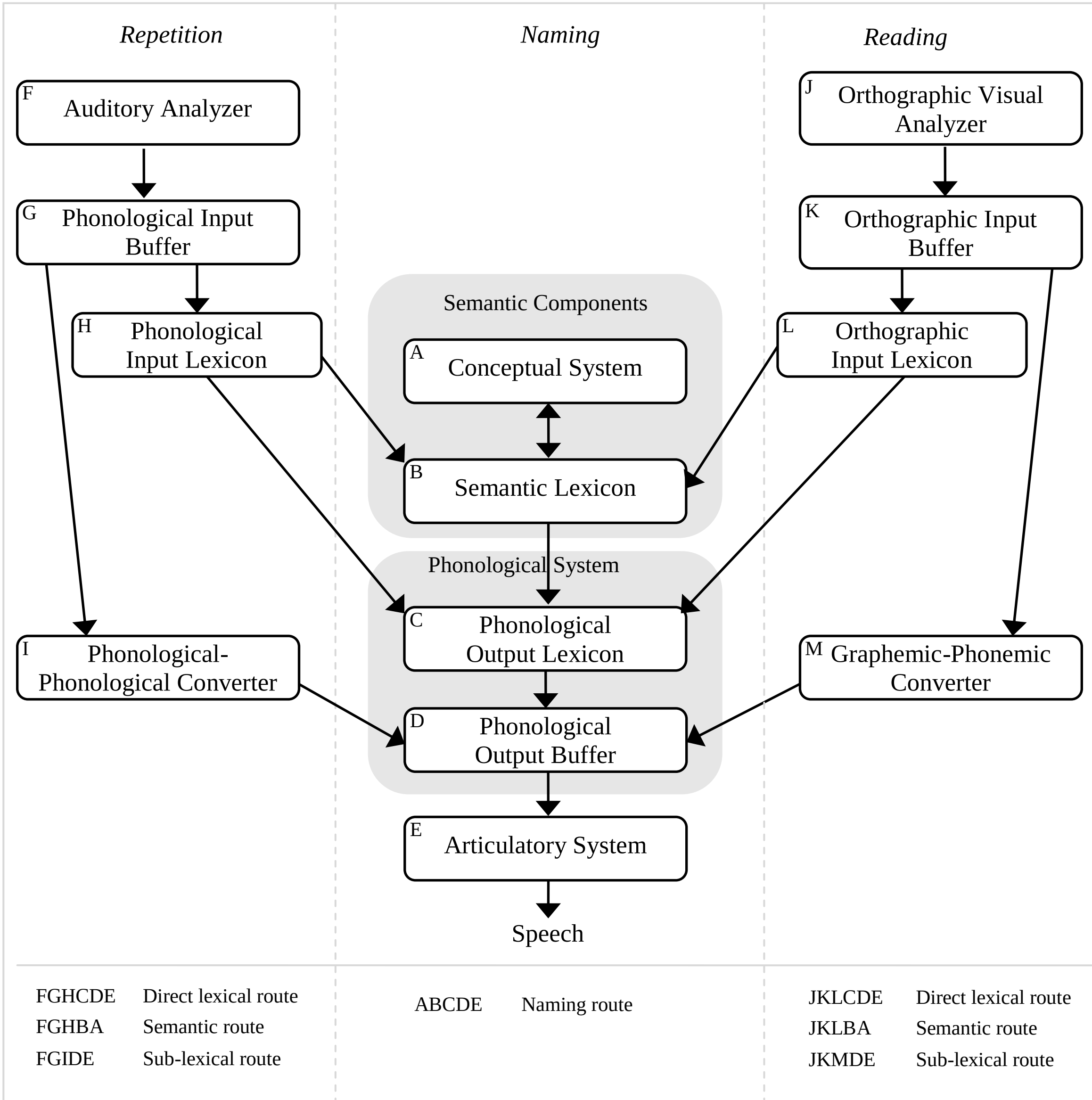


Figure 1. A model for speech production

The phonological output lexicon stores words as disassembled units: for each word, it stores metrical information (number of syllables and stress pattern) and segmental information (the phonemes of the word – consonants and vowels and their relative position) (Butterworth, 1992; Levelt, 1992). These two types of phonological information are merged at the phonological output buffer, where the full phonological form of the word is created and maintained until the word is fully articulated. A possible reason for the lexicon to keep the words disassembled rather than as already-assembled words is that words may have different phonological forms in different contexts – for example, “Come on” may be sometimes pronounced as “Com’on”. If the speech production system is phonologically productive, namely, if the phonological form of the word is constructed each time we use it, it becomes easier for the speech production system to be productive and to generate the word in many different ways, depending on the context (Levelt, 1992). Some researchers (Kohn & Melvold, 2000; Lavro, Ben-Or, & Friedmann, 2006) argue that the phonological parts of the speech production system are also *morphologically* productive: according to this view, words are retrieved from the phonological output lexicon without the morphological affixes (i.e., the lexicon stores only the word's stem). A subsequent stage, implemented perhaps by the phonological output buffer, inflects the word by adding the morphological affix to the stem that was retrieved from the phonological output lexicon.

Each of the components described above is a separate module that can be selectively impaired following a focal brain lesion. Selective deficits in different components would result in different types of errors. Thus, a deficit in the semantic lexicon may cause semantic paraphasias – replacing a word with a semantically related word (e.g., *pineapple* → *coconut*). A deficit in the phonological output lexicon or in the phonological output buffer would result in phonological errors, which may take various forms. They may be formal paraphasias – replacing the word with a phonologically similar word (*bell* → *cell*; we defined "phonologically similar" as containing at least 50% of the target word's phonemes); phonemic paraphasias – a nonword that is phonologically similar to the target word (*computer* → *lomputer*); or errors in the order of insertion of the segments into the metrical structure (*coconut* → *conocut, nococut*) (Biran & Friedmann, 2005). In tip-of-the-tongue situations partial information is often accessible, with the result of approximations – faithful (*pineapple* → *pi… app… pine… pine… pineapple*) and unfaithful (*pineapple* → *pi… mi… pike… pineapple*). A deficit in the phonological lexicon may also lead to semantic errors in addition to phonological errors (Caramazza & Hillis, 1990; Rapp & Goldrick, 2000).

Selective deficits in different speech production modules will be manifested not only in different types of errors but also in different factors affecting performance, because the various modules have sensitivity to different parameters such as the word length or frequency. The different types and effects of errors can be used to assess the functional locus of impairment in cases of brain lesions. They are discussed in more detail in section 3.2, where we assess the locus of deficit for each of the participants in the present study.

1.2.2 Reading aloud

The dual-route model for reading single words (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Coslett, 1991; Ellis & Young, 1988; Forster & Chambers, 1973; Frederiksen & Kroll, 1976; Friedmann & Gvion, 2001; Funnell, 1983; Jackson & Coltheart, 2001; Marshall & Newcombe, 1973; Paap & Noel, 1991; Patterson & Morton, 1985) describes the different components involved in word reading (the right column in Figure 1). The first stage, which is shared by all reading routes, is the orthographic-visual analysis system. This stage is responsible for encoding the abstract letter identities and the relative position of the letters within the word. The information from the orthographic-visual analyzer flows through an orthographic input buffer to a lexical and a sublexical route. In the lexical route, the activation flows to the orthographic input lexicon and then further diverges into two sub-routes. In the direct lexical route, the main route used by skilled readers for oral reading, the orthographic input lexicon directly activates the representation of the word in the phonological output lexicon, and from there the activation flows to the phonological output buffer. The word's representation in the orthographic input lexicon also activates its semantic representation in the semantic lexicon, which in turn activates the entry in the conceptual system. This is the route responsible for written word comprehension. The sub-lexical route bypasses the orthographic input lexicon and the phonological output lexicon, and uses a direct conversion of graphemes in the graphemic input buffer to phonemes in the phonological output buffer by the graphemic-phonemic converter. This sub-lexical route allows reading of unfamiliar words, which are not stored in the lexicons.

1.2.3 Repetition

Word repetition (the left column in Figure 1) is similar in principle to word reading: it begins in analyzing the auditory-phonological input and holding it in a phonological input buffer. The information can then flow through either a lexical or a sub-lexical route. In the lexical route, the word is activated in the phonological input lexicon, and repetition is done via direct activation of the phonological output lexicon. Comprehension is done via the semantic route: the phonological input lexicon activates the semantic lexicon, which in turn activates the conceptual system. Like in reading, there exists a sub-lexical route for repetition. This route does not use the lexicons, but rather forms a direct connection between the phonological input buffer and the phonological output buffer, via the phonological-phonological converter – a component that can "hear" phonemes and produce a copy of them (Goldrick & Rapp, 2007; Nickels, 1997).

1.3 Research goals

The first question we asked in this study was whether individuals with aphasia who make phonological errors in production of content words systematically produce number words with semantic errors. We showed that this is indeed the case. Following this, we investigated the origin of STEPS and showed that the mechanism underlying this phenomenon is the fact that the speech production system handles number words as pre-assembled phonological building blocks.

2 Methods

2.1 General Procedure

Each of the participants performed the tasks in a series of 8-15 sessions. Each session lasted between one and two hours, in a quiet room – usually at the participant's home. In the tasks that required oral response, the participant's response was written down in real time by the experimenter, and was also audio-recorded and later transcribed by two judges. In some tasks the stimuli were presented on a computer screen. In these cases, we used a laptop with a 15" monitor, and the participant clicked the mouse to present each stimulus. In the reading tasks, stimuli were presented in David 40 point font in the computerized tasks, and in Arial 14 point font when shown on paper. In tasks with orally-presented stimuli, the experimenter repeated the stimulus as many times as the participant requested. There was no time limit in any task. The verbal material in all tasks was in Hebrew, except for a few cases in which this is explicitly indicated.

2.2 Statistics

To compare a participant’s success rate between two different tasks or conditions, we used χ^2 or Fisher’s exact test (according to the number of items in each expected cell). Group-level comparisons were done using Wilcoxon's Signed Ranks test, the compared values being the error rates of each participant. All *p* values reported throughout the article are one-tailed.

3 Participants

Six Hebrew-speaking aphasic patients participated in this study. They were 3 men and 3 women, aged 52-84. They all had left hemisphere lesions – five of them had CVA, and one (GE) had a hemorrhage following a brain tumor removal surgery. All had normal or corrected-to-normal vision. The participants’ background information appears in Table **2**, including the lesion location of each participant and the functional locus of deficit according to our diagnosis (which is presented in detail below). Their CT images appear in Appendix A.

SZ, GE, and ZH were native speakers of Hebrew. The other participants lived in Israel and spoke Hebrew for more than fifty years. Prior to the brain damage, all of them could read well. Five of them read Hebrew, but RB read mainly in English, so she performed the word reading tasks in English rather than in Hebrew. ZC, whose reading was severely impaired following her stroke, did not perform the reading tasks at all.

We used several tasks to identify the exact functional locus of deficit of each participant. The analysis showed that all six participants had a deficit in the phonological output buffer. For SZ and GE the deficit was a pure deficit that selectively impaired their phonological output buffer. The four other participants also had a deficit in the phonological output lexicon. All participants had intact input modules.

The remainder of this chapter describes the process we used to select participants for this study and to assess their exact locus of functional deficit.

Table 2. Background information of the participants.

	Age	Sex	Dominant Hand	Years of Educ-ation	Etiology	Lesion Location	Post-onset Time[a]	Phonological Output Buffer	Phonological Output Lexicon	Comments
SZ	69	M	Right	12	CVA	Left fronto-parietal	3 months	Impaired	Intact	
GE	51	F	Left	16	Hemorrhage following brain tumor surgery	Left fronto-parietal	4 years	Impaired	Intact	
YL	84	M	Right	University professor	CVA	Not available	6 months	Impaired	Impaired	Uses a hearing aid
ZH	52	M	Left	16	CVA	Left temporo-parietal	18 months	Impaired	Impaired	Reported attention difficulties before the CVA
RB	78	F	Left	17	CVA	Left temporo-parietal	6 months	Impaired	Impaired	
ZC	81	F	Right	Not available	CVA	Not available	1 year	Impaired	Impaired	Cannot read at all following the CVA

[a] The "post-onset time" column refers to the time that passed from the brain damage and onset of aphasia until we met the participant.

3.1 Inclusion criteria

The first research question in this study was whether aphasic patients who make phonological errors in content words produce number words with semantic errors. To examine this question in an unbiased manner, the inclusion criteria were based only on the existence of phonological errors in the participants' production of content (non-number) words. Their production of number words was assessed only as a second stage.

The inclusion criteria were based on three tasks: picture naming, word repetition, and word reading.

Task 1. Picture naming (SHEMESH test, Biran & Friedmann, 2004): The test includes pictures of 100 objects of various semantic categories, and the participant is requested to say the object name. The target words are one to four syllable long nouns, with ultimate and penultimate stress and with various first phonemes. They include both masculine and feminine nouns, with regular and irregular gender morphology. The frequency of the words, judged by 75 Hebrew-speaking participants with no language deficits, ranges from 2.39 to 6.84 on a scale of 1-7 (M = 4.90, SD = 1.09). The performance of adults without a language deficit in this test is 95.6% (SD = 4.2%; these data are based on 52 control subjects, aged 50-84, Biran & Friedmann, 2004, 2005). RB performed this task in Hebrew and in English.

Task 2. Word repetition: the participants repeated 40 words (verbs/nouns) that were said by the experimenter. The words included 2-12 phonemes (M = 5.7, SD = 2.2). The task included no number words or function words. Ten of the words included an inflectional suffix and the rest were monomorphemic. The average orthographic frequency of these words was 26 in a million (SD = 46) (Linzen, 2009).

Task 3. Word reading: the participants read aloud 100 content words from the dyslexia screening test of the TILTAN battery (Friedmann & Gvion, 2003). The words included 2-11 letters (M = 4.7, SD = 1.4) and 2-12 phonemes (M = 5.8, SD = 1.8). SZ read the words from paper and the other participants saw them on a computer screen, one at a time. These 100 words were presented within a larger list, which included 46 more words: 28 function words from the TILTAN battery, and 18 words that will be described in detail in Section 4.1.3.

RB performed the reading task in English. She read a simple English text printed in Arial 11 font, with large spacing between the lines. The text included 153 words: 77 content words, 63 function words, 11 number words and two acronyms. Only the 77 content words were analyzed. They included 3-13 letters (M = 6.9, SD = 2.4) and 3-11 phonemes (M = 6.2, SD = 2.2).

Table 3 summarizes the rates of the various types of errors in these word production tasks. **Semantic error** was defined as a substitution with a word that is semantically related to the target word (such as *giraffe* → *zebra*). **Phonological error** was defined as a response that is phonologically related to the target word, containing at least 50% of the phonemes of the target word. In phonological errors, the produced word could be a nonword (**phonemic paraphasias**, e.g., *zebra* → *zebla*) or an existing word (**formal paraphasias**, e.g., *dog* → *fog*). A **faithful approximation** is when the participant repeatedly produces only some of the word's phonemes, but does not produce any incorrect phoneme (e.g., *pineapple* → *pi… app… pine… pine… pineapple*). **Mixed error** is the production of a word that is both semantically and phonologically related to the target word (e.g., *cat* → *rat*). **Layered error** is the production of a semantically related word with a phonological error (e.g., *giraffe* → *zebla*). **Unrelated word error** is the production of a word that is neither semantically nor phonologically related to the target word (e.g., *giraffe* → *door*). Last, **neologism** is a nonword that is phonologically unrelated to the target word, i.e., includes fewer than 50% of the target word’s phonemes (e.g., *giraffe* → *klobaf*).

The total number of phonological errors (“all phonological errors” column) includes phonemic and formal paraphasias, layered errors, and faithful approximations. Neologisms were not counted as phonological errors because they may sometimes result from a non-phonological deficit (Butterworth, 1979; Nickels, 1997). Mixed errors were not counted as phonological because they could have a semantic origin (Rapp & Goldrick, 2000).

We included in the study participants who (1) produced phonological errors in at least 15% of the words in each of the three production tasks above; (2) had at least partial access to the phonology of 50% of the words or more in the 3 tasks, pooled together (we considered a word as "partially accessed" if the participant produced at least half of the word's phonemes, even if the word was produced with phonological errors); and (3) most of the participant’s errors in the 3 tasks together were phonological. As Table 3 shows, all six participants meet these criteria.

Table 3. Rates of errors of different types in word production tasks

Task		Word production attempt[a]	All errors[b]	Semantic errors	All semantic errors	Phonemic paraphasias	Formal paraphasias	Faithful approxim-ations	All phonolog-ical errors	Mixed errors	Layered errors	Unrelated word	Neologisms	Word accessed[c]
Picture naming (Task 1)	SZ	98%	62%	5%	6%	35%	4%	33%	55%	1%	2%	1%	3%	94%
	GE	99%	42%	0%	0%	21%	0%	24%	41%	1%	0%	0%	1%	98%
	YL	94%	55%	0%	3%	35%	6%	2%	43%	0%	3%	5%	10%	84%
	ZH	74%	62%	15%	20%	19%	4%	8%	30%	4%	5%	3%	7%	55%
	RB (Heb)	78%	68%	6%	9%	28%	3%	0%	31%	0%	3%	3%	13%	51%
	RB (eng)	90%	58%	8%	10%	27%	8%	0%	36%	1%	3%	3%	7%	71%
	ZC	96%	61%	11%	17%	43%	7%	8%	42%	2%	6%	1%	4%	83%
Word repetition (Task 2)	SZ	100%	23%	0%	0%	23%	0%	5%	23%	3%	0%	0%	0%	100%
	GE	100%	28%	0%	0%	13%	8%	8%	28%	0%	0%	0%	0%	100%
	YL	100%	33%	0%	0%	23%	8%	3%	33%	0%	0%	0%	0%	100%
	ZH	83%	64%	0%	0%	45%	21%	0%	61%	0%	0%	0%	3%	83%
	RB (Heb)	100%	23%	0%	0%	15%	8%	0%	23%	0%	0%	0%	0%	100%
	ZC	100%	48%	3%	3%	38%	5%	0%	43%	0%	0%	0%	5%	95%
Word reading (Task 3)	SZ	100%	70%	0%	0%	37%	35%	18%	70%	0%	0%	0%	0%	100%
	GE	100%	18%	0%	0%	4%	7%	11%	18%	0%	0%	0%	0%	100%
	YL	100%	21%	0%	0%	13%	7%	1%	21%	0%	0%	0%	0%	100%
	ZH	100%	56%	0%	0%	33%	12%	18%	55%	0%	0%	1%	1%	99%
	RB (Eng)	99%	33%	0%	0%	25%	4%	0%	29%	1%	0%	1%	0%	97%

[a] Items to which there was any verbal response (excluding definitions and “don’t know”). Except the “word accessed” column, percentages are specified out of these verbal responses.

[b] The sum of all error types may exceed the total number of errors, because some words were produced with more than one error.

[c] The phonological form of the word was accessed (≥50% phonemes), even if the word was produced with phonological errors.

3.2 Assessment of the participants' locus of deficit

To assess the participants' functional locus of deficit, we first showed that their speech production errors result from a deficit in the phonological stages of speech production and not from a deficit in earlier stages (the input modules, the conceptual system, or the semantic stages). We then assessed the exact locus of the phonological impairment for each participant – whether it is the phonological output lexicon that is impaired, the phonological output buffer, or both.

All tasks in this assessment process included no numbers or function words.

3.2.1 Ruling out an input deficit

To examine whether the participants had deficits in the input modules (auditory, visual, and orthographic input), which could give rise to errors in repetition, naming, and reading, we compared their performance in tasks that involve speech output with their performance in corresponding comprehension tasks that do not require spoken output. If the deficit is in the input stages, we expect the performance in the input-only tasks to be impaired as well.

