<RH-v> *Bilinguals' retention of phonemes and their serial order.*

<RH-r> *Kirsten Schraeyen, Wim Van der Elst, Astrid Geudens, Pol Ghesquière and Dominiek Sandra*

10 Tables

<CT>

Beyond global differences between monolingual and bilingual children on the nonword repetition task: retention skills for phonemes' identity and serial order.

<CA>

KIRSTEN SCHRAEYEN

Thomas More University College, Department of Speech Language Therapy and Audiology, Research group Language and Learning, Antwerp, Belgium

University of Leuven, Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, Leuven, Belgium

University of Antwerp, Department of Linguistics, Antwerp, Belgium

WIM VAN DER ELST

Hasselt University, Institute for Biostatistics and Statistical Bioinformatics, Hasselt, Belgium

ASTRID GEUDENS

Thomas More University College, Department of Education & Department of Speech language Therapy and Audiology, Research group Language and Reading Education, Antwerp, Belgium

POL GHESQUIÈRE

University of Leuven, Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, Leuven, Belgium

DOMINIEK SANDRA

University of Antwerp, Department of Linguistics, Antwerp, Belgium

<DATES>

(Received: September 11, 2016; final revision received: April 21, 2017; accepted: April 25, 2017)

<ADDRESS>

Address for correspondence:

Kirsten Schraeyen

Molenstraat 8

2018 Antwerp

Belgium

Kirsten.schraeyen@thomasmore.be

<Abstract>

This study compared NRT-performance in monolingual Dutch and bilingual Turkish–Dutch third-graders using a Dutch Nonword Repetition Task (NRT). Several novel response analyses at the phoneme level were applied to further understand the earlier reported overall accuracy differences in NRT-performance between bilinguals and monolinguals. Analyses in which the retention of phonemes and the retention of their serial order were disentangled revealed that monolinguals outperform bilinguals with respect to the retention of the phonemes themselves. However, both groups did not differ in their retention of the serial order of correctly recalled phonemes. Furthermore, this study confirms that expressive vocabulary skills do affect overall NRT-performance. The results are discussed in light of current short-term memory (STM) models and the role of long-term phonological knowledge in NRT tasks.

<Keywords> nonword repetition, bilingualism, serial order, phonological knowledge, short-term memory

<HA>Introduction

In recent years, nonword repetition performance has received a great deal of attention, both in language and literacy research. In a nonword repetition task (NRT), the participant first listens to a nonword¹, then temporarily stores the novel phonological information in short-term memory, and finally reproduces it. Although there is no consensus yet on what the NRT actually measures (for a review, see Coady & Evans, 2008), this task has been explicitly used in language and literacy research as a measure of the storage, coding, and retrieval of

phonological information, i.e., phonological representations in short-term memory (STM). It has been documented that there is a positive correlation between NRT performance and vocabulary knowledge (Gathercole, 2006; Munson, Kurtz & Windsor, 2005) and literacy level (Conti-Ramsden & Durkin, 2007).

A considerable amount of literature has been published on the association between NRT performance and lexical (i.e., long-term) knowledge in typically developing and language impaired MONOLINGUAL CHILDREN (Edwards, Beckman & Munson, 2004; Gathercole, 2006; Thordardottir, Kehayia, Lessard, Sutton & Trudeau, 2010). Different types of studies demonstrated that effective STM encoding of nonwords relies on long-term knowledge of lexical properties, and thus, makes active use of many cues related to both lexical-semantic features (i.e., meaning) and sublexical phonological/phonotactic features (i.e., form) of words. For instance, studies have shown that words are recalled significantly better than nonwords. This semantic advantage in STM has been called the lexicality effect (Gathercole, Pickering, Hall & Peaker, 2001; Hulme, Maughan & Brown, 1991). Additionally, it has been demonstrated that structural knowledge of the words in a language (restrictions on possible word forms, such as phonotactic regularities) also facilitates immediate recall. This is known as the word-likeness effect, i.e., the finding that nonwords with high phonotactic probability are recalled better than nonwords with low phonotactic probability (Edwards et al., 2004).