To assess the orthographic and auditory input modules, we used a word definition task in which the participants read or heard the words. Auditory input was also assessed by a word matching task, in which the participants judged whether two pairs of words had an identical order or not. Picture naming (Task 1) was used to assess the visual input module: a deficit in this module would be reflected in substituting a word with the name of a visually similar object, e.g., saying *sun* instead of *lemon*[2].

Task 4. Word definition. In the written version of this task, the participants silently read 40 words and were asked to define each word or use it in a sentence. In the auditory version, they heard the same 40 words and defined them (three of the participants did both task versions, which were administered several months apart). Each definition was coded by three independent judges as correct (even if mispronounced) or incorrect. Inter-judge agreement was high: disagreements on coding a word as correct or incorrect occurred in only 3% of the cases, and in these cases the majority opinion was used. We excluded responses that did not include enough information to determine whether the participant understood the target word or not. The frequencies and phoneme lengths of words in this list were comparable to those of the words in the word reading and word repetition tasks (in unpaired t-tests, all $p > 0.19$; frequency data from Linzen, 2009).

Task 5. Word matching (FriGvi battery, Friedmann & Gvion, 2002; Gvion & Friedmann, 2012a). Each item in this task consists of two word pairs. There were 8 items: four with same word order (e.g., *mouse flower; mouse flower*) and four with different word order (e.g., *house garden; garden house*). The participant was requested to determine whether the word order was the same in both pairs or not.

---

[2] Note that the picture naming results presented in Table 3 are sufficient to refute the possibility that the only deficit underlying naming errors was in the visual input modules. This is because each of the participants had many phonological errors in the picture naming task, and such errors cannot be explained by impaired visual processing of the pictures.

Table 4 compares the participants' performance in input tasks with their performance in the corresponding input+output tasks, separated per input channel. This comparison clearly shows a difference between the participants’ relatively spared input and their impaired performance in the tasks that require speech production, thereby indicating that the input modules were intact. The visual error rate in picture naming was low (GE: 1%, ZH: 2%, RB: 1% in English and 3% in Hebrew; ZC: 8%, SZ and YL: no errors), and none of the errors could be positively categorized as visual – all of them could also be explained as semantic or phonological substitutions. This shows that the participants’ visual input module was also intact.

Table 4. Error rates in word comprehension tasks vs. tasks that require both comprehension and production

	Orthographic input (written word)		Auditory input (heard word)		
	Input	Input + Output	Input		Input + Output
	Comprehension measured by definition	Reading aloud	Comprehension measured by definition	Word matching	Word repetition
SZ	0%	67%	0%	13%	23%
GE	0%	18%	0%	0%	23%
YL	5%	21%	–	13%	30%
ZH	0%	56%	0%	0%	66%
RB (Heb)	12%	–	–	0%	23%
RB (Eng)	–	32%	–	–	–
ZC	–	–	3%	–	48%

ZH performed two additional reading tasks that directly tap the orthographic input lexicon – lexical decision and homophone selection.

Task 6. Lexical discrimination (Friedmann & Lukov, 2008): ZH saw 68 pairs of items, each including an existing word and its pseudo-homophone (a nonword that sounds like the target word, e.g., *chiken – chicken*, or *hoze – hose*), and chose the correctly-spelled word in each pair. Sub-lexical reading cannot distinguish such distracters from the target words, so to succeed in this task ZH had to rely on the lexical knowledge in the orthographic input lexicon.

Task 7. Homophones and palm trees (from the TILTAN battery, Friedmann & Gvion, 2003): ZH saw 41 pictures of nouns, with 2-4 words written next to each picture, and was asked to choose the word matching the picture. One of the words matched the picture, and the distracters were potentiophones (words that are written differently from the target word *and* sound differently, but may sound the same if sub-lexical reading is used, e.g., *dessert – desert*, with a picture of a cake; Friedmann & Lukov, 2008; Gvion & Friedmann, 2001) or heterographic homophones (words that sound like the target word but are written differently, e.g., *which – witch*; see Marshall, 1984 regarding a homophone comprehension task). In this task too, sub-lexical reading cannot distinguish the distracters from the target words, so ZH had to rely on the lexical knowledge in the orthographic input lexicon.

ZH made no errors in the lexical discrimination task and merely two errors in the homophones and palm trees task, which is within the normal range (Friedmann & Lukov, 2008). This further shows that his orthographic input lexicon was intact.

3.2.2 Assessing the conceptual system and the semantic lexicon

The conceptual system and the semantic lexicon are involved in word comprehension. Thus, the good performance of SZ, GE, YL, ZH, and ZC in the word definition tasks (Table 4), as well as ZH's good performance in the homophones and palm trees task (Task 7 above), indicate that their conceptual system and the semantic lexicon were intact.

Another way to examine the conceptual system and the semantic lexicon is an analysis of error types in picture naming (Task 1): a deficit in the conceptual system would result in words that are neither semantically nor phonologically related with the target word, and possibly in semantic errors too. A deficit in the semantic lexicon causes semantic errors. A deficit in the phonological output lexicon, or in the connection from the semantic lexicon to it, causes either phonological or semantic errors (Caramazza & Hillis, 1990; Rapp & Goldrick, 2000). Deficits in later stages cause phonological errors. Thus, if an individual makes only phonological errors, with no semantic or unrelated-word errors, this indicates that her deficit is in the phonological output lexicon or in later stages, and her semantic lexicon and conceptual system are intact.

For this analysis of error types, error percentages were counted out of the verbal naming attempts, namely, gestures and "don't know" responses were excluded. As shown in Table 3, all participants had many phonological errors and almost no unrelated-word errors. All but ZH and ZC had a very low rate of semantic errors (the total number of semantic errors included semantic and layered errors). The scarcity of semantic and unrelated-word errors indicates that at least for SZ, GE, YL, and RB, there is no deficit at the semantic lexicon and at the conceptual system.

To further assess the semantic system, we analyzed semantic errors in the picture naming task for situations in which the participants indicated that they knew that they responded incorrectly (e.g., cases in which they explicitly said so, or continued looking for the word and made another attempt for verbal response, etc.). Such self-monitoring situations suggest that the participant was aware of the word's meaning and that the deficit underlying the semantic error was not in the semantic lexicon but in a later stage (Rapp & Goldrick, 2000). This analysis showed that five of the participants were usually aware of the few semantic errors they made (SZ had only 1 unmonitored error out of 6 semantic errors; YL: 1/3; RB: 2/6 in Hebrew and 2/7 in English; ZC: 2/16; GE had no semantic errors), so their semantic errors probably originated in a post-semantic stage. Only ZH had many unmonitored semantic errors (10/15), a finding that does not exclude a semantic deficit.

ZH's conceptual system and semantic lexicon were therefore assessed by two additional tasks. In the *Picture Association* task (Biran & Friedmann, 2007), he saw 35 target pictures and had to choose which of two other pictures had a semantic relation with the target picture (e.g., target picture = egg, other pictures = chicken, dog). In the *object name recognition* task he saw 25 sets of four pictures depicting semantically unrelated nouns, and had to point at the picture named by the experimenter. ZH's performance was flawless in both tasks. As described above (section 3.2.1), he was also flawless in word definition and almost flawless in the homophones and palm trees task. All these are indicators to good conceptual and lexical-semantic abilities.

These data indicate that none of the participants had a deficit at the semantic lexicon or the conceptual system, and their deficit is either at the phonological output lexicon or at the phonological output buffer. ZH and ZC's semantic errors may indicate that they probably had an additional, earlier deficit, on top of their phonological deficit – probably in the access to the phonological output lexicon (but an even-earlier deficit is unlikely).

3.2.3 Phonological output lexicon or phonological output buffer impairment?

We thus narrowed down the participants' impairment to the phonological output stages. Phonological output deficits can be caused by an impairment in the phonological output lexicon, in the phonological output buffer, or in both. Thus, we separately assessed the status of the participants' phonological output lexicon and phonological output buffer.

3.2.3.1 Phonological output lexicon deficit

A deficit in the phonological output lexicon was identified by three criteria: Word frequency effect, unfaithful errors, and sub-lexical reading.

3.2.3.1.1 Word frequency effect

The phonological output lexicon is organized by frequency, so access to high-frequency words is easier than access to lower-frequency ones. The phonological output buffer and the semantic lexicon are insensitive to word frequency (Biran & Friedmann, 2012; Jescheniak & Levelt, 1994; Nickels, 1997). Thus, a frequency effect is an indication of a phonological output lexicon deficit. Frequency effect was calculated as the point biserial correlation between the word frequency and the success in producing it. We analyzed the frequency effect in picture naming (Task 1). Frequency and length did not confound in this task (r = -.045).

Table 5 shows that YL, ZH, RB, and ZC had significant word frequency effects, indicating a phonological output lexicon deficit. SZ and GE had no frequency effect, which indicates that their phonological lexicon was intact.

3.2.3.1.2 Faithful errors and approximations

We analyzed the phonological errors and approximations in the participants’ picture naming (Task 1) according to whether they were faithful to the phonemes in the target word or not. A response was categorized as faithful when the participant produced only phonemes that exist in the target word (even if not all the necessary phonemes were produced, or not in the right order). Such faithful errors were taken to indicate that the lexicon activated the correct phonemes, which later decayed or changed their order in the impaired phonological output buffer, because the buffer is a short-term memory component and the deficit lowered its capacity to the extent that it could not maintain all the phonemes of the target word (Biran & Friedmann, 2012; Laganaro & Zimmermann, 2010). A high rate of faithful errors and approximations suggests that the phonological output lexicon functions correctly and is not impaired[3]. Table 5 shows that SZ and GE had many faithful errors. This supports the conclusion that their lexicon was intact whereas the other participants had a lexicon deficit.

Table 5. Measures for phonological output lexicon deficit

	Frequency effect in picture naming (r_{pb})	Faithful errors in picture naming	Surface errors in reading
SZ	.16	55%	8%
GE	.13	67%	11%
YL	.33 ***	16%	37%
ZH	.35 ***	14%	40%
RB	.30 ***	1%	–
ZC	.38 ***	14%	–

*** $p \leq .001$

Note. The percentages of errors in this table are out of the total number of *errors*, not out of the total number of items in the task.

3.2.3.1.3 Sub-lexical reading

The phonological output lexicon participates in reading aloud, where it is part of the lexical reading route (see Figure 1), the direct route from the orthographic input lexicon to the phonological output lexicon (Coltheart et al., 2001). Therefore, phonological output lexicon impairment would not allow for reading via the lexical route (Gvion & Friedmann, 2012b), and would cause reading via the sublexical route and the grapheme-to-phoneme converter. As a result, the reader would make surface errors – reading words according to the grapheme-to-phoneme conversion rules rather than according to lexical knowledge (e.g., reading the word *listen* with a pronounced *t*, or reading the word *now* sounding like *no*). In Hebrew, such errors are especially easy to detect because there is no word that can be unambiguously read via the sublexical route. Thus, if a person reads Hebrew words aloud correctly, this means that she can read via the lexical route, and her phonological output lexicon is intact. Unlike Hebrew, many English words can be read correctly even via the sublexical route. This was the case for most of the words presented to RB in her reading task, and for this reason her reading was not analyzed for surface errors.

Sub-lexical reading was assessed by counting the number of surface errors in the word reading task (Task 3) out of the total number of errors. Table 5 shows that SZ and GE had low rates of surface errors, which suggests that their lexical route (and the phonological output lexicon) was relatively spared. YL and ZH had a high rate of surface errors, which suggests that they used the sublexical route for reading aloud and thus indicates a deficit in the lexical route. The lexical route includes two components – the orthographic input lexicon and the phonological output lexicon – and ZH's and YL's deficit can be in either

[3] Whereas the existence of faithful errors is indicative, the existence of unfaithful ones, in which the response includes phonemes that did not appear in the target word, is not indicative. This is because unfaithful errors could result from failed retrieval from the phonological output lexicon, but they could also result from a deficit in later stages.

of these two lexicons or in the connection between them. Among these three possible deficits, a phonological output lexicon deficit seems to be the most likely alternative. The orthographic input lexicon of YL and ZH was intact, as shown by their good performance in several tasks: both performed well in the word definition task, and ZH performed two additional reading tasks in which he succeeded – lexical decision and homophones and palm trees (section 3.2.1). A selective deficit in the connection between the orthographic and phonological lexicons is possible but less likely: YL and ZH had very high error rates in picture naming, a task that does not involve the orthographic input lexicon[4].

3.2.3.1.4 Conclusions of the phonological lexicon assessment

Clearly, SZ and GE did not have a phonological output lexicon deficit. They had no word frequency effect, an effect that is expected in cases of phonological output lexicon impairment. They had a high rate of faithful errors, which indicates that in many cases of retrieval failure, the phonological output lexicon selected the correct phonemes of the target word. Unlike them, the other four participants had a phonological output lexicon deficit. They showed word frequency effect and had a lower rate of faithful errors.

The findings from the analysis of surface errors further support this conclusion: surface errors, which suggest a phonological output lexicon deficit, were scarce in the reading of SZ and GE and more frequent in the reading of YL and ZH.

3.2.3.2 Phonological output buffer deficit

A deficit in the phonological output buffer was identified by several criteria: the existence of a syllable frequency effect and length effect, difficulty in nonword reading and repetition, and limited phonological short-term memory span.

3.2.3.2.1 Syllable frequency effect

It is assumed that speech production uses a stored representation of syllables, the mental syllabary, which is located between the phonological output buffer and the phonetic articulation stages (Cholin & Levelt, 2009; Levelt, Roelofs, & Meyer, 1999). Higher error rate in low-frequency syllables than in high-frequency syllables is attributed to deficits in this syllabary or in the access to it, and is typical to patients with a deficit in the phonological output buffer (Laganaro, 2005, 2008) and to patients with apraxia of speech (Laganaro, 2008). Because there is no syllable frequency data in Hebrew, we used CV-segment frequency instead (Schocken, 2008). We counted the number of within-CV phonological errors in picture naming (Task 1), and calculated the point biserial correlation between the CV segment frequency and the success in producing it. We excluded from the analysis semantic errors, unproduced CV segments, and few cases of unclear mapping between the stimulus and the response syllables. The CV-segment frequency in this task did not confound with the lexical frequency of words (r = .10).

The results (first column in Table 6) showed significant CV frequency effect for all participants. This indicates that they all have a phonological output buffer deficit.

---

[4] An alternative explanation of these findings is that SZ and GE had a phonological output lexicon deficit, and the absence of surface errors from their reading was caused by an additional deficit in the sub-lexical reading route. However, assuming that both the lexicon and the sub-lexical route are impaired is less parsimonious than assuming than neither is impaired, so the latter assumption should be favored.

Table 6. Measures for phonological output buffer deficit

	CV frequency effect (r_{pb})	Nonword production errors		Length effect (r_{pb}) [a]	
		Reading	Repetition	Picture naming	Nonword repetition
SZ	$.20^{***}$	57%	38%	$-.08^{**}$	$-.24^{***}$
GE	$.11^{*}$	47%	64%	$-.10^{***}$	$-.31^{***}$
YL	$.11^{*}$	23%	70%	$-.04$	$-.14^{**}$
ZH	$.12^{*}$	40%	Many [b]	$-.16^{***}$	
RB	$.24^{***}$		45%	$-.06$	$-.06$
ZC	$.12^{*}$		87%	$-.07^{+}$	$-.30^{***}$

$^{+}$ $p \leq .1$ * $p \leq .05$ ** $p \leq .01$ *** $p \leq .002$

[a] Point biserial correlation between successful production of each phoneme and the word length in phonemes

[b] ZH had major difficulties in this task, with no nonword repeated correctly, and the task was terminated after few items.

3.2.3.2.2 Nonword production

To further assess the phonological output buffer, we tested the participants' ability to read and repeat nonwords. Nonwords are produced via the sub-lexical route (see Figure 1), which involves the phonological output buffer but not the phonological output lexicon. Thus, errors in nonword production could result from a phonological output buffer impairment but not from a phonological output lexicon impairment.

Task 8. Nonword reading (from the TILTAN battery, Friedmann & Gvion, 2003): the participants read aloud 30 nonwords that were printed on paper. Twenty nonwords had diacritic marks, which provide information about the vowels, and ten did not (skilled Hebrew readers tend to ignore the diacritic marks, which are typically only used for reading acquisition up to second grade, therefore diacritic mark errors were not counted as errors).

Task 9. Nonword repetition (from the BLIP battery, Friedmann, 2003): the experimenter said a nonword, and the participants repeated it. There were 53 nonwords: 8 nonwords with a single syllable, and 15 nonwords with 2, 3, and 4 syllables.

All participants performed poorly in nonword production in both tasks (Table 6), which supports the conclusion of a phonological output buffer deficit.[5]

3.2.3.2.3 Length Effect

The buffer is a short term memory component, and is therefore sensitive to word length: when the buffer is impaired, longer words are produced with more phonological errors than shorter words (Biran & Friedmann, 2012; Franklin, Buerk, & Howard, 2002; Nickels & Howard, 2004; Nickels, 1997). Here we took an even more conservative approach to the analysis of length effect. According to this view, any phonological impairment – either in the lexicon or in the buffer – may result in a fixed probability to

[5] An alternative explanation, which attributes these results to additional deficits in the grapheme-to-phoneme converter and in the phoneme-to-phoneme converter (on top of a phonological output lexicon deficit), and hence requires impairments in three separate components, is far less likely than the assumption of a single deficit in the phonological output buffer.

err in each phoneme it retrieves, so the probability to produce an erroneous word increases almost linearly with the word length (Olson, Romani, & Halloran, 2007). However, a lexicon deficit cannot explain a situation in which the *per-phoneme* error rate increases with word length. Such a finding could only be attributed to a phonological output buffer deficit. A similar argument can be made with respect to nonword production – the sub-lexical conversion modules may have a fixed probability to err in each phoneme they process, and this may result in a per-word error rates that increase with the word length. However, a sub-lexical conversion deficit cannot explain per-phoneme error rates that increase with the word length – such a finding could only result from a phonological output buffer deficit. Note that absence of length effect does not indicate that a person's buffer is intact: earlier deficits, e.g., in the phonological output lexicon, may cause errors in shorter words too, rendering the length effect non-significant.

We therefore assessed whether per-phoneme length effect existed in the participants' production of words (picture naming, Task 1) and nonwords (nonword repetition, Task 9). We counted – for each target word – how many of its phonemes the participant produced correctly, and in the correct order (e.g., in *joker* → *kojer* only the phonemes o, e, and r were classified as correct). Ambiguities were resolved in a way that maximizes the number of correct phonemes (e.g., in *joker* → *jojaker*, all 5 phonemes of the target word appear in correct order). We classified only the first response to each target word, but if the participant made a partial response and then extended it, the extension was used to classify the remaining phonemes (e.g., in *cartoon* → *cal…martoon*, the phoneme classification was based on the response *caltoon*). Semantic errors were excluded from this analysis. Length effect was calculated as the point biserial coefficient between correct/incorrect production of each phoneme and the number of phonemes in the word.

This analysis (Table 6) showed a significant per-phoneme length effect for SZ, GE, YL, ZH, and ZC, thereby indicating that the phonological output buffer of these participants was impaired.

3.2.3.2.4 Memory span

Phonological output buffer deficit may be reflected in reduced STM spans in recall tasks, and indeed the participants' digit spans and nonword spans were lower than the spans of age-matched controls (digit spans: SZ, GE, and RB: 3; YL and ZC: 2; ZH: 1; nonword spans: GE and RB: 2; YL: 1; SZ: ½; ZH and ZC: 0; $p < .001$ for all. The span tasks and the control group data here and in the rest of this section are from the FriGvi battery, Gvion & Friedmann, 2012a).

In certain cases, limited STM spans may also be caused by other deficits (Martin & Lesch, 1996), e.g., in the phonological *input* buffer, yet the low digit spans still provide some corroborating evidence to the conclusion that the participants had a phonological output buffer deficit. Furthermore, the performance pattern of four participants specifically suggests that the origin of their deficit was not in an input module. The digit span of YL and ZC was 3 when they responded by pointing to digits written on paper, but it was only 2 when they responded orally – a pattern that indicates an output deficit. For two other participants, SZ and ZH, we compared spans between a serial recall task, which required oral production of words, and a word matching task (deciding whether two lists of words were presented in the same order or not), which required no oral production.

SZ’s span was significantly lower than the control group in the recall task (span = 2.5, $t(10) = 2.9$, $p < .01$) but not in the matching task (span = 4, $t(9) = 1.4$, $p = .1$). ZH had low results in both tasks, however, the pattern was still similar – his performance in the recall task (span = 1, $t(9) = 6.03$, $p < .001$) was poorer than in the matching task (span = 4, $t(9) = 2.3$, $p = .02$). Thus, the performance of both SZ and ZH significantly dropped when the task involved the phonological output buffer.

3.2.3.2.5 Conclusions of the phonological output buffer assessment

The above analyses show that all participants had a phonological output buffer deficit. They all had CV frequency effect, and all but RB had a per-phoneme length effect. These results clearly point to a buffer rather than a lexicon deficit. Further support to this assessment conclusion comes from the difficulty they had in nonword production (in both reading and repetition tasks), which could not be accounted for by a phonological lexicon deficit.[6] Finally, the participants’ low digit spans, as well as their better performance in the span tasks when no oral output was involved, are also in accord with a buffer deficit.

3.2.4 Interim summary: the locus of deficit

The series of experiments described above showed that all participants had a deficit in the phonological stages of the speech production system. More specifically, they all had a phonological output buffer deficit. For SZ and GE this was the only deficit, and the other participants had a phonological output lexicon deficit on top of their buffer deficit.

The participants’ orthographic, visual, and phonological input modules were intact. Their semantic and conceptual abilities were also intact, with the exception of ZH and ZC, who may have had an additional deficit in a slightly earlier stage – in the access to the phonological output lexicon.

4 Experimental Investigation

Having established that our participants have a deficit in the phonological output buffer, the current chapter examines the STEPS phenomenon in detail. The chapter is organized as follows. Section 4.1 will show that the participants, who make phonological errors in

[6] A possible question may arise regarding the different error rates in the production of words and nonwords. According to some researchers, a buffer deficit should cause similar error rates in words and nonwords (Goldrick & Rapp, 2007), and in our results this seems to be the case only in the reading tasks and not in the repetition tasks. However, comparing error rates in the words and nonwords that we used is not very reliable, because the stimuli were not matched for phonological parameters such as length, syllable frequency, syllable structure, etc. (especially in the case of the repetition tasks). Furthermore, the theoretical claim that a phonological output buffer deficit should cause similar error rates in word and nonword production could be questioned. An alternative view is that the earlier processing stages, which are different for words and nonwords, may activate the buffer in slightly different manners and consequently cause more errors in nonwords than in words. For example, the nonword production pathway could provide the buffer with weaker or noisier activation than the word production pathway, or have greater short-term memory requirements, whereas the lexical route and particularly the phonological output lexicon can support and re-activate the representation of words in the phonological output buffer. Indeed, some previous studies also found higher phonological error rates in nonwords than in words in cases of a phonological buffer deficit (see Caramazza, Miceli, & Villa, 1986; Romani, 1992; Shallice et al., 2000).

content words, tend to make semantic rather than phonological errors in number words. Section 4.2 aims to discover why STEPS happens: is there something special about number words, which is unique to them? Apparently not – STEPS reflects a more general property of the speech production system. We show this by identifying more categories of words that, like numbers, exhibit the STEPS phenomenon: function words, morphological affixes, and letter names. We conclude that the phonological forms of these words are stored in the phonological output buffer in pre-assembled form. Section 4.3 shows that certain changes in the task and context causes STEPS to disappear, so that number words and function words are produced with phonological errors, just like content words. Last, Section 4.4 considers some alternative accounts for STEPS and rules them out.

4.1 Demonstration of the STEPS phenomenon

4.1.1 Production of words and numbers

The first step in this part of the study was to assess whether our participants exhibited STEPS. Namely, we tested whether individuals with impairment in the phonological output buffer, who were selected for this study only because they produced phonological errors in non-number words, and who were not yet tested in number production, consistently make semantic rather than phonological errors in number words. The analysis of 57 aphasic patients by Messina et al. (2009) indicates that STEPS was observed at the group level. We wish to extend these findings by testing whether this pattern exists also within individual patients.

4.1.1.1 Tasks and analysis

The participants were tested in various word and number production tasks. We used three tasks to assess the production of words (here and henceforth, we use the term *words* alone, or *content words*, to refer to non-number words): picture naming, word repetition, and oral reading of single words (Tasks 1, 2, and 3, reported in section 3.1). The participants' performance in these tasks was compared with two number production tasks: reading Arabic numbers and number repetition.

Task 10. Reading Arabic numbers: the participants read aloud 150 Arabic numbers, one to five digits long, 30 numbers of each length. The digit zero appeared in 63 numbers. RB read the numbers both in Hebrew and in English.

Task 11. Number repetition: the experimenter read aloud 80 numbers and the participants repeated each number. The list included 18 single digits, 22 two-digit numbers, and 40 three-digit numbers. Longer numbers were not used in order not to exceed the participants' low digit span. ZH, whose digit span was extremely low, could not perform this task.

The classification of errors in word production was described in section 3.1. In number production, phonological errors were defined like in word production, and we used standard classification of semantic errors in numbers (McCloskey, Sokol, & Goodman, 1986; Sokol & Mccloskey, 1988): Lexical errors are substitutions of a number word with another existing number word. They include class errors, when the target and error number words have the same 1-9 value but belong to a different number word class (e.g., *five* → *fifty*), within-class errors, which maintain the class but not the 1-9 value (e.g., *five*

→ *six*, or *forty* → *seventy*), and substitutions of multiplier words (e.g., *five hundred* → *five thousand*). Syntactic errors are errors that completely change the number structure (e.g., *103* → *thirteen*, or 407,000 → *four hundred thousand and seven*), or errors that create an invalid number syntax (e.g., *915* → *nine hundred one five*, or *70* → *seventy zero*)[7].

Rates of the various error types were calculated as the total number of semantic or phonological errors out of the total number of verbal responses in each task (i.e., "don't know" responses were excluded from this analysis). The total number of semantic errors included semantic and layered errors in word production, and lexical and syntactic errors in number production. The total number of phonological errors included phonemic paraphasias, formal paraphasias, and layered errors, and in picture naming and repetition – also approximations. Mixed errors, unrelated word errors, and neologisms were not counted (but counting neologisms as phonological errors yielded essentially the same results).

4.1.1.2 Results

The detailed results are presented in Table 3 (Section 3.1) and Table 7. We compared equivalent tasks of word and number production. Number repetition was compared with word repetition. As for the Arabic number reading task, determining the equivalent word production task depends on the theory that one adopts regarding Arabic-to-verbal transcoding, the process of converting numbers from their digital-Arabic representation to their verbal representation: if the assumption is, along the lines of the triple-code model of numerical processing (Dehaene & Cohen, 1995; Dehaene, 1992), that Arabic-to-verbal transcoding is an a-semantic process that resembles the direct lexical reading route in word reading (see Figure 1), number reading should be compared with word reading. However, an alternative assumption, along the lines of McCloskey's number production model (McCloskey et al., 1986; McCloskey, 1992; see also Cipolotti & Butterworth, 1995), is that Arabic-to-verbal transcoding involves an intermediate semantic representation (similarly to the semantic route for word reading, see Figure 1). If one adopts this latter assumption, number reading should be compared with picture naming, because in both cases the produced words originate in a semantic representation. To avoid committing to any of the above models, we compared number reading versus both tasks – word reading and picture naming.

[7] We classified only errors that were clearly syntactic as such. Specifically, substitutions of single number words were not classified as syntactic errors, because such errors may be a lexical substitution resulting from the STEPS phenomenon. For this purpose, a "single number word" could be single words such as "three", "forty", and "hundred", but it could also be any lexicalized phonological sequence. For example, the numbers 11-19 in Hebrew consist of two separate words (e.g., 13 is /shlosh/-/esre/), but these word pairs are usually pronounced as connected (/shlosre/), and might therefore be a lexicalized sequence. Thus, reading 30 as 13 (or even 40 as 13) might be a single-word substitution and was not counted as a syntactic error. A similar non-Hebrew example is the number 70 in French, which is pronounced as "soixante-dix" (literally, "sixty-ten"). Although *soixante dix* are two words, they may be a single lexicalized sequence.

Table 7. Error rates in number production tasks

Task		All errors	Phonological	Class	Within-class	All lexical	Syntactic
Number reading	SZ	69%	4%	15%	59%	67%	3%
	GE	24%	3%	2%	14%	14%	1%
	YL	24%	4%	4%	11%	12%	4%
	ZH	38%	2%	10%	24%	29%	10%
	RB (Heb)	46%	8%	18%	16%	35%	12%
	RB (Eng)	41%	3%	16%	8%	25%	18%
Number repetition	SZ	64%	3%	0%	48%	48%	5%
	GE	26%	1%	0%	18%	18%	3%
	YL	38%	4%	4%	31%	33%	6%
	RB (Heb)	38%	21%	0%	28%	28%	4%
	RB (Eng)	40%	13%	4%	31%	36%	3%
	ZC	64%	18%	9%	34%	41%	9%

The comparison of phonological vs. semantic errors is shown in Figure 2 (words with no phonological or semantic errors were not counted; layered errors such as *giraffe* → *zebka* were counted twice). Figure 2 shows that the STEPS pattern was observed for each of the participants: for each of the word production tasks and the equivalent number production task, the ratio of phonological errors out of the total number of errors was significantly higher in words than in numbers ($p < .001$ per task and participant). These findings cannot be explained by suggesting that word production is more difficult than number production, or vice versa, because the overall error rates in words and numbers were similar: the average error rates in reading were 39% in both numbers and words; in repetition – 46% in numbers and 37% in words; and in picture naming – 57%.

In another analysis we compared the phonological and semantic error rates between word production tasks and number production tasks. The results (Table 8) were clear: all participants had significantly more phonological errors in content words than in number words, and significantly fewer semantic errors in content words than in number words, in all compared tasks. In the number repetition task, RB and ZC did make phonological errors but their rate was still lower than the semantic errors and lower than the rate of phonological errors they made in words.

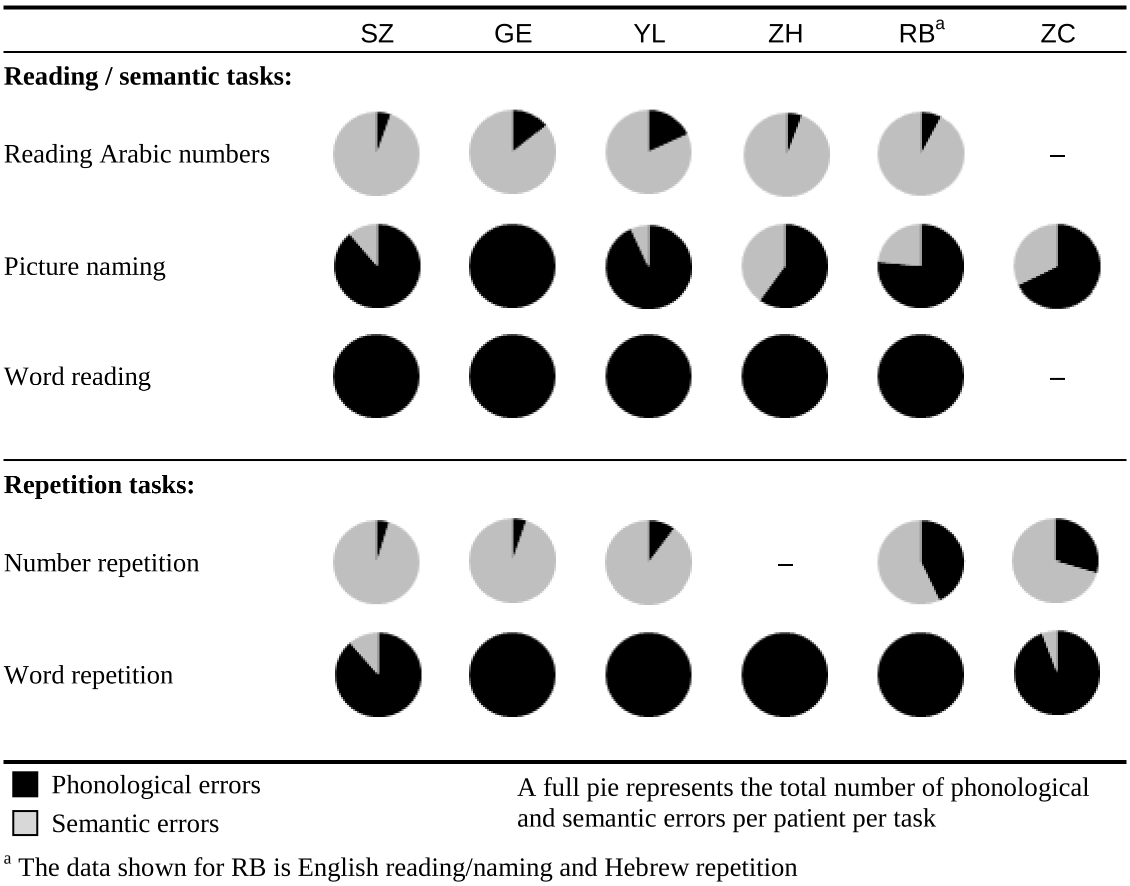


Figure 2. Semantic vs. phonological errors in speech production: numbers vs. words

An alternative explanation, according to which the semantic errors in number words are a coincidental result of random phonological substitutions, is unlikely. In Hebrew, single-phoneme manipulation on number words never result in another number word, yet they can often result in a non-number word. Thus, random phonological substitutions in number words should yield a content word more often than a number word. The results, however, are the complete opposite: in all number production tasks the participants made many semantic errors resulting in number words, but never produced a content word[8].

---

[8] A special case is class errors: in Hebrew, like in English, decade names are derived from the unit names by a simple phonological manipulation (usually adding the suffix /im/ and changing a vowel or two). Thus, class errors (e.g., *three* → *thirty*, which in Hebrew is /shalosh/ → /shloshim/) could be phonological errors in disguise. However, even when excluding the class errors, STEPS was observed for all participants (semantic error rate in number reading – SZ: 65%, GE: 25%, YL: 15%, ZH: 31%, RB Hebrew: 27% and English: 28%; in number repetition, RB English: 34%, ZC: 41%, and for the other participants like in Table 8; the results of the statistical comparisons are the same as reported in Table 8).

Table 8. Error rates in the basic production tasks of words vs. numbers

	Picture Naming		Word Reading		Word Repetition		Number Reading		Number Repetition	
	Phono.	Semantic	Phono.	Semantic	Phono.	Semantic	Phono.	Semantic	Phono.	Semantic
SZ	55%***	6%***	70%***	0%***	23%***	0%***	4%	69%	3%	60%
GE	41%***	0%***	18%***	0%***	28%***	0%*	3%	18%	1%	18%
YL	43%***	3%***	21%***	0%***	33%***	0%***	4%	18%	4%	35%
ZH	30%***	20%*	55%***	0%***	61%	0%	2%	33%	–	–
RB_{heb}	31%***	9%***	–	–	23%	0%***	8%	39%	21%	28%
RB_{eng}	36%***	10%***	29%***	0%***	–	–	3%	35%	13%	36%
ZC	42%	17%	–	–	43%***	3%***	–	–	18%	43%

Corresponding word and number tasks were compared: picture naming and word reading vs. number reading, and word repetition vs. number repetition: * $p \leq .05$ *** $p \leq .001$

Phono. = Phonological

4.1.1.3 Production of single digits

Patients who show STEPS often produce single digits better than multi-digit numbers, with virtually no semantic errors in single digits (Delazer & Bartha, 2001; Girelli & Delazer, 1999; Marangolo et al., 2004; see Table 1). Our participants too had low semantic error rate in single digits, both in reading (GE and ZH: 7%, YL: 3%, RB: 13%) and in repetition (no errors for GE, YL and RB, and a single error for SZ and ZC). The error rates in single-digit numbers were significantly lower than in longer numbers (per participant, $p \leq .01$ in reading and $p \leq .002$ in repetition) and were not significantly higher than zero ($p > .05$ for RB's reading, $p > .1$ for all other comparisons). The only exception was SZ, who made 37% errors even in reading single digits (which is significantly higher than zero, Fisher's $p < .001$). However, his error rate in longer numbers was even higher (78%, $\chi^2 = 24.5$, $p < .001$).

4.1.1.4 Syntactic errors in number production

An interesting pattern with respect to syntactic errors in number production can be noticed in Table 7: the proportion of these errors out of all errors was low for participants who had only a phonological output buffer deficit (SZ: 3%, GE: 3%) and high for participants who had an additional, earlier deficit (YL: 14%, ZH: 28%, RB: 28% in Hebrew and 53% in English). The implications of this pattern are elaborated in the discussion (in Section 5.3).

4.1.2 Number production from the semantic system

To further assess the error patterns in number production, we used two tasks in which the numbers originate in the semantic system rather than in reading or repetition.

Task 12. Reading analog clock: the participants saw 15 analog clocks printed on paper, and were asked to say the time shown in each clock. The clock diameter was 6.6 cm, and the numbers 1-12 were printed on them. The hour and minute hands were painted blue and red, respectively.

Task 13. Mental calculation: the participants saw simple arithmetic exercises printed on paper, and were requested to say the result without reading aloud the exercise itself. The task included a block of 21 short exercises (7 addition, 7 subtraction, and 7 multiplication exercises) in which the operands were single digits and the answer had one or two digits; and a block of 20 longer exercises (10 additions and 10 subtractions) in which the answer had 3 digits, one of which was zero. The exercises did not require carry or borrow operations. SZ and RB performed only the 20 longer exercises.

As Table 9 shows, the pattern of errors in these tasks was similar to the pattern in other number production tasks – the participants produced predominantly semantic errors, and almost no phonological errors.

Table 9. The error rates in semantic number production tasks

	Reading Analog Clock		Mental Calculation	
	Semantic	Phonological	Semantic	Phonological
SZ	73%	0%	75%	0%
GE	20%	0%	5%	2%
YL	–	–	10%	7%
ZH	13%	0%	10%	2%
RB	13%	7%	45%	15%
ZC	40%	0%	–	–

4.1.3 Great words: Does the error pattern in numbers result from their phonological features?

Is the special pattern of errors in number words really a result of their being numbers and processed differently in the word production system, or is it somehow related with phonological characteristics of the number words? As STEPS was found in several languages, the latter alternative seems less likely. Still, we assessed this alternative by asking the participants to read words in which the phonological form of a number word was embedded, e.g., the Hebrew word *xoshesh*, worries, in which the number *shesh*, six, is embedded (relevant examples in English are *great* (8), *tomorrow* (2), *before* (4), *content* (10), etc.). Eighteen such gr8 (great) words were embedded among other words in the word reading task (Task 3). The words were written in their ordinary format (i.e., “great” and not “gr8”). If the STEPS phenomenon is caused by phonological properties of the Hebrew number words, we should expect no phonological errors, and possibly expect semantic errors, in the number part of the gr8 words. Table 10 compares the phonological error rate in the number part of the gr8 words with their rate in Arabic number reading (task 10) and on the reading of 100 words (task 3). It shows that except for YL, the phonological error rate in gr8 words resembles words (in fact, it was even slightly higher than in words). There was no case in which the embedded number word was substituted with another number word. Thus, STEPS is related to the classification of a produced word as a number word, rather than to its phonological characteristics

(Section 4.3 will further elaborate on this classification). A similar analysis of gr8 words in Italian yielded essentially the same results (Bencini et al., 2011).