Another indication of an association between long-term linguistic knowledge and the efficacy with which STM encodes phonological information comes from studies in BILINGUAL POPULATIONS (e.g., Thorn & Gathercole, 1999; Cheung, 1996; Yoo & Kaushanskaya, 2012). Several studies have shown that bilingual children perform better on a NRT task in their

native language compared to one in their second language, in which they do not have the same level of proficiency (Gathercole & Baddeley, 1990; Thorn & Gathercole, 1999; Cheung, 1996; Masoura & Gathercole, 1999). In line with this finding, bilingual children perform worse on a NRT based on their second language than monolinguals who speak this language as their first language (Windsor, Kohnert, Lobitz & Pham, 2010; Messer, Leseman, Boom & Mayo, 2010).

These conclusions were mainly based on the overall accuracy scores in NRT tasks, i.e., scores based on a dichotomy between fully correct vs. incorrect repetitions of nonwords (see for instance Lee & Gorman, 2012; Messer et al., 2010; Summers, Bohman, Gillam, Pena & Bedore, 2010; Thordardottir & Brandeker, 2013). In such studies the unit of analysis is the number of nonwords that a participant recalls correctly. Even though a lot has been learnt by using this metric, analyses of this type can only reveal THAT there is a difference between two groups, not WHY there is a difference.

A first step towards addressing the latter question has been made in some bilingual studies, which also included (more) detailed analyses by relying on proportional scores (see for instance Gutiérrez-Clellen & Simon-Cereijido, 2010). These scores were obtained by dividing the number of phonemes repeated correctly by the number of input phonemes. This type of scoring is more fine-grained than the use of overall accuracy scores, making it possible to focus on types of errors such as phoneme omissions, additions and deletions (see also Sorenson Duncan & Paradis (2016) which, for instance, considered onsets versus codas and Topbaş, Kaçar-Kütükçü & Kopkalli-Yavuz (2014), who considered consonants versus vowels). By comparing the proportion of correctly recalled phonemes one can find out whether the group difference at the level of the overall scores is matched by a difference in

the proportion of correctly recalled phonemes. If so, this would suggest that differences in phoneme retention lie at the basis of the group effect in overall NRT performance.

The Gutiérrez-Clellen and Simon-Cereijido study demonstrates that, in order to better understand the cause of inferior performance on a NRT, it is important to perform more fine-grained analyses of the responses, i.e., to explore the sublexical processing of nonwords in more detail by focusing at the phoneme level. However, even though proportional scores shed more light on the question of why two groups differ in their NRT performance, there are two reasons why an even more sensitive measure is desirable. First, the proportional scores mentioned above only reflect participants' skill in remembering the identity of phonemes. Other phoneme properties (see below) are not captured. Second, by analyzing the average proportion of correctly recalled phonemes (aggregation across items) a lot of information is lost: more particularly, how systematically participants of both groups differ with respect to their recall of phoneme identities.

A more systematic study of nonword repetition performance is to make a task analysis. In which processes do participants have to engage when repeating nonwords, i.e., which types of information do they have to encode and retrieve? In order to correctly reproduce a nonword, participants must have a correct retention of (1) all phonemes appearing in the target nonword, i.e., the phonological content of the nonword (henceforth: 'item information') and (2) the serial position associated with each phoneme, i.e., the serial order of the phonemes in the nonword (henceforth: 'serial order information').

Interestingly, this distinction between processing item information and serial order information has already received some attention in current models of STM, based on memory span tasks in which a series of stimuli (i.e., words, letters, and digits) have to be recalled (for a

review, see Majerus & Cowan, 2016). Most of this research has been carried out in a monolingual context (Burgess & Hitch, 1999, 2006; Page & Norris, 1998; Gupta, 2003) with the exception of one bilingual study with adults (Majerus, Poncelet, Van der Linden & Weekes, 2008). Although there is some controversy about the exact way in which serial order information is represented (Hitch, Fastame & Flude, 2005; Majerus et al., 2008), there seems to be a consensus that a specialized STM system is dedicated to the representation of the serial order between items, whereas representation of item information depends directly on the activation of the language system (i.e., long-term memory knowledge).