Table 10. The phonological error rate in gr8 words vs. words and number reading

	Gr8 words	Words (Task 3)	Numbers (Task 10)
SZ	67%	70%	4%
GE	22%	18%	3%
YL	11%	21%	5%
ZH	61%	55%	2%

4.1.4 Interim summary

The results of all the comparisons between the error patterns in words and numbers were clear. Whereas numbers were produced with semantic errors (mostly substitution of digits), words were produced with phonological errors, and with fewer or almost no semantic errors. These results were consistent for each of the six participants, although the participants were included in the study solely on the basis of phonological errors in their word production.

4.2 The origin of STEPS: The building blocks hypothesis

An interesting explanation for STEPS, suggested by Cohen et al. (1997), is that the speech production system includes phonological building blocks of different sizes. According to such an approach (which will be hereby called *the building blocks hypothesis*), the basic phonological units in the speech production system are not only phonemes, but may also be larger pre-assembled phonological units. The building blocks of non-number words are phonemes, but the building blocks of number words are whole words – digits like *six*, multipliers like *thousand*, and other number words like *sixty* and *sixteen*. Crucially, a substitution error in the phonological production system, which substitutes a phonological unit with another, will result in different error types for building blocks of different sizes, and hence will give rise to different errors in words and numbers. When producing words, in which the phonological building blocks are phonemes, substituting a phonological unit by another will result in a phonological error. When producing number words, the phonological building block is a whole word, and substituting one unit with another will result in a semantic error. This hypothesis assumes that substitution errors involve whole, atomic, already-assembled phonological units, which would rarely be broken into parts.

The patient described by Cohen et al. (1997) made semantic errors not only in numbers, but also in letter naming. Cohen et al. suggested that this finding supports the building blocks view: in French, like in English, the letter names are the “building blocks” used for producing a more complex structure – acronyms. Phonemes, number words, and letter names, they said, are “the building blocks of speech”. The term “building block” should be interpreted in this context as a phonological building block, larger than one phoneme, which belongs to a limited set of phonological units. The units from this limited set are used in a productive phonological process that assembles them into a more complex phonological structure (e.g., assembling letter names into acronyms, or number words into multi-digit numbers). The building blocks hypothesis assumes that when this is the case, the building blocks would become atomic phonological units.

We assessed three additional types of possible building blocks, to test whether being a building block indeed turns these words, or phonological sequences, into atomic phonological units: *function words* can be viewed as building blocks in the process of sentence construction. *Morphological affixes* may be the building blocks in the morphological composition of words. We also assessed the processing of *letter names*, which were already found as atomic phonological units in French by Cohen et al. (1997), but testing them in Hebrew can shed further light on the matter, as we explain below.

4.2.1 Function words

Function words may be building blocks in the productive process of forming phrases and sentences. Some function words are free, namely, appear as separate words (e.g., *on* in "He relies on his instincts"), and are therefore building blocks in the formation of phrases and sentences. Other function words are part of words, termed "bound function words". Such bound function words can appear at the beginning or the end of the word (e.g., She'<u>ll</u> be there, or French and Italian clitics like *l'* in *l'opera*). In Hebrew, both bound and free function words are frequent, and bound function words are always attached to the beginning of a word.

As building blocks, function words may also be atomic phonological units. To assess this possibility, we analyzed the participants' production of free and bound function words in two tasks: sentence production and function word elicitation.

4.2.1.1 Sentence production

Task 14. Sentence production: the participants read or repeated 68 short sentences of 3 to 6 words, which included 254 words: 60 free function words, 93 bound function words, 173 content words, and 21 other words that will be discussed later (in Section 4.3.2).

Errors in function words were classified into semantic errors, phonological errors, and omissions. Semantic errors were defined as a substitution of a function word with another function word (e.g., *<u>ha</u>-bayit*, the house → *<u>le</u>-bayit*, to a house), or an addition of a function word. Phonological errors were defined as a substitution of a function word with a phoneme sequence that is not an existing function word. Note that although each bound function word is phonologically attached to a content word, the function word and the content word were analyzed separately. For example, the substitution *<u>ha</u>-bayit* → *<u>le</u>-gayit* was counted as a semantic error in the bound function word (substituting the function word *ha* with the function word *le*) and a phonological error in the content word (*bayit* → *gayit*). Morphological errors, namely, substitutions of one morphological affix by another (e.g., *relying* → *reliable*, *walked* → *walking*), were excluded from this analysis, which focused on content and function words. We also excluded from this analysis surface errors in reading. YL and ZH used the sublexical route in reading (see Section 3.2.3.1), which resulted in many errors typical to surface dyslexia (e.g., reading "now" as "no", or "deaf" as "deef"). These errors were excluded from this analysis, both in content words and in function words.

As predicted by the building blocks hypothesis, the participants produced significantly fewer phonological errors in bound and free function words than in content words (Table 11). Also in accord with the building blocks hypothesis, there were more semantic errors in function words than in content words (i.e., substitutions of one function word with another), a difference that was significant in the group level as well as for some individual participants. Overall, the semantic error rate in function words was not as high as in number words. In this respect the function words, being single words, resemble single digits, in which the semantic error rate was also very low (see section 4.1.1.3). Finally, there were no omissions of content words but there were omissions of bound function words. The omissions of whole function words are also in line with the building blocks hypothesis, because like semantic errors, whole-word omissions suggest that the word is processed as a single atomic unit.

Table 11. Errors in function words vs. content words in sentence production

		Phonological errors			Semantic errors			
	Task	Content words	Free function words	Bound function words	Content words	Free function words	Bound function words	Omissions of bound function words
SZ	Reading	19%	0% ***	0%***	0%	0%	7%**	3%
	Repetition	41%	3% ***	0%***	2%	5%	2%	4%
GE	Reading	19%	3% **	1%***	0%	3% $^{+}$	4%*	5%
	Repetition	20%	2% ***	0%***	0%	2%	2%	8%
YL	Reading	12%	2% **	0%**	1%	0%	4%$^{+}$	0%
ZH	Reading	40%	3% ***	0%***	0%	3% $^{+}$	4%**	2%
RB	Reading [a]	30%	0% **		3%	3%		
ZC	Repetition	44%	1% ***	0%***	2%	12% **	4%	4%
Group average		28%	2% **	0%**	1%	4%*	4%*	4%

Function words vs. content words: $^{+}\ p \leq .1$ $^{*}\ p \leq .05$ $^{**}\ p \leq .01$ $^{***}\ p \leq .001$

[a] RB’s reading results are of her English text reading in Task 3.

The high rate of semantic errors in bound function words is unlikely to result from random phonological substitutions. If this were the case, we could expect that phonological errors in content words will also accidentally result in bound function words (e.g., for the content word *xarak*, bug, the phonological error *barak*, lightning, includes the sub-string “ba”, a bound function word that did not exist in the target word). However, this phenomenon was not frequent in content words: only 9% of the group’s phonological errors in content words resulted in a response that included a phoneme sequence that is an existing bound function word, as opposed to 97% of the errors in bound function words ($p < .02$ per participant).

4.2.1.2 Function word elicitation

The second task that was used to assess the participants' production of function words was an elicitation task. In this task, the participants did not hear or see the target function words, but had to find the appropriate function word on their own. ZC did not perform this task.

Task 15. Function word elicitation. The participants saw 30 pictures depicting simple situations. A short sentence describing each picture was written under it, and a function word was missing from the sentence (marked with an underlined gap). The task included 30 function words, 28 free and 2 bound. The experimenter read the sentence aloud, and the participant repeated it and completed the missing function word. To make the required answer clearer, the pictures were presented in pairs, and the two missing function words in each pair of sentences were opposites. For example, one pair of pictures showed an airplane flying above or under a cloud, and both pictures had the same sentence under them: *The airplane is flying _____ the cloud*.

We compared the phonological versus semantic error rates in this task, excluding mixed errors. As Table 12 shows, four participants had more semantic than phonological errors and this difference was significant for three of them. This finding further supports the notion that function words are phonological building blocks, stored as pre-assembled units.

Table 12. Phonological vs. semantic error rates in function word elicitation

	Phonological errors	Semantic errors	Mixed errors
SZ	0%*	17%	3%
GE	3%	0%	0%
YL	3%*	20%	3%
ZH	17%	20%	3%
RB	7%**	47%	0%

* $p \leq .01$ ** $p \leq .005$

4.2.2 Morphological affixes

Hebrew is a Semitic language with rich morphology. Most words are constructed of a 3-letter root and a morphological template. Derivational and inflectional affixes in Hebrew are prefixes, suffixes, and infixes. The building blocks hypothesis predicts that morphological affixes may also be atomic phonological units, because morphemes, similarly to number and function words, are building blocks participating in a specific productive process – the construction of morphologically complex words.

According to this view, morphologically complex words (e.g., *drinking*) are processed in the following way: the word's stem is processed as separate phonemes (*d-r-i-n-k)*, whereas the non-root morpheme (the morphological affix) is processed as a single atomic phonological unit (-*ing*) – similarly to bound function words. The building blocks hypothesis predicts that errors in the stem would be phonological – omission or substitution of a phoneme (*drinking* → *driking*, *drilking*), whereas the errors in the morphological affix would be morphological – omission of a whole morphological affix, or substitution by another existing affix. A morphological error may result in an existing word (e.g., *drinkable*) or in a nonword (e.g., *drinkly*).

We assessed the production of morphological affixes in three tasks: sentence production (Task 14), word reading (Task 3), and another task in which the participants read and repeated morphologically complex words (Task 16). The results of the three tasks were pooled together, as the pattern of errors was similar across tasks.

Task 16. Multi-morphemic word production: A list of 19 morphologically complex words (and 18 filler monomorphemic words) was presented in two modes: a word repetition task and a reading aloud task. RB and ZC performed only the repetition task.

Table 13. Error rates in production of morphological affixes

	Explained morphologically			Not explained morphologically
	Morphological errors	Morphological Omissions	Total	Phonological errors
SZ	14%	3%	17%	2% ***
GE	7%	1%	8%	1% ***
YL	6%	0%	6%	1% ***
ZH	21%	5%	26%	1% ***
RB	15%	0%	15%	4% **
ZC	15%	2%	17%	2% ***

Morphological vs. phonological errors: ** $p \leq .02$ *** $p < .001$

The results, shown in Table 13, were very clear: in average, 91% of the errors in the morphological affixes were either omissions or morphological substitutions, and only 9% of them were phonological errors that could not be explained morphologically. The difference between the phonological error rate and the morphologically-explained error rates was significant for each of the participants. Thus, the STEPS phenomenon extends to morphological affixes too, as predicted by the building blocks hypothesis.

Given the very high rate of morphologically-explained errors, it seems unlikely that they result from random phonological substitutions. Nevertheless, we assessed this possibility. If indeed random phonological substitutions result in valid morphological affixes when the substitution occurs in the affix part of the word, we can expect the same to happen in the word stems, i.e., phonological substitutions in word stems should also yield valid morphological affixes (e.g., for the content word *taklit*, a record, the phonological error *taknit* includes the sub-string “ni”, a valid morphological affix that did not exist in the target word). However, the results did not confirm this prediction: in the sentence production task, averaged over participants, 95% of the errors in morphological affixes resulted in a valid morphological affix, but only 20% of the errors in the stems of content words included a valid morphological affix (p < .001 per participant; omissions and the creation of single-phoneme affixes were excluded from this analysis). This indicates that the morphological errors in morphological affixes were not random phonological substitutions.

Does the speech production system process morphological affixes as building blocks (and as atomic phonological units) only when they are a part of an existing word, or does it identify morphemes in nonwords too? If morphological affixes are identified in nonwords, this would imply that the morphological affixation process, and presumably the representation of morphological templates, does not have to depend on lexical representations and mechanisms. To answer this question, we requested the participants to produce *zicklings* – nonwords with valid morphological affixes, such as the noun “quattners” or the verb “nimzed”.

Task 17. Production of zicklings (morphologically complex nonwords): SZ and GE saw 60 morphologicaly complex nonwords – 30 pseudo-verbs and 30 pseudo-nouns, all with valid Hebrew word templates and morphological affixes. The experimenter read each word aloud, and the participant then read it from paper. The task was administered in a mixed repetition/reading design because hearing the nonword disambiguates the vowels, which are under-represented in written Hebrew (because skilled Hebrew readers tend to ignore the diacritic marks), and seeing the nonword may prevent some confounding effects, e.g., limited input phonological working memory. The pseudo-verbs in the task were preceded by pronouns (e.g. “I nimzed” rather than just “nimzed”), which were not analyzed.

The zickling production was analyzed like the word production was – by classifying the errors in the morphological affixes into morphological, phonological, and omissions. The results were similar to the production of real morphologically complex words: 81% of SZ’s errors and 75% of GE’s errors were morphological substitutions or omissions, and less than 25% were phonological errors. This shows that the speech production system identifies morphological affixes not only in words but also in nonwords.

Taken together, the findings described in this Section show that morphological affixes, similarly to number and function words, are stored as atomic pre-assembled phonological units. Importantly, this indicates that the phonological forms of morphological affixes are stored separately from the word stems, and that morphological affixation is a productive process rather than lexicalized knowledge.

4.2.3 Letter names

When Cohen et al.’s (1997) patient named the 26 alphabet letters in French, he made no phonological errors. Cohen et al. suggested that letter names are atomic phonological units too, because they are building blocks in the productive process of uttering acronyms, which are pronounced in French (like in English) as a sequence of letter names. Interestingly, Hebrew shows a completely different pattern with respect to acronyms. Most Hebrew acronyms are pronounced as whole words rather than as a sequence of letter names. These words are usually created by adding vowels to the acronym’s letters. For example, the acronym פ.ק.מ. (P.K.M., Pikadon Kcar Moed, which means a short-term deposit) is pronounced as /pakam/ rather than as the sequence of letter names (pe – kuf – mem). There are few acronyms in Hebrew that are pronounced, like in English, as the sequence of letter names, but they are far less frequent than in English and French, and far less frequent than acronyms pronounced as words. For this reason, we assumed that the role of letter names as building blocks in acronym production is weaker in Hebrew than in English and French. It is therefore possible that English or French letter names are phonological building blocks whereas Hebrew letter names are not.

To assess this possibility, we asked the participants to produce Hebrew letter names using an acronym reading task. If they make phonological errors in naming Hebrew letters, this would support the building blocks hypothesis. ZH named letters both in Hebrew and in English, so we could compare his performance between the two languages.

Task 18. Acronym production: the participants saw sentences that included acronyms and were asked to read the whole sentence. SZ saw sentences in which the acronyms were underlined, and was instructed to silently read the whole sentence but say aloud only the underlined words. The Hebrew task included 19 sentences with 14 two-letter acronyms that are pronounced as a sequence of letter names. The sentences also included 21 filler acronyms that are pronounced as whole words, but they were not analyzed. The English task included 17 sentences with 30 acronyms, 2-4 letters long. SZ, GE, and YL performed the Hebrew task, RB performed the English task, and ZH performed both tasks.

Table 14 shows the error rates in acronym production, excluding letters for which no verbal response was made. The results in Hebrew acronym production were clear: the four participants who performed the task in Hebrew had more phonological than semantic errors. Conversely, phonological errors were scarce in English acronyms (a task that was done by ZH and RB), similarly to the pattern observed in French letter names (Cohen et al., 1997). This difference between languages is in accord with the building blocks hypothesis: phonological errors were frequent in naming Hebrew letters, where letter names seldom function as building blocks, but did not appear in naming English and French letters, which *are* building blocks.

Table 14. Phonological vs. semantic error rates in acronym production

	Hebrew		English	
	Phonological	Semantic	Phonological	Semantic
SZ	25%*	4%		
GE	25%*	7%		
YL	52%***	4%		
ZH	32%**	7%	2%	14%**
RB			3.5%	6%

* $p < .05$ ** $p < .01$ *** $p < .001$

Note. Percentages are specified with respect to the total number of letters.

The most interesting finding in this context is the performance of ZH, who produced letters both in Hebrew and in English. ZH, whose dominant error type in Hebrew acronyms was phonological, made only 2% phonological errors in the English acronym task, significantly fewer than in Hebrew ($\chi^2 = 21.8$, $p < .001$). This within-participant difference provides strong support to the notion that English letter names are indeed phonological building blocks whereas Hebrew letter names are not.

Another task we used to assess letter production was **single letter naming**: the participants saw 50 Hebrew letters and said aloud the letter name. The trend in this task was similar to the acronym production task, although not as clear: SZ, GE, YL, and ZH had too few errors for a reliable analysis of error types (12% or less). The two other participants, RB and ZC, indeed had significantly more phonological than semantic errors (RB: 18% vs. 4%, ZC: 24% vs. 8%, two-tailed $p < .03$).

4.2.4 Interim summary: The origin of STEPS

The above series of experiments showed that there are several other categories of words that, like number words, are produced with semantic rather than phonological errors. It is clear now that STEPS is not a phenomenon specific to numbers, but reflects a more general property of the speech production system. The findings are consistent with the hypothesis that several word categories are phonological building blocks that are stored as pre-assembled phonological units: function words, morphological affixes, English and French letter names, and number words. Certain phonological deficits result in substitutions of one phonological unit by another. In the case of monomorphemic content words, the phonological unit is a phoneme, resulting in a phonological error – a substitution of a phoneme by another phoneme. In the case of number words and function words, the phonological unit is a whole word, resulting in a semantic error – the substitution of a number word by another number word and of a function word by another function word. For morphological affixes, the phonological unit is the whole affix, resulting in a morphological error.

Where are these building blocks stored? In Section 3.2 we showed that the only deficit common to all participants was a phonological output buffer deficit. We conclude that the phonological output buffer is the module responsible for STEPS, and hence that the pre-assembled phonological representations of number words, function words, morphological affixes, and English and French letter names are stored in the phonological output buffer. This conclusion is also supported by the results of the zickling production task (Task 17): The morphological errors in zickling production can only originate in a module that is a part of the nonword production pathway, and the phonological output buffer, but not the phonological output lexicon, is such a module (see Figure 1). Furthermore, if we concluded that morphological affixes are stored as pre-assembled units in the phonological output buffer, it follows that the affixes cannot be assembled with the word's phonemes in an earlier processing stage. Thus, morphological affixation is done in the phonological output buffer and not earlier.

4.3 If it functions as a building block, it is a building block

Is it the case that number words, function words, and morphemes are simply burned as preassembled atomic phonological units, or are they atomic phonological units only when the speaker classifies the word as functioning as a building block in a productive process? Namely, if a speaker says the sequence "eight" but not with numeric meaning, would this sequence still be a pre-assembled phonological unit?

To assess this question, we evaluated the way the participants produce number words and function words when they appear in their "natural" role, functioning as building blocks, and when they function as ordinary content words. The word role hypothesis assumes that the pre-assembled phonological representations of number words and function words are used only when the words actually serve as building blocks. This hypothesis predicts that when number or function words do not appear in their natural role, they would be produced with phonological rather than semantic errors, just like any other word.

The productive process in which function words participate is syntactic – the construction of sentences. Thus, to grant function words their syntactic role or deprive them of it, we showed them either as part of sentences or in a plain list of words. For number words, we created a task in which the numbers appeared without a numeric meaning.

4.3.1 Function words

We compared the phonological error rate in function words when produced in syntactic context (in Task 14, sentence production) with the error rate when the same function words were produced without syntactic context, in a word list.

Task 19. Producing function words and content words in a word list: the participants were presented with a list of 160 single words. SZ, GE, and ZH read them from paper, and ZC repeated them. The list included 59 free function words (the same function words that were included in Task 14, sentence production), 53 filler content words, and 48 other filler words.

As Table 15 shows, the results confirm the prediction of the word role hypothesis: all four participants had fewer phonological errors in function words when they were embedded in sentences than when they were presented as isolated single words. This pattern was significant for three of the participants.

Table 15. Phonological errors in function words

	With syntactic context (in a sentence)	Without syntactic context (in a list)
SZ	0%	14%***
GE	3%	8%
ZH	8%	29%***
ZC	5%	29%***

*** $p \leq .002$

If function words are not phonological building blocks when produced without syntactic context, we should expect the phonological error rates in function and content words to be similar when reading a word list. We therefore compared function words and content words in the word-list reading tasks: Task 19 (36 content words and 11 function words) and Task 3 (63 content words and 17 function words). To control for word length we included only words with 4-6 phonemes, because many of the function words were short. This comparison, shown in Table 16, provides full support of the word role hypothesis: when function words were read in a list (without syntactic context), the phonological error rate in them was as high as in the length-matched content words – in fact, for some participants it was even slightly higher. Thus, function words exhibit the STEPS phenomenon only when they are produced in syntactic context, embedded in sentences. Finally, there were almost no semantic errors when producing function words without syntactic context: no errors for SZ, GE, ZC, and 8% for ZH (for ZC this error rate was lower than in sentence production, Fisher's $p < .01$, and for the other participants the error rates were comparable, Fisher's $p > .21$).

Table 16. Phonological errors in words with 4-6 phonemes without syntactic context (in a list)

	Function words	Content words
SZ	50%	47%
GE	18%	17%
YL	18%	19%
ZH	50%	57%
ZC	45%	42%

4.3.2 Numbers

How are number words processed when they are deprived of their numeric meaning? To assess this question, we used number words that were a part of a familiar name or expression. For example, the number 7, which is pronounced in Hebrew "sheva", appeared as "Be'er Sheva" – the name of a city in Israel. Corresponding examples in English are the brand name “*Nine* West”, the expression “give me high *five*”, and names of places like *Three* Forks Montana. In such a context, the numeric meaning of the number words is expected to be less salient. The sentence production task (Task 14) included 21 such number words, which were embedded in 18 of the sentences. The word role hypothesis predicts that the “Nine West” number words would be produced with phonological errors, like content words.

Table 17 compares the production of “Nine West” number words with the production of content words in the sentence production task (Task 14), and with the production of single digits in Arabic number reading (Task 10) and in number repetition (Task 11). The group-level comparison confirmed the prediction of the word role hypothesis – there were more phonological errors in the “Nine West” numbers than in the production of single digits. The rate of semantic errors in the “Nine West” task was low (no errors in SZ's and GE's repetition and ZH's and YL's reading; one error in GE's reading; and 2 errors in SZ's reading and ZC's repetition).

Table 17. Phonological errors in production of “Nine West” numbers vs. single digits and content words

	Task	Nine West	Single digits	Content words
SZ	Reading	10%	0%	20%
	Repetition	24%	0% *	44% *
GE	Reading	5%	0%	20% *
	Repetition	5%	0%	21% +
YL	Reading	5%	2%	14%
ZH	Reading	29%	0% *	43% *
ZC	Repetition	33%	6% *	47%
Group		16%	1% **	32% **

+ $p < .1$ * $p < .05$ ** $p < .01$

An alternative explanation could attribute the high phonological error rate in the “Nine West” task to the fact that in this task, unlike the reading and repetition tasks, the numbers were embedded in sentences. We assessed this alternative account using another task in which numbers were embedded in sentences, but contrary to the “Nine West” numbers, they had salient numeric meaning – e.g., "the man ate *two* apples".

Task 20. Numeric numbers in sentences: the participants produced 29 sentences that included 31 single-digit number words, all with numeric meaning. The sentences also included 69 content words and 14 function words. Like in the “Nine West” task, the numbers were written as words and not as digits. ZC repeated the sentences and the other participants read them.

The “sentence vs. single words” alternative explanation predicts that number words in the “numeric numbers” task would be produced with many phonological errors, because they are embedded in sentences. The word role hypothesis, however, focuses on the fact that the “numeric numbers” have numeric meaning, and therefore predicts that they be produced with only few phonological errors.

Table 18 confirms the prediction of the word role hypothesis: the phonological error rate in the "numeric numbers" task was similar to the phonological error rate in production of single digits, and lower than the phonological error rate in the “Nine West” numbers and in content words (Task 14). Thus, the reason for the high phonological error rate in “Nine West” number words was indeed the absence of their numeric meaning rather than the fact that they were embedded in sentences.

Table 18. Phonological errors in "numeric numbers" vs. other words

	Numeric numbers	Single digits	Nine West	Content words
GE	0%	0%	5%	22%***
ZH	10%	0%	29%$^{+}$	42%***
ZC	10%	6%	33%*	32%***
Group	6%	1%$^{+}$	22%***	32%***

$^{+}$ $p \le .1$ * $p \le .05$ *** $p \le .002$

4.3.3 Interim summary: The word role effect

The experiments presented in this section showed that the STEPS phenomenon depends on the role in which a word appears. When number words and function words appear in the relevant role (number words with numeric meaning and function words with syntactic role), they are produced with semantic rather than phonological errors. Conversely, when the number or function words were deprived of their role by changing the task and context, they were produced with many phonological errors, and without semantic errors, just like content words[9]. In the case of number words, it could be argued that the findings reflect a lexical difference between “*nine* children” and “*Nine West*” – namely, that the two *nine*s are two separate lexical units (homophones). Under this view, the word role effect observed in the “Nine West” task is an extension of the gr8 words phenomenon (Section 4.1.3). However, this interpretation cannot account for the word role effect in function words: the manipulation we used on function words – presenting them either in a sentence or in a list – changed their grammatical environment but not their meaning, so it is unlikely that the different presentation modes activated two different lexical units.

The word role effect suggests that number and function words are stored not only as pre-assembled phonological units but also as separate phonemes, like content words. The speech production system selects one of these two representations according to the role in which the word appeared: it selects the pre-assembled phonological unit when the word actually functions as a building block, but the segmented representation when the word functions as an ordinary content word.

4.4 Ruling out alternative accounts of STEPS

In Section 4.2 we showed that the STEPS phenomenon is not limited to numbers, but extends to function words, morphological affixes, and English (but not Hebrew) letter names. These findings strongly support the building blocks hypothesis. However, several alternative accounts could be suggested for STEPS. We assessed three such alternative hypotheses and ruled them out.

[9] The low rate of semantic errors could be interpreted as further evidence to the word role effect. However, an alternative explanation is that semantic errors were absent from the "role-deprived" tasks (presenting the function words in a list rather than in sentences, and using single-digit rather than multi-digit numbers) because these tasks were easier than the previous tasks.

4.4.1 Ruling out the frequency hypothesis

One possible alternative account could ascribe our findings to frequency. Such account would suggest that number words are produced without phonological errors because they are more frequent than other words. A first important problem with such an account is that it only explains the reduction in phonological errors in number words, and cannot explain the elevated number of semantic errors. Previous studies have refuted the frequency hypothesis in two ways: Bachoud-Lévi and Dupoux (2003) compared the phonological error rate between number words and non-number words with similar frequencies. They found that contrary to the frequency hypothesis, number words were still produced with fewer phonological errors than their frequency-equivalent word counterparts. Cohen et al. (1997) looked for correlation between the word frequency and phonological error rate, and found none. Such a correlation (namely, frequency effect) was also not found for two of our patients (SZ and GE), the patients who were impaired exclusively in the phonological output buffer. In fact, we do not expect to find frequency effect in any patient who has a selective deficit in the phonological output buffer, because frequency affects the phonological output lexicon, but not the phonological output buffer (Jescheniak & Levelt, 1994; Nickels, 1997).

Another method we used to assess the frequency hypothesis was analyzing the errors in number production in a foreign language for a person who does not master it and does not speak it fluently. For such a person, we assumed that the frequency of number words in the foreign language is lower than the frequency of frequent non-number words in Hebrew. The frequency hypothesis therefore predicts a high rate of phonological errors when saying number words in a second language that is not frequently used. We tested this for ZH, who is not a native speaker of English. He studied English in school, his level of English is average, and at the time he participated in the study it was in no way comparable with his level of Hebrew. This was evident also from his performance in picture naming: when he named the 50 most frequent words in the picture naming task, we had indications of him knowing only 34% of the English words compared with 96% of the same words in Hebrew (counting the object names he was able to retrieve, even with phonological errors, long hesitations, or approximations).

Task 21. Number reading in English: ZH read aloud in English 80 Arabic numbers with 1-5 digits (15-16 numbers of each length), printed on paper; 48 of the numbers included the digit zero.

ZH had only 7.5% phonological errors in reading English numbers – significantly less than in the Hebrew word production tasks – picture naming (22% phonological errors, $\chi^2 = 6.3$, $p = .006$) and word reading (55% phonological errors, $\chi^2 = 45$, $p < .001$). It was also significantly less than his phonological error rate in naming 50 pictures in English (23% errors, $\chi^2 = 4.1$, $p = .02$). This finding is completely unexplainable under the frequency hypothesis. Although the frequency of English numbers in ZH's mental lexicon is presumably low, his phonological error rate in these words is lower than in Hebrew words. It is more similar to his phonological error rate in Hebrew number reading (2%).

To complete the picture, ZH’s semantic error rate in reading English numbers was 60%, i.e., he showed the full STEPS phenomenon in English. Clearly, the frequency hypothesis cannot explain this phenomenon either.

Another crucial finding that cannot be explained by the frequency hypothesis is the word role effect. The frequency hypothesis, as well as any hypothesis that assumes a single phonological representation of number and function words, cannot account for the finding that the exact same function words trigger different rates of phonological errors when produced in a list of single words or within sentences, and the same is true for number words.

Thus, existing findings in the literature and results from our patients, as well as conclusions drawn from cognitive models, indicate that the frequency hypothesis cannot explain STEPS.

4.4.2 Ruling out the number word frame generator hypothesis

Another possibility we explored concerns the number word frame generator – the component responsible for creating the number word frame, a pre-phonological verbal representation of the number (Cohen & Dehaene, 1991; McCloskey et al., 1986; McCloskey, 1992; Power & Dal Martello, 1990). This hypothesis assumes an architecture in which the output of the semantic lexicon follows two different routes for numbers and for words. For words, the semantic lexicon directly activates the phonological output lexicon, whereas for number words, it activates the number word frame generator, which in turn activates the phonological output lexicon. Although this specific pathway was not suggested by existing models of number processing, it is consistent with the model of McCloskey et al. (1986), and with minor modifications – also with the model suggested by Cohen and Dehaene (1991).

This hypothesis would explain STEPS by assuming a deficit in the output of the semantic lexicon. Such a deficit would impair the access to phonological output lexicon, yielding phonological errors in words. It will also impair the access to the number word frame generator, yielding semantic errors for number words.

However, this hypothesis has a major caveat when we come to examine it as an explanation for our patients' pattern of performance. It assumes an early deficit, in a position that is even prior to the phonological output lexicon. Thus, there is no way the hypothesis would be able to explain the error pattern for at least three of the patients in this study, who had a deficit in a later stage: SZ and GE were impaired only in the phonological output buffer, and YL was impaired in the phonological output lexicon itself rather than in the access to it from the semantic lexicon, as indicated by his reading pattern (see Section 3.2.3.1).

Furthermore, the hypothesis predicts that a deficit in the phonological output buffer would cause phonological errors in both words and numbers. Our findings clearly refute this prediction: none of the participants had phonological errors in number words, although they all had phonological output buffer deficits.

Moreover, the number word frame generator hypothesis cannot explain the STEPS phenomenon in function words, French/English letter names, and morphological affixes.

4.4.3 Ruling out the separate lexicon hypothesis

A third hypothesis that we considered for the reduced rate of phonological errors in number words is that there are two separate phonological output lexicons, one for words and the other for number words. According to such an account, individuals with impairment in the phonological lexicon of words, and without impairment in the number phonological lexicon, would make phonological errors in words but not in number words. This hypothesis also predicts the other direction of dissociation: individuals with impairment in the phonological lexicon of number words, with unimpaired word lexicon, would make phonological errors in number words but not in words. Such a patient, with phonological errors only in number words, has not been reported, as far as we know.

A major drawback of the separate lexicon hypothesis is that, like the frequency hypothesis, it can explain the reduction in phonological errors in number words, but it cannot explain the elevated number of semantic errors. Moreover, our findings contradict the separate lexicon hypothesis in several ways. First, this hypothesis, which focuses on the phonological lexicon, cannot explain why patients with a selective deficit in the phonological output buffer, like SZ and GE, do not make phonological errors in number words. Even if there are separate phonological output lexicons for numbers and for words, there is still a single phonological output buffer, and a deficit in this buffer should induce phonological errors in both words and numbers. Another reason to rule out this hypothesis is probabilistic: the participants in this study were selected based only on their performance in the production of non-number words, and yet none of them had phonological errors in numbers. It is possible to assume that coincidentally they all had an impaired phonological output lexicon and a spared numbers lexicon, but such an assumption does not seem likely. The separate lexicon hypothesis also cannot explain STEPS in function words, letter names, and morphemes. Even if the hypothesis is extended to cover these word categories, it still cannot explain the morphological errors in production of morphologically complex nonwords, because nonword production bypasses the lexicon. Finally, the separate lexicon hypothesis cannot explain the word role effect, namely, why number and function words are produced without phonological errors when they appear in their natural (syntactic/numeric) role, but are produced with phonological errors when they are deprived of their syntactic/numeric role.

The existence of two separate phonological output lexicons, one for words and one for numbers, was already suggested by Marangolo et al. (2005). According to them, the hypothesis is supported by the performance of the patient RA, who made semantic errors in numbers and no errors in content words. A similar error pattern was reported for patient FA (Marangolo et al., 2004). However, we believe that RA and FA do not make a good case for the separate lexicon hypothesis, because their errors in number words were semantic, whereas the separate lexicon hypothesis aims to explain phonological errors.[10] Indeed, we could extend the separate lexicon hypothesis to allow for semantic errors too, because a phonological output lexicon deficit might cause such errors (Caramazza & Hillis, 1990). However, even this extended hypothesis does not explain why a deficit in

[10] Interestingly, Marangolo et al.'s patient RA made errors not only on number words but also in function words and verbs. Although the type of errors was not detailed, this might resemble the errors in function words and morphemes in our patients.

the word lexicon causes phonological errors, whereas a deficit in the number lexicon causes semantic errors.

Marangolo et al. (2005) explicitly argued that the phonological output buffer cannot be responsible for RA’s errors because his buffer was intact, as proved by his spared repetition. However, this argument does not hold if number words and content words have slightly separate processing pathways within the phonological output buffer (as we suggest) and only the number processing pathway was impaired in RA’s case, or if RA's semantic errors in numbers originated in a processing stage earlier than the phonological output buffer.

Denes and Signorini (2001) also suggested that the phonological output lexicon stores number words separately from content words, because their patient AD made lexical errors in reading Arabic numbers and number words, but could read content words and repeat both words and numbers. Denes and Signorini concluded that AD had a selective deficit in the access from “central” transcoding mechanisms to a specific phonological storage of number words. However, while Denes and Signorini make a good argument for the existence of this separate phonological storage, their data does not clearly show whether this storage is in the phonological output lexicon or in the phonological output buffer. Thus, their study is in accord both with the separate lexicon hypothesis and with the view we suggest in the present study – that the phonological forms of number words are stored in the phonological output buffer.

In spite of all this, we do not completely dismiss the separate lexicon hypothesis. On the contrary – our findings support the major crux of this hypothesis: the assumption of a separate phonological storage for number words, function words, morphological affixes, and (English and French) letter names, which is separate from the phonological output lexicon of content words. However, this separate phonological storage is different from the phonological output lexicon in two respects: it includes pre-assembled units rather than unassembled phonological information, and it is functionally located in the phonological output buffer level rather than in an earlier stage. If the separate lexicon hypothesis is amended to acknowledge these two differences, it would be completely supported by our findings.

5 Discussion

5.1 Understanding STEPS

This research investigated the STEPS phenomenon, the *Stimulus Type Effect on Phonological and Semantic errors*. STEPS is the situation in which aphasic patients, who produce words with phonological errors, produce number words with semantic rather than phonological errors. The participants in this study were selected only on the basis of their phonological errors in non-number words, yet STEPS was observed for all of them. This indicates that STEPS, which was already mentioned in previous studies, is a robust phenomenon rather than an anecdotal or coincidental finding. We also showed that STEPS is not limited to Germanic and Romance languages but also extends to Hebrew, a Semitic language. An analysis of the participants' functional locus of deficit showed that the phonological output buffer was the only component in which all participants had a deficit, and for two of the participants (SZ and GE) this was the only impaired module. This indicates that the phonological output buffer is the functional component responsible for STEPS.

We showed that STEPS is not limited to numbers – there are several other word categories that are produced, like number words, with semantic rather than phonological errors: function words, morphological affixes, and English letter names. This finding suggests that all these word categories, including number words, are pre-assembled phonological units, which are stored in the phonological output buffer. A deficit in the phonological output buffer causes substitutions of one phonological unit by another, and therefore the resulting error is determined by the size of the phonological unit: In the case of monomorphemic content words, the unit is a phoneme, which is substituted by another phoneme, and the result is a phonological error. In the case of number words, function words, and English/French letter names, the substituted unit is a whole word, substituted by another whole word of the same type, and the result in a semantic error. Finally, morphological affixes are pre-assembled phonological units as well, and their substitution results in morphological errors, namely, the substitution of a morpheme with another existing morpheme, or the omission of a whole morpheme.

The specific word categories for which STEPS was found support the theoretical hypothesis suggested by Cohen et al. (1997): a word, or a phonological sequence, becomes a pre-assembled phonological unit when it serves as a building block in a specific productive process that creates a more complex construct: single number words are building blocks in creating multi-digit numbers, function words are building blocks because of their role in the syntactic process of sentence construction, morphological affixes are building blocks in creating morphologically complex words, and letter names are building blocks in the creation of acronyms. Thus, single phonemes, number words, function words, letter names, and morphological affixes – are all indeed “the building blocks of speech”.

We suggest that the phonological output buffer includes several phonological mini-stores, each containing the pre-assembled phonological forms of words belonging to one of the above categories. These mini-stores resemble the notion of a syllabary, the storage of pre-assembled syllables (Cholin & Levelt, 2009; Laganaro, 2005, 2008; Levelt et al., 1999), in the sense that they contain pre-assembled sequences of phonemes, ready for articulation. However, the syllabary is believed to belong to later stages, i.e., it is a stage subsequent to the phoneme-level representation in the buffer (Cholin, Levelt, & Schiller, 2006; Cholin & Levelt, 2009; Laganaro & Alario, 2006; Laganaro, 2008), whereas the mini-stores we suggest here are a part of the phonological output buffer, and the pre-assembled phonological unit *replaces* the segmented phoneme-level representation.

The results indicate that STEPS is a context-dependent phenomenon: semantic errors in number and function words occur mainly when the number or function word is used in the appropriate role – function words in a syntactic role (within sentences), and number words with numeric meaning. When the number or function words are deprived of their relevant role, they are produced with phonological errors, similarly to content words. This indicates that although number and function words are stored in pre-assembled format in the phonological output buffer, they are also stored in the phonological output lexicon in a decomposed manner, as separate phonemes, like content words. Interestingly, studies of acquired and developmental phonological impairment also noted the same dissociation: the two young women described by Temple and Marshall (1983) and Temple (1984) and the two patients described by Friedman (1996) also reported a word role effect for function words and morphological affixes in two patients with phonological impairment.

The phonological output buffer is commonly described as having two functions: phonological working memory and phonological composition (see Figure 1). In its role as a working memory component it holds phonological units active until their production. In its composition function it receives phonological information, phonemes and metrical information, from the phonological output lexicon, and assembles them into words and word sequences (a mechanism that has been termed "segment-to-slot insertion", cf. Nickels & Howard, 1999). This model is sufficient to account for the production of monomorphemic content words. However, the results of the present study suggest that the phonological output buffer has additional storage and assembly roles, which are necessary to account for the production of number and function words and of morphologically complex words. These roles are the mini-stores of pre-assembled phonological units, and morphological composition.

The mini-stores contain the pre-assembled number and function words and morphological affixes (and in some languages, also letter names). These units are activated by abstract identity information that arrives in the phonological output buffer.

When number words appear in the relevant role (with numeric meaning), the phonological output buffer receives their abstract identity, which activates the pre-assembled phonological unit in the relevant mini-store. These pre-assembled units of number words can be used to produce a multi-digit number.

Function words are produced similarly to number words: their pre-assembled phonological form is taken from the relevant mini-store. Namely, the phonological output buffer is aware of the special role of the function words, handles them as phonological building blocks, and can embed them in a sentence.

The production of morphologically complex words involves an additional stage – morphological assembly, i.e., combining the pre-assembled morphemes with the phonemes of the word’s stem. To produce the word "drinking", for example, the phonological output lexicon activates the phonemes of the word’s stem (/*d*/, /*r*/, /*i*/, /*n*/, /*k*/) in the phonological output buffer. The morphological affix (/*ing*/) is activated in the morpheme mini-store as a single unit. The phonological output buffer then assembles the phonemes of the word’s stem with the pre-assembled morpheme, to yield *drinking*. Bound function words are produced in a similar manner. Thus, whereas the phonological output buffer is sometimes described as responsible for phonological assembly of words, we suggest that it is also responsible for morphological assembly.

Morphologically complex nonwords (e.g., *zicklings*) are processed in a similar way: the buffer receives the stem of the nonword as separate phonemes (/*z*/, /*i*/, /*k*/, /*l*/). It also receives the abstract identities of the morphological affixes (/*ing*/ and /*s*/) and retrieves the pre-assembled phonological form of the morphemes from the morpheme mini-store. It then assembles everything into a single morphologically inflected nonword. The finding that STEPS applies to morphological affixes in nonwords as well as in words provides further support to the claim that the pre-assembled morphological affixes are stored in the phonological output buffer and not in an earlier lexical stage, because the buffer is the earliest speech production module that participates in the production of both words and nonwords. This finding is a solid indication that morphological structures can be represented independently of the lexical entry, and suggests that the stems and morphological templates of words are represented separately and handled by separate processes.

The phonological storage of pre-assembled units in the buffer has strict categorical organization, i.e., each of the phonological mini-stores is dedicated to one category of words or morphemes. This is evident from the finding that the participants never made between-category errors such as substitution of a number word for a function word or vice versa.

It is therefore our view that each mini-store receives the abstract identities of certain words or word parts (number words, morphological affixes, and function words), and provides their phonological content. We hypothesize that this information is received directly from the components that handle each of these word types: the abstract identity of morphological affixes and function words could arrive from morpho-syntactic processes. The abstract identity of number words could be received from a number-specific process that creates the sequence of abstract identities of number words to be uttered. However, these are merely hypotheses, and further research is needed to examine the exact relationship between the phonological output buffer and the earlier stages in the speech production system.

Figure 3 illustrates the extended model we suggest for the phonological output buffer, together with the hypothesized information flow from previous stages to the buffer.

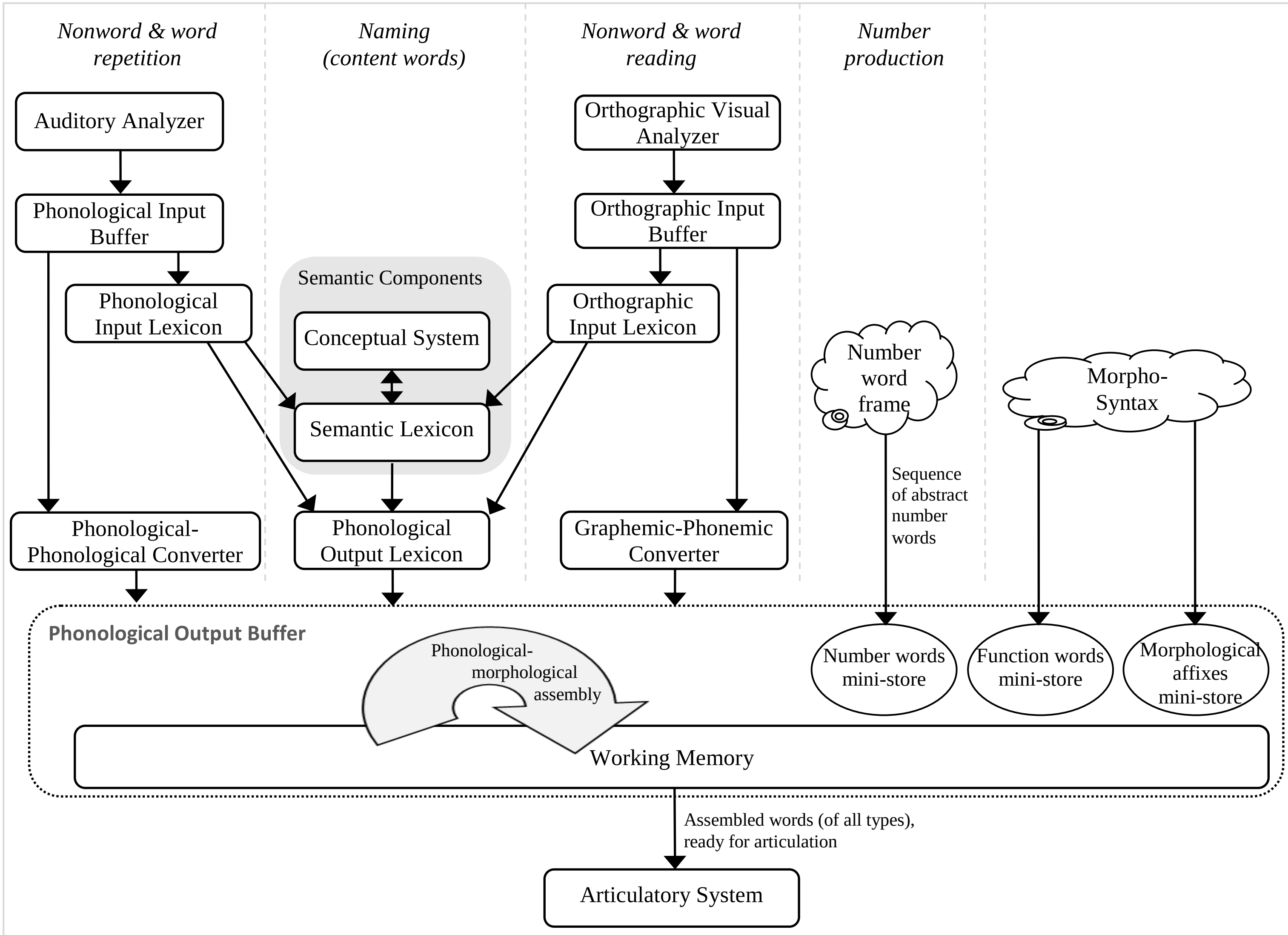


Figure 3. An extended model of speech production.

5.2 Morpho-syntactic processing in peripheral modules

The co-occurrence of phonological and morphological errors, which may be called "morphological STEPS", was already reported in previous studies (Badecker & Caramazza, 1991; Kohn & Melvold, 2000; Miceli, Capasso, & Caramazza, 2004; Miceli & Caramazza, 1988). A similar phenomenon is the production of neologisms with valid morphological affixes, Semenza, Butterworth, Panzeri, & Ferreri, 1990). Patterson (1982) even hypothesized that morphological affixes and function words are represented in the phonological system as whole units rather than segmented.

We are also not the first to have suggested that the phonological assembly of morphemes is done in the phonological output buffer: this was probably the opinion of Kohn and Melvold (2000), who said that morphological affixation is done at the "phonological planning" stage, and it is implied by the speech production model of Levelt et al. (1999). Job and Sartori (1984) claimed that the phonological representations of morphological affixes are separate from those of word stems, and that the morphological assembly occurs only in a late stage in speech production. Garrett (1975, 1980) claimed that morphological affixes are a part of the syntactic frame of the sentence, and are assembled with the content words only in a late stage in speech production, when the content words are inserted into this syntactic frame. Garrett's description of this stage as a limited-capacity memory store (Garrett, 1975, p. 166) may remind one of the phonological output buffer. Lavro et al. (2006) found morphological errors in patients with a phonological output buffer deficit, including the assembly of existing stem and affix into non-existent forms (e.g., *drinkly*, *walkedable*), and suggested that morphological affixes may be stored as pre-assembled phonological units and assembled with the word stems in the phonological output buffer. Finally, a slightly different opinion was presented in Cohen-Goldberg, Cholin, Miozzo, and Rapp (2013): like us they talk about morphological assembly of distinct phonological representations of stems and affixes, yet their findings suggest that this process belongs to a lexical rather than a post-lexical stage.

The notion that the phonological output buffer performs sentence-level operations (such as incorporating a function word into the word string) is also not revolutionary: recent syntactic theories also suggest that some word order variations occur at this phonological stage. For example, it was suggested that the movement of the verb to the second position of the sentence (before the subject) does not take place in narrow syntax, but rather in the phonological component (Chomsky, 1995, 2001; Zwart, 2001; see also Friedmann et al., 2013). This component, the phonological form (PF), is assumed to be the level of processing where the sentences are assigned with the phonological representation, after the construction of their syntactic structure (see Neeleman & Reinhart, 1998). It seems reasonable to assume that this word-order change occurs in the phonological output buffer, or at least that the buffer is responsible for holding the words in the correct order[11]. Thus, such a theory, much like the conclusions we reached in the present research, suggests that the phonological buffer performs sentence-level lexical operations

[11] One may assume that the PF corresponds with the phonological output lexicon too; but the lexicon is less good as a candidate for the word-order-change operation, because the lexicon operates in the scope of single words, not in the scope of whole sentences.

and operates on phonological units of various sizes – in this case, the words being re-ordered in the sentence. It may also be that the syntactic operations give specifications for certain features, which are then retrieved in the phonological output buffer as function words.

From a broader perspective, our findings agree with many studies that showed that morphological processing is not limited to central processes, but extends to peripheral modules as well. Several studies have reached such a conclusion with respect to peripheral modules of writing, and proposed a writing model in which the graphemic output buffer is aware of morphology (Badecker, Hillis, & Caramazza, 1990; Badecker, Rapp, & Caramazza, 1996; Yachini & Friedmann, 2008). According to these researchers, the orthographic output lexicon stores the written form of words separately from their morphological inflections, and the inflections are attached to the word's stem only in the next stage, the graphemic output buffer – very similarly to the model we proposed here. (Our proposal differs from Badecker et al.'s proposal in that they do not make the extra assumption of the morphological units being atomic building blocks). Other studies investigated reading and have found evidence for morphological processing in early peripheral stages of reading, the orthographic-visual analyzer (Beyersmann, Castles, & Coltheart, 2011; Friedmann & Gvion, 2012; Friedmann, Kerbel, & Shvimer, 2010; Longtin & Meunier, 2005; Rastle, Davis, & New, 2004; Rastle & Davis, 2008; Reznick & Friedmann, 2009; Sternberg & Friedmann, 2007; Velan & Frost, 2011).

STEPS and morphological processing are not the only two examples for high-level encoding in peripheral modules: early peripheral modules involved in reading Arabic numbers were shown to be sensitive to the digits 0 and 1, presumably because they reflect the morpho-syntactic structure of the number (Dehaene & Cohen, 1991; Dotan & Friedmann, 2009). Another example comes from working memory: verbal working memory is usually thought to rely on purely phonological encoding, but it seems that numbers are encoded in verbal WM also semantically and not only phonologically (Knops, Nuerk, Fimm, Vohn, & Willmes, 2006).

Interesting open questions relate to the extent and organization of the morpho-phonological information stored in the phonological output buffer. Does the buffer store pre-assembled phonological forms also for derivational morphemes, or only for inflections? How are irregularly-inflected words stored? Are the pre-assembled forms organized by morphological factors, e.g., prefixes vs. suffixes? An in-depth analysis of morphological errors made by patients on various types of morphologically complex word could be one of the methods to look into such questions.

5.3 Processing stages in number production

The model we described implies that number production involves at least two stages, as suggested by McCloskey and his colleagues (McCloskey, Sokol, Caramazza, & Goodman-Schulman, 1990; McCloskey et al., 1986; Sokol & Mccloskey, 1988). According to them, a syntactic stage creates a sequence of abstract identities of number words. Presumably, this sequence is created based on the number's syntactic features (its length and the positions of the digits 0 and 1), which define a *number word frame* (Cohen & Dehaene, 1991), in which the digit identities are embedded. A subsequent phonological stage retrieves the phonological representation of these number words.

Our data indicate that the phonological stage is implemented by the phonological output buffer. The notion of mini-stores in the phonological output buffer is also consistent with McCloskey et al.'s (1986) claim that number words such as "sixty" are morphologically pre-assembled rather than decomposed into /six/ and /-ty/, even if McCloskey et al. did not explicitly talk about number words being *phonologically* pre-assembled.

Our data also support the notion of an earlier, syntactic processing stage: according to the two-stage model, only a deficit in the syntactic stage would cause syntactic errors, namely, errors that completely change the number structure or create an invalid number syntax (e.g., *103* → *thirteen*; *70* → *seventy zero*). Indeed, in the number reading task, syntactic errors were relatively rare for the participants who had only a phonological output buffer deficit (SZ and GE), and were more common for the participants who also had an additional earlier deficit (see Section 4.1.1.4) [12].

The exact nature of the interface between the syntactic and phonological stages, namely, the nature of the abstract representation of number words that activate the number word mini-stores, was not investigated in this study. It seems plausible to rely on the model suggested by McCloskey et al. (1986), who suggested that the phonological form of number words is accessed with two parameters: the class of the number word (tens, teens, or ones) and the identity of the digit, i.e., its serial position within the class. Another question, which remains open for future research, is whether a selective deficit in the syntactic stage, with a spared phonological output buffer, can also cause single-word substitutions.

5.4 STEPS and other phenomena

Two previous studies have reported cases of "partial STEPS", with patients who had phonological errors in word production, and no errors at all in multi-digit number production (Bencini et al., 2011; Lochy et al., 2004; see Table 1). We can explain this as a situation in which the number words mini-store and the access to it were spared, leaving the number production pathway fully intact. The phonological errors of these patients may result from a deficit either in the phonological output lexicon or in the connection from it to the phonological output buffer. Because we assume that the activation to the mini-stores does not arrive from the phonological output lexicon, such deficits should not affect number production.

The situation of semantic errors in number production with spared production of content words may be explained by a selective deficit in the access to the mini-stores, or a deficit in earlier modules that are specific to number production (e.g., the syntactic module that creates the abstract identities of number words before sending them to the phonological output buffer). This may have been the case for patients FA and RA, reported by Marangolo et al. (2004, 2005).

---

[12] Class errors (e.g., saying 5 as *fifty*) may originate in a syntactic deficit, but, crucially, they could also originate from a phonological output buffer deficit, because producing "five" instead of "fifty" may be a substitution of one unit with another in the phonological output buffer. This is the reason for which the analysis of syntactic errors in section 4.1.1 followed the error classification method of researchers such as Michael McCloskey, who counted class errors separately from syntactic errors and lexical substitutions.

Two case studies reported semantic errors in number production accompanied by phonological errors in reading number words (patients BP and GS, Delazer & Bartha, 2001; Girelli & Delazer, 1999). This pattern may be explained if the word reading task did not activate the number-specific phonological stores. One possibility is that these individuals had dyslexia that caused sub-lexical reading (surface dyslexia), so the number words were treated as phonological sequences rather than as semantic entities. If this was indeed the case, the phonological errors in reading number words resemble the word role effect that we observed in the present study. The possibility of surface dyslexia is supported at least for one of these individuals: GS had both phonological and semantic errors in word production, which may suggest a phonological output lexicon deficit that caused sub-lexical reading (Gvion & Friedmann, 2012b).

The model we suggested predicts that we will not discover a "reversed STEPS" phenomenon, namely, a situation in which number words are produced with phonological errors but content words are produced with semantic errors. It is still possible for phonological errors to occur in number words that are presented without numeric meaning, or in function words without syntactic context. This may have been the case for patient GBC (Bencini et al., 2011), who made phonological errors in content words as well as in function words that were presented in a list (without syntactic context), but made no errors in numbers.

Another prediction of the model is that phonological errors in number and function words may occur in cases of a post-buffer deficit. Patients with apraxia of speech, whose deficit is in a stage later than the phonological output buffer, are expected to make phonological (or phonetic) errors also in number and function words. Indeed, we have recently met several apraxic patients whose pattern of errors may confirm the latter prediction: they make phonological errors in number words as well as in content words (Dotan, Friedmann, & Dehaene, 2014; Shalev, Ophir, Gvion, Gil, & Friedmann, 2014). Similarly, manipulations on normally-speaking participants could induce phonological errors in number/function words if they operate on post-buffer processes. This is one possible explanation for Dell’s (1990) findings, of similar rates of induced phonological errors in content and function words. Crucially, our model predicts that phonological errors in number or function words (in the relevant roles) can originate only in a post-buffer deficit or manipulation.

The model could also explain a bundle of interesting phenomena related with phonological dyslexia. Phonological dyslexia is defined as a selective deficit in the sub-lexical reading route (see Figure 1), which results in a specific difficulty in reading nonwords compared to existing words. The selective deficit in nonwords is explained by the fact that nonwords can be read only via the impaired sub-lexical route, whereas existing words are read primarily via the lexical route. Interestingly, many individuals with phonological dyslexia are reported to also make morphological errors in reading morphologically complex words and semantic errors in function words (Guggenheim & Friedmann, 2014; Gvion & Friedmann, 2010; Job & Sartori, 1984; Kendall, McNelil, & Small, 1998; Patterson, Suzuki, & Wydell, 1996; Patterson, 1982; Temple & Marshall, 1983; Temple, 1984, 1990). The account that we suggest here may explain this phenomenon. If we suggest that the deficit of these individuals in the sublexical route is at the phonological output buffer, this could explain their difficulty with nonword reading as well as their morphological and function word errors. Conversly, if a person with phonological dyslexia does not make morphological errors or errors in function words (and no phonological errors in word production and nonword repetition), his phonological output buffer is probably intact, and the deficit should be ascribed to the grapheme-to-phoneme conversion stage in the sublexical route. Indeed, several previous reports of individuals with phonological dyslexia seem to be in line with this hypothesis: individuals with phonological dyslexia who make semantic errors in function words and morphological errors often show some evidence for phonological impairment and difficulty in blending or nonword repetition (Friedman, 1996; Job & Sartori, 1984; Patterson et al., 1996), whereas others show no phonological deficit and no tendency for semantic errors in function words or for morphological errors (Cuetos, 1996; Vliet, Miozzo, & Stern, 2004).

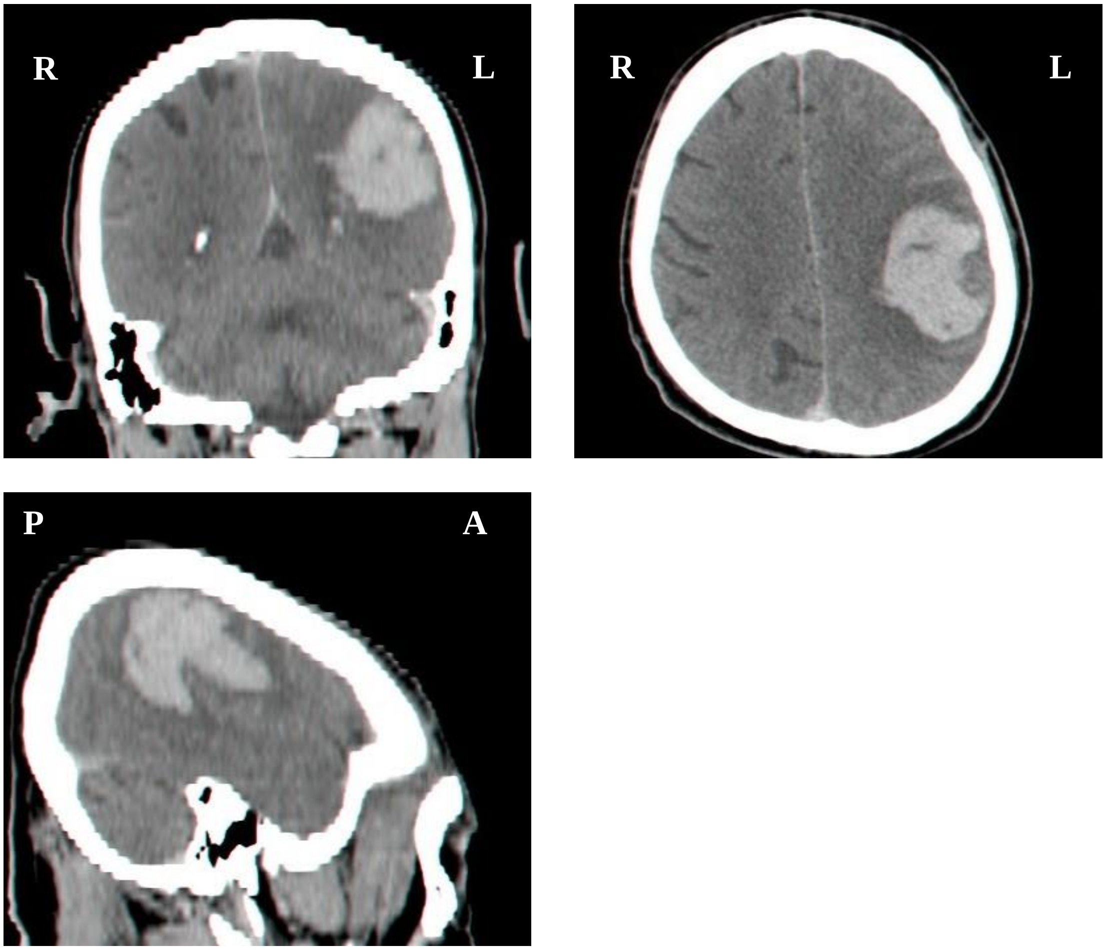
The model also makes an interesting prediction regarding the phonological flexibility of word forms: if the phonological form of certain words is pre-assembled, these words are less likely to be produced in several different phonological forms. Although we did not test this hypothesis directly, such a phenomenon in fact occurs for number words in French, as noted by Cohen et al. (1997): in many words in French, when the final letter is a consonant, it is not overtly produced. However, the final consonant is produced if the following word begins with a vowel (e.g., “gros”, fat, is produced as /gro/, but “gros homme”, fat man, is produced as /gro-zom/, because “h” is considered a vowel). This phenomenon is called *liaison*. Interestingly, if the second word (which begins with a vowel) is a number word, no liaison occurs and the last consonant of the first word remains unuttered (e.g., “les huit”, the eight, is produced as /le-wit/ and not as /le-zwit/). This lack of liaison in number words could be easily explained if liaison requires some phonological manipulation of the second word by the phonological output buffer, a manipulation that is impossible if the word is phonologically pre-assembled. A similar phenomenon occurs in Hebrew. Numbers between 11 and 19 are formed as construct state nominals, something like "the three of ten" for 13. In Hebrew, when a construct state nominal that includes content words is definite, it is always marked with the definite article "ha-" before the second word (Borer, 1999; Danon, 2008; Shlonsky, 2004; Siloni, 1997). However, because these numbers are pre-assembled, it is impossible to insert a definite article within them, and the “ha-“ article is produced before the whole CSN rather than before the second word.

Thus, it is possible that words with pre-assembled forms are not “phonologically flexible”, namely, they cannot be subject to phonological manipulations and therefore are unlikely to be produced in several different phonological forms. The opposite could also be true: it could be that pre-assembled phonological forms are created only for words that are not “phonologically flexible”, i.e., words that are always produced more or less in the same manner.

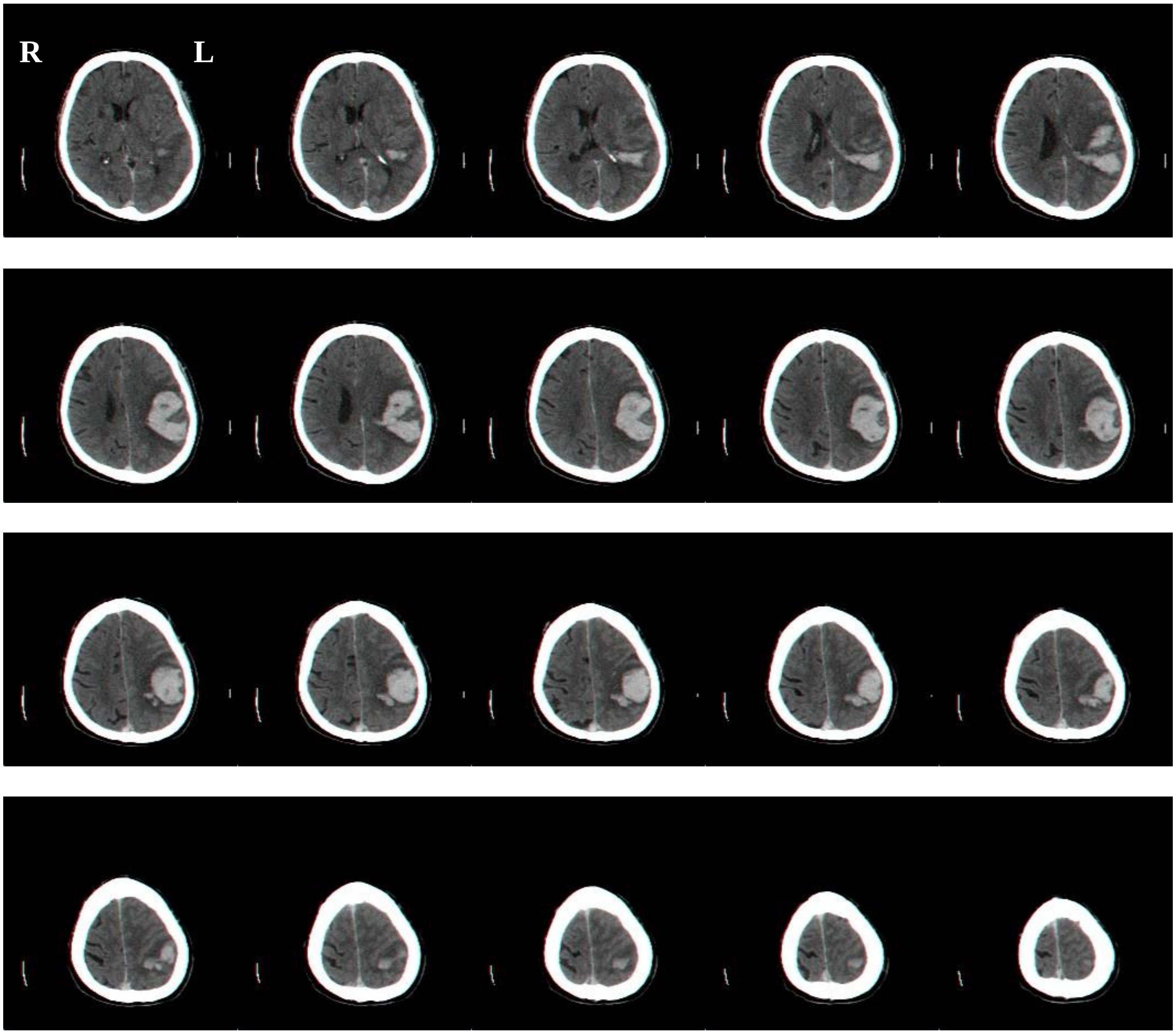
5.5 Epilogue

We believe that the findings we have presented here show that the phonological output buffer is a more complex module than was usually described. It is not only a buffer, because it has other important roles of composition and storage. It is phonological only in a broader sense, because it handles much more than just single phonemes.

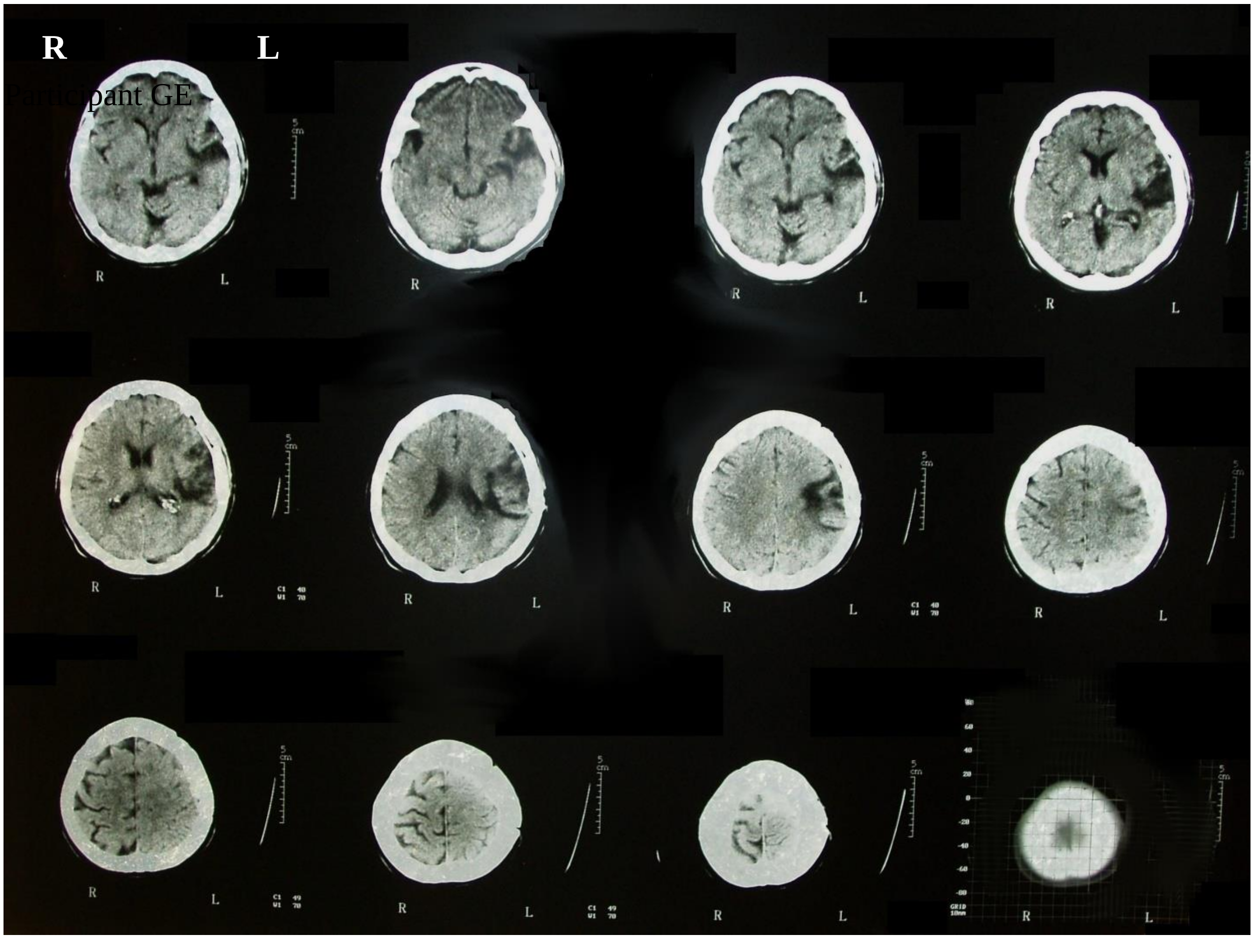
Appendix A. CT Images



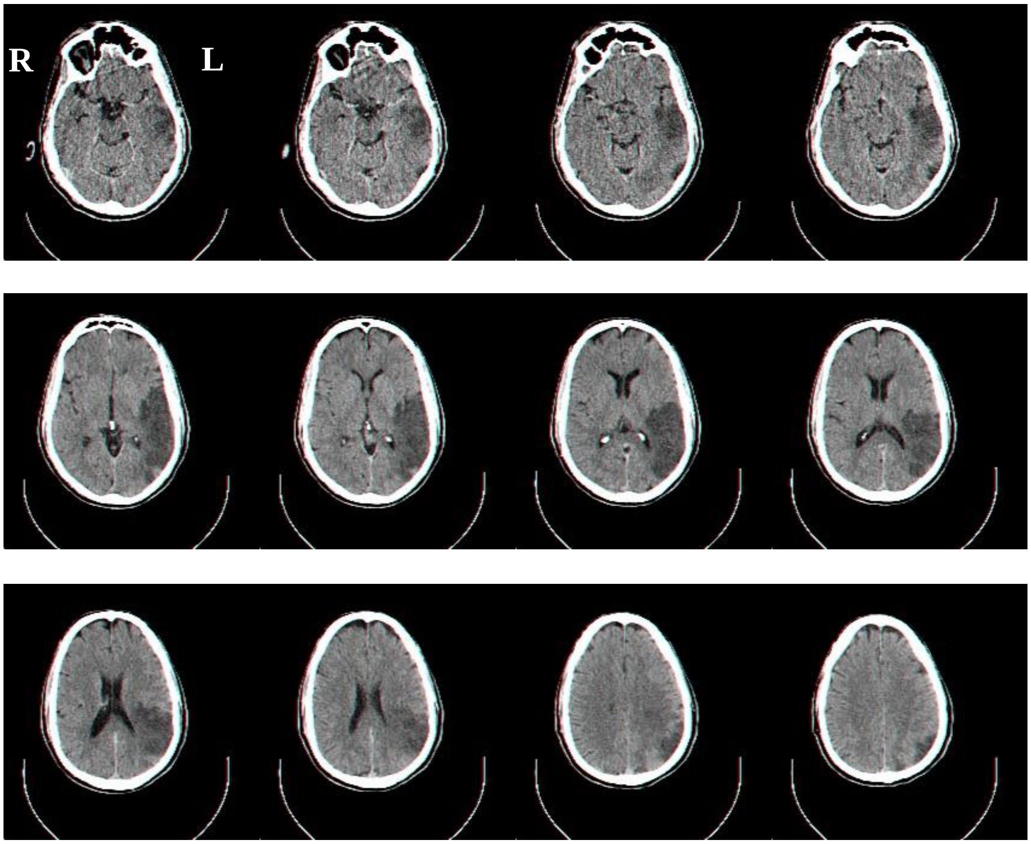
Participant SZ



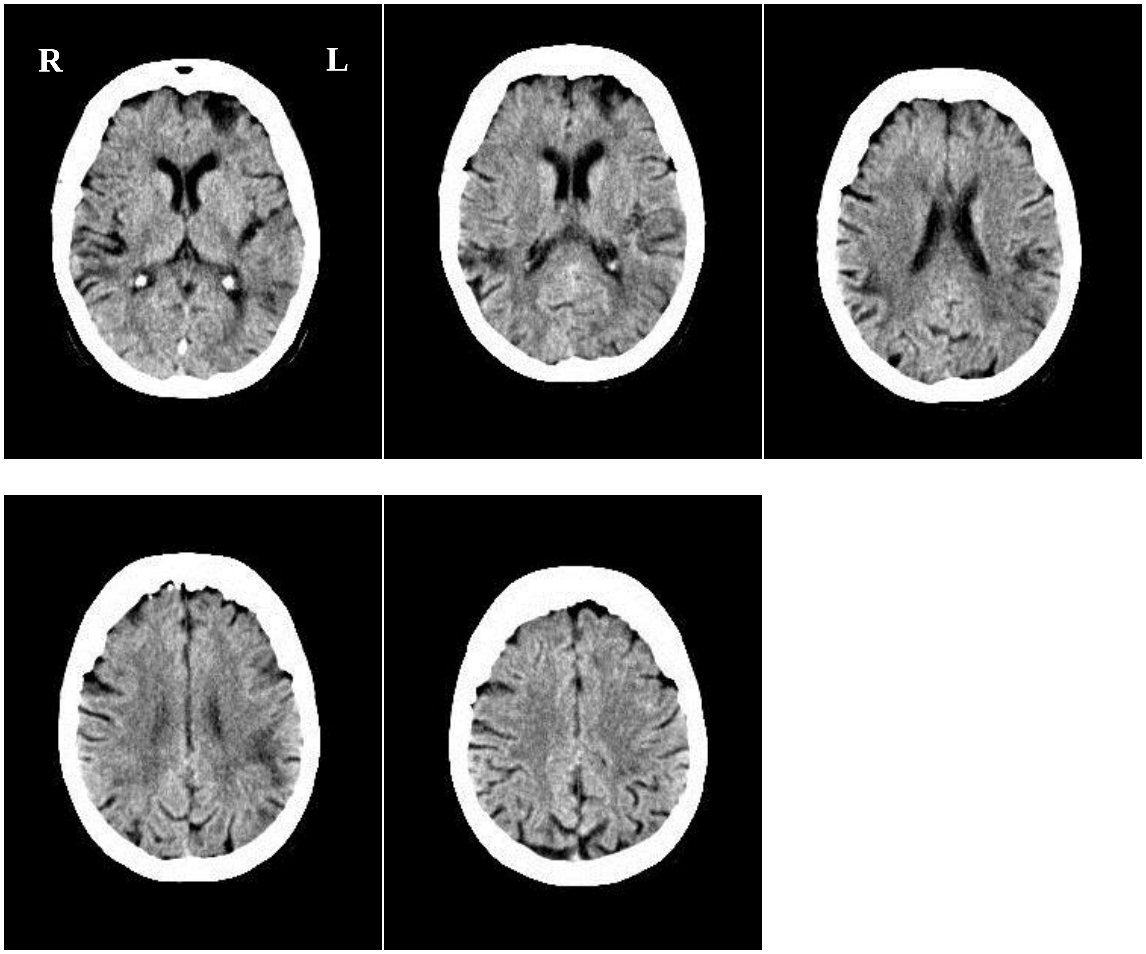
Participant SZ (continued)



Participant GE



Participant ZH



Participant RB

References

Bachoud-Lévi, A. C., & Dupoux, E. (2003). An influence of syntactic and semantic variables on word form retrieval. *Cognitive Neuropsychology*, *20*(2), 163–188. doi:10.1080/02643290242000907

Badecker, W. B., & Caramazza, A. (1991). Morphological composition in the lexical output system. *Cognitive Neuropsychology*, *8*(5), 335–367. doi:10.1080/02643299108253377

Badecker, W. B., Hillis, A. E., & Caramazza, A. (1990). Lexical morphology and its role in the writing process: Evidence from a case of acquired dysgraphia. *Cognition*, *35*(3), 205–243.

Badecker, W. B., Rapp, B., & Caramazza, A. (1996). Lexical morphology and the two orthographic routes. *Cognitive Neuropsychology*, *13*(2), 161–176. doi:10.1080/026432996381980

Bencini, G. M. L., Pozzan, L., Bertella, L., Mori, I., Pignatti, R., Ceriani, F., & Semenza, C. (2011). When two and too don't go together: A selective phonological deficit sparing number words. *Cortex*, *47*(9), 1052–1062. doi:10.1016/j.cortex.2011.03.013

Beyersmann, E., Castles, A., & Coltheart, M. (2011). Early morphological decomposition during visual word recognition: Evidence from masked transposed-letter priming. *Psychonomic Bulletin & Review*, *18*(5), 937–942. doi:10.3758/s13423-011-0120-y

Biran, M., & Friedmann, N. (2004). *SHEMESH: Naming a hundred objects*. Tel Aviv: Tel Aviv University.

Biran, M., & Friedmann, N. (2005). From phonological paraphasias to the structure of the phonological output lexicon. *Language and Cognitive Processes*, *20*(4), 589–616. doi:10.1080/01690960400005813

Biran, M., & Friedmann, N. (2007). *MA KASHUR: Picture associations test*. Tel Aviv: Tel Aviv University.

Biran, M., & Friedmann, N. (2012). The representation of lexical-syntactic information: Evidence from syntactic and lexical retrieval impairments in aphasia. *Cortex*, *48*(9), 1103–1127. doi:10.1016/j.cortex.2011.05.024

Bock, K., & Levelt, W. J. M. (1994). Language production. Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Borer, H. (1999). Deconstructing the construct. In K. Johnson & I. G. Roberts (Eds.), *Beyond Principles and Parameters* (pp. 43–89). Dordrecht: Kluwer. doi:10.1007/978-94-011-4822-1_3

Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language*, *8*(2), 133–161. doi:10.1016/0093-934X(79)90046-4

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition*, *42*(1-3), 261–286. doi:10.1016/0010-0277(92)90045-J

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208. doi:10.1080/026432997381664

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, *26*(1), 95–122.

Caramazza, A., Miceli, G., & Villa, G. (1986). The role of the (output) phonological buffer in reading, writing, and repetition. *Cognitive Neuropsychology*, *3*(1), 37–76. doi:10.1080/02643298608252669

Cholin, J., & Levelt, W. J. M. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes*, *24*(5), 662–684. doi:10.1080/01690960802348852

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, *99*(2), 205–235. doi:10.1016/j.cognition.2005.01.009

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, *124*(4), 375–390. doi:10.1037/0096-3445.124.4.375

Cohen, L., & Dehaene, S. (1991). Neglect dyslexia for numbers? A case report. *Cognitive Neuropsychology*, *8*(1), 39–58. doi:10.1080/02643299108253366

Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology*, *14*(7), 1029–1061. doi:10.1080/026432997381349

Cohen-Goldberg, A. M., Cholin, J., Miozzo, M., & Rapp, B. (2013). The interface between morphology and phonology: Exploring a morpho-phonological deficit in spoken production. *Cognition*, *127*(2), 270–286. doi:10.1016/j.cognition.2013.01.004

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi:10.1037/0033-295X.108.1.204

Coslett, H. (1991). Read but not write “idea”: Evidence for a third reading mechanism. *Brain and Language*, *40*(4), 425–443. doi:10.1016/0093-934X(91)90141-M

Cuetos, F. (1996). A case of phonological dyslexia in Spanish. *Cognitive Neuropsychology*, *13*(1), 1–24. doi:10.1080/026432996382042

Danon, G. (2008). Definiteness spreading in the Hebrew construct state. *Lingua*, *118*(7), 872–906. doi:10.1016/j.lingua.2007.05.012

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1-2), 1–42. doi:10.1016/0010-0277(92)90049-N

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia*, *29*(11), 1045–1074. doi:10.1016/0028-3932(91)90076-K

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Delazer, M., & Bartha, L. (2001). Transcoding and calculation in aphasia. *Aphasiology*, *15*(7), 649–679. doi:10.1080/02687040143000104

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. doi:10.1037/0033-295X.93.3.283

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*(2), 124–142. doi:10.1016/0749-596X(88)90070-8

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349. doi:10.1080/01690969008407066

Denes, G., & Signorini, M. (2001). Door but not four and 4: A category specific transcoding deficit in a pure acalculic patient. *Cortex*, *37*(2), 267–277.

Dotan, D., & Friedmann, N. (2009). Morpho-syntactic effects in the visual analysis of numbers. *Language and Brain*, *9*, 143–158 (in Hebrew).

Dotan, D., Friedmann, N., & Dehaene, S. (2014). Breaking down number syntax: Spared comprehension of multi-digit numbers in a patient with impaired digit-to-word conversion. *Cortex*, *59*, 62–73. doi:10.1016/j.cortex.2014.07.005

Ellis, A. W., & Young, A. W. (1988). *Human Cognitive Neuropsychology*. Hove, UK: Erlbaum.

Ellis, A. W., & Young, A. W. (1996). *Human cognitive neuropsychology: A textbook with readings*. Hove, East Sussex: Psychology Press.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. doi:10.1016/S0022-5371(73)80042-8

Franklin, S., Buerk, F., & Howard, D. (2002). Generalised improvement in speech production for a subject with reproduction conduction aphasia. *Aphasiology*, *16*(10-11), 1087–1114. doi:10.1080/02687030244000491

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 361–379. doi:10.1037/0096-1523.2.3.361

Friedman, R. B. (1996). Phonological Text Alexia: Poor Pseudoword Reading Plus Difficulty Reading Functors and Affixes in Text. *Cognitive Neuropsychology*, *13*(6), 869–885. doi:10.1080/026432996381845

Friedmann, N. (2003). *BLIP: Battery for assessment of phonological abilities*. Tel Aviv: Tel Aviv University.

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics* (pp. 350–374). Cambridge, UK: Cambridge University Press.

Friedmann, N., & Gvion, A. (2001). Letter position dyslexia. *Cognitive Neuropsychology*, *18*(8), 673–696. doi:10.1080/02643290143000051

Friedmann, N., & Gvion, A. (2002). *FriGvi: Friedmann Gvion battery for assessment of phonological Working Memory*. Tel Aviv: Tel Aviv University.

Friedmann, N., & Gvion, A. (2003). *TILTAN: A test battery for dyslexias*. Tel Aviv: Tel Aviv University.

Friedmann, N., & Gvion, A. (2012, August). *Morphological decomposition and letter position*. Presented at the LPD Day, McMahons Point, NSW, Australia.

Friedmann, N., Kerbel, N., & Shvimer, L. (2010). Developmental attentional dyslexia. *Cortex*, *46*(10), 1216–1237. doi:10.1016/j.cortex.2010.06.012

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, *44*(9), 1146–1160. doi:10.1016/j.cortex.2007.09.005

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology*, *74*(2), 159–180. doi:10.1111/j.2044-8295.1983.tb01851.x

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *Psychology of Learning and Motivation, 9* (pp. 133–177). New York, NY: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New Approaches to Language Mechanisms* (pp. 231–256). Amsterdam: North-Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production* (pp. 177–220). London: Academic Press.

Garrett, M. F. (1992). Disorders of lexical selection. *Cognition*, *42*(1-3), 143–180. doi:10.1016/0010-0277(92)90042-G

Girelli, L., & Delazer, M. (1999). Differential effects of verbal paraphasias on calculation. *Brain and Language*, *69*, 361–364.

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260. doi:10.1016/j.cognition.2005.12.010

Guggenheim, R., & Friedmann, N. (2014, July). *Previously undescribed type of developmental phonological dyslexia*. Presented at the INS mid-year meeting, Jerusalem, Israel.

Gvion, A., & Friedmann, N. (2001, July). *Surface dyslexia in a deep-orthography language*. Presented at the Israeli Association for Literacy 16th SCRIPT conference, Shfayim, Israel.

Gvion, A., & Friedmann, N. (2010). Letter position dysgraphia. *Cortex*, *46*(9), 1100–1113. doi:10.1016/j.cortex.2009.09.008

Gvion, A., & Friedmann, N. (2012a). Phonological short-term memory in conduction aphasia. *Aphasiology*, *26*(3-4), 579–614. doi:10.1080/02687038.2011.643759

Gvion, A., & Friedmann, N. (2012b, September). *The interaction between lexical retrieval and oral reading: Evidence from acquired and developmental anomia and surface dyslexia*. Presented in 13th Science of Aphasia Conference, Groningen, The Netherlands.

Hillis, A. E. (2001). The organization of the lexical system. In B. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology: What Deficits Reveal about the Human Mind* (pp. 185–210). New York, NY: Psychology Press.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*. (Macquarie Monographs in Cognitive Science). Philadelphia, PA: Psychology Press/Taylor and Francis.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. doi:10.1037/0278-7393.20.4.824

Job, R., & Sartori, G. (1984). Morphological decomposition: evidence from crossed phonological dyslexia. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *36*(3), 435–458. doi:10.1080/14640748408402171

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*(2), 185–209. doi:10.1016/0010-0277(83)90029-X

Kendall, D. L., McNelil, M. R., & Small, S. L. (1998). Rule-based treatment for acquired phonological dyslexia. *Aphasiology*, *12*(7-8), 587–600. doi:10.1080/02687039808249560

Knops, A., Nuerk, H. C., Fimm, B., Vohn, R., & Willmes, K. (2006). A special role for numbers in working memory? An fMRI study. *NeuroImage*, *29*(1), 1–14. doi:10.1016/j.neuroimage.2005.07.009

Kohn, S. E., & Melvold, J. (2000). Effects of morphological complexity on phonological output deficits in fluent and nonfluent aphasia. *Brain and Language*, *73*(3), 323–346. doi:10.1006/brln.2000.2179

Laganaro, M. (2005). Syllable frequency effect in speech production: Evidence from aphasia. *Journal of Neurolinguistics*, *18*(3), 221–235. doi:10.1016/j.jneuroling.2004.12.001

Laganaro, M. (2008). Is there a syllable frequency effect in aphasia or in apraxia of speech or both? *Aphasiology*, *22*(11), 1191–1200. doi:10.1080/02687030701820469

Laganaro, M., & Alario, F. X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, *55*(2), 178–196. doi:10.1016/j.jml.2006.05.001

Laganaro, M., & Zimmermann, C. (2010). Origin of phoneme substitution and phoneme movement errors in aphasia. *Language and Cognitive Processes*, *25*(1), 1–37. doi:10.1080/01690960902719259

Lavro, I., Ben-Or, G., & Friedmann, N. (2006). Inflectional morphology in individuals with conduction aphasia. *Language and Brain*, *5*, 46–50 (in Hebrew).

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1–22. doi:10.1016/0010-0277(92)90038-J

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*(5), 553–567. doi:10.1162/089892998562960

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75.

Linzen, T. (2009). *Corpus of blog postings collected from the Israblog website*. Tel Aviv: Tel Aviv University.

Lochy, A., Domahs, F., Bartha, L., & Delazer, M. (2004). Specific order impairment in arabic number writing: A case-study. *Cognitive Neuropsychology*, *21*(5), 555–575. doi:10.1080/02643290342000618

Longtin, C. M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, *53*(1), 26–41. doi:10.1016/j.jml.2005.02.008

Marangolo, P., Nasti, M., & Zorzi, M. (2004). Selective impairment for reading numbers and number words: A single case study. *Neuropsychologia*, *42*(8), 997–1006. doi:10.1016/j.neuropsychologia.2004.01.004

Marangolo, P., Piras, F., & Fias, W. (2005). “I can write seven but I can’t say it”: A case of domain-specific phonological output deficit for numbers. *Neuropsychologia*, *43*(8), 1177–1188. doi:10.1016/j.neuropsychologia.2004.11.001

Marshall, J. C. (1984). Toward a rational taxonomy of the developmental dyslexias. In R. N. Malatesha & H. A. Whitaker (Eds.), *Dyslexia: A global issue*. The Hauge: Martinus Nijhoff.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*(3), 175–199.

Martin, R. C., & Lesch, M. F. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 149–178). Hove, UK: Psychology Press.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1-2), 107–157. doi:10.1016/0010-0277(92)90052-J

McCloskey, M., Sokol, S. M., Caramazza, A., & Goodman-Schulman, R. A. (1990). Cognitive representations and processes in number production: Evidence from cases of acquired dyscalculia. In A. Caramazza (Ed.), *Cognitive neuropsychology and neurolinguistics: Advances in models of cognitive function and impairment* (pp. 1–32). Hillsdale, NJ, England: Erlbaum.

McCloskey, M., Sokol, S. M., & Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General*, *115*(4), 307–330. doi:10.1037/0096-3445.115.4.307

Messina, G., Denes, G., & Basso, A. (2009). Words and number words transcoding: A retrospective study on 57 aphasic subjects. *Journal of Neurolinguistics*, *22*(5), 486–494. doi:10.1016/j.jneuroling.2009.04.001

Miceli, G., Capasso, R., & Caramazza, A. (2004). The relationships between morphological and phonological errors in aphasic speech: Data from a word repetition task. *Neuropsychologia*, *42*(3), 273–287. doi:10.1016/j.neuropsychologia.2003.09.002

Miceli, G., & Caramazza, A. (1988). Dissociation of inflectional and derivational morphology. *Brain and Language*, *35*(1), 24–65. doi:10.1016/0093-934X(88)90100-9

Neeleman, A., & Reinhart, T. (1998). Scrambling and the PF interface. In M. Butt & W. Geuder (Eds.), *The Projection of Arguments* (pp. 309–353). Stanford: CSLI Publications.

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*. Hove: Psychology Press.

Nickels, L., & Howard, D. (1999). Effects of lexical stress on aphasic word production. *Clinical Linguistics & Phonetics*, *13*(4), 269–294. doi:10.1080/026992099299086

Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It’s the number of phonemes that counts. *Cognitive Neuropsychology*, *21*(1), 57–78. doi:10.1080/02643290342000122

Olson, A. C., Romani, C., & Halloran, L. (2007). Localizing the deficit in a case of jargon aphasia. *Cognitive Neuropsychology, 24*(2), 211–238. doi:10.1080/02643290601137017

Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research, 53*(1), 13–24. doi:10.1007/BF00867328

Patterson, K. (1982). The relation between reading and phonological coding: Further neurpsychological observations. In A. W. Ellis (Ed.), *Normality and Pathology in Cognitive Functions* (pp. 77–112). London: Academic Press.

Patterson, K., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface Dyslexia: Neurpsychological and Cognitive Studies of Phonological Reading* (pp. 335–360). Hove, UK: Erlbaum.

Patterson, K., & Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The Cognitive Neuropsychology of Language* (pp. 273–294). Hove, UK: Erlbaum.

Patterson, K., Suzuki, T., & Wydell, T. N. (1996). Interpreting a case of Japanese phonological alexia: The key is in phonology. *Cognitive Neuropsychology*, *13*(6), 803–822. doi:10.1080/026432996381818

Power, R. J. D., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*(3), 237–254. doi:10.1080/01690969008402106

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460–499. doi:10.1037//0033-295X.107.3.460

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7-8), 942–971. doi:10.1080/01690960802069730

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11*(6), 1090–1098. doi:10.3758/BF03196742

Reznick, J., & Friedmann, N. (2009). Preliminary morphological assessment in the early stages of visual-orthographic perception: Evidence from neglexia. *Language and Brain*, *8*, 31–61 (in Hebrew).

Rodriguez, J., & Laganaro, M. (2008). Sparing of country names in the context of phonological impairment. *Neuropsychologia, 46*(7), 2079–2085. doi:10.1016/j.neuropsychologia.2008.02.009

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142. doi:10.1016/0010-0277(92)90041-F

Romani, C. (1992). Are there distinct input and output buffers? Evidence from an aphasic patient with an impaired output buffer. *Language and Cognitive Processes*, *7*(2), 131–162. doi:10.1080/01690969208409382

Schocken, B. (2008). Characterization and evaluation of speech-reading support system for hard-of-hearing students in the class. In Y. Eshet-Alkalai, A. Caspi, & N. Geri (Eds.), *Learning in the technological era: Proceedings of the Chais conference on instructional technologies research*. Raanana, Israel: The Open University of Israel (in Hebrew).

Semenza, C., Bencini, G. M. L., Bertella, L., Mori, I., Pignatti, R., Ceriani, F., … Caldognetto, E. M. (2007). A dedicated neural mechanism for vowel selection: A case of relative vowel deficit sparing the number lexicon. *Neuropsychologia*, *45*(2), 425–430. doi:10.1016/j.neuropsychologia.2006.07.006

Semenza, C., Butterworth, B., Panzeri, M., & Ferreri, T. (1990). Word formation: New evidence from aphasia. *Neuropsychologia*, *28*(5), 499–502. doi:10.1016/0028-3932(90)90075-Y

Shalev, N., Ophir, E., Gvion, A., Gil, M., & Friedmann, N. (2014, February). *Dissociations between production processes of numbers and words*. Presented at the first Conference on Cognition Research of the Israel Society for Cognitive Psychology, Akko, Israel.

Shallice, T., Rumiati, R. I., & Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, *17*(6), 517–546. doi:10.1080/026432900050110638

Shlonsky, U. (2004). The form of Semitic noun phrases. *Lingua*, *114*(12), 1465–1526. doi:10.1016/j.lingua.2003.09.019

Siloni, T. (1997). *Noun Phrases and Nominalizations: The Syntax of DPs*. Dordrecht: Kluwer.

Sokol, S. M., & Mccloskey, M. (1988). Levels of representation in verbal number production. *Applied Psycholinguistics*, *9*(03), 267–281. doi:10.1017/S0142716400007840

Sternberg, T., & Friedmann, N. (2007). Developmental graphemic buffer dyslexia. *Language and Brain*, *6*, 91–96 (in Hebrew).

Temple, C. M. (1984). Developmental analogues to acquired phonological dyslexia. In R. N. Malatesha & H. A. Whitaker (Eds.), *Dyslexia: A Global Issue* (pp. 143–158). The Netherlands: Springer.

Temple, C. M. (1990). Foop is still floop: A six year follow-up of phonological dyslexia and dysgraphia. *Reading and Writing*, *2*(3), 209–221. doi:10.1007/BF00257972

Temple, C. M., & Marshall, J. C. (1983). A case study of developmental phonological dyslexia. *British Journal of Psychology*, *74*(4), 517–533. doi:10.1111/j.2044-8295.1983.tb01883.x

Velan, H., & Frost, R. (2011). Words with and without internal structure: What determines the nature of orthographic and morphological processing? *Cognition*, *118*(2), 141–156. doi:10.1016/j.cognition.2010.11.013

Vliet, E. C., Miozzo, M., & Stern, Y. (2004). Phonological dyslexia without phonological impairment? *Cognitive Neuropsychology*, *21*(8), 820–839. doi:10.1080/02643290342000465

Yachini, M., & Friedmann, N. (2008). Which is better – snacks or chocolate? On morphology in graphemic buffer dyslexia. *Language and Brain*, *7*, 149–160 (in Hebrew).

Zwart, C. J.-W. (2001). Syntactic and Phonological Verb Movement. *Syntax*, *4*(1), 34–62. doi:10.1111/1467-9612.00036