Evidence for the distinction between these two types of information comes, for instance, from studies indicating an increased performance in retrieving item information (fewer item errors) but not in retrieving the items' serial order. This was observed in tasks revealing higher recall of semantically related vs. unrelated words, i.e., the semantic similarity effect, or higher recall of high vs. low frequency words, i.e., the word frequency effect (Poirier & Saint-Aubin, 1996; Nairne & Kelley, 2004). Research from Gupta and colleagues suggests the existence of an ordering mechanism in STM (Gupta, 2003, 2009; Gupta, Lipinski, Abbs & Lin, 2005). Evidence for this idea was furnished by their finding of syllable primacy and recency effects in a NRT task (Gupta, 2003; Gupta et al., 2005). Syllable primacy effects refer to the finding that first syllables in polysyllabic words are better recalled than syllables in mid-word positions, whereas syllable recency effects refer to the finding that final syllables are better recalled than syllables in mid-word positions. Based on their results, and the fact that primacy and recency effects are the hallmark characteristics of immediate serial recall, Gupta et al. (2005) postulated the existence of a "sequence memory" mechanism

that underlies performance in STM tasks. They argue that this mechanism plays an important role in temporarily maintaining and repeating syllable sequences that form a nonword.

Where does this leave us with respect to understanding NRT performance in bilinguals? The aforementioned effects suggest the importance of distinguishing between item and serial order retention. However, as explained above, the focus in NRT experiments on participants' number of correct responses or on their average proportion of correctly recalled phonemes makes it impossible to distinguish between the retention of a nonword's phonemes and the retention of their serial order. Indeed, item and serial order information are by definition collapsed in correct responses, as these are instances where all phonemes are recalled in their correct serial order. Analyzing the proportion of correctly recalled phonemes is already a first step in the direction of an error analysis, but is restricted to participants' ability to recall phonemes' identity, not their serial position. As nonword repetition requires the retention of both item and serial order information we need a way to distinguish between errors that result from a failure to recall all phonemes in a nonword and errors that result from a failure to recall phonemes' serial position (some errors will probably reflect both types of failure). The explicit aim of the current study is to disentangle participants' retention of the phonemes themselves (item STM processing) from their retention of the phonemes' serial order (serial order STM processing).

Recently, Schraeyen, Geudens, Ghesquière, Van der Elst and Sandra (in press) conducted a NRT study in which they compared more skilled and less skilled young monolingual readers. They analyzed both nonword level performance (i.e., number of correct nonwords responses) and phoneme level performance (i.e., an analysis of participants' recall performance at the sub-item level). More precisely, they distinguished between phoneme

identity performance, defined as the number of responses without phoneme identity errors (but with possible serial order errors), and serial order performance, defined as the number of responses without serial order errors (but with possible phoneme identity errors). Literacy skill was not linked to the ability to retain phonemes (item information). However, literacy skill was linked to the retention of phonemes' serial order, higher serial order scores being significantly associated with higher literacy scores).

In the current study we will use the same technique to compare NRT performance in bilinguals and monolinguals. We want to assess whether participants' language group (monolinguals vs. bilinguals) predicts their phoneme identity and serial order performance when repeating nonwords. In order to accomplish this goal we will first investigate whether we can detect an effect of Language Group on the retention of nonwords, as has been shown in many previous studies (e.g., Lee & Gorman, 2012). Next, we will investigate whether Language group has an effect on the retention of the identity of phonemes and/or on the retention of their serial order. Finally, we will analyze whether the two groups differ with respect to the distribution of the different types of error responses (only phoneme identity errors, only serial order errors, both error types combined).

Our major objective is to find out whether the difference between bilinguals and monolinguals will be reflected in different NRT performance patterns at the phoneme level, more particularly, in their retention of phonemes' identity and their serial order. Given the idea that item information depends on the activation of the language system (cf. Majerus et al., 2008), it is reasonable to hypothesize that participants with a better knowledge of Dutch vocabulary (monolinguals) will have a better retention of phonemes compared to participants with a smaller vocabulary in Dutch (bilinguals). At the same time, we want to know whether

this expected difference will hold for the retention of phonemes' serial order as well. If retention of serial order is related to a specialized system in STM, as suggested by previous literacy research (e.g., Majerus et al., 2008), one would expect a negative answer. There is no reason to assume that typically developing monolinguals and bilinguals will differ in the way this STM mechanism operates.

However, the prediction with respect to serial order retention is complicated by the fact that the retention of an unfamiliar phoneme sequence (a nonword) is not quite identical to the retention of a sequentially ordered set of random items. For the purposes of coding and retention of phonemes' serial order, linguistic LTM knowledge may also play a role: more particularly, knowledge of the combinatorial possibilities of the phonemes in the language (so-called phonotactic restrictions). Knowledge of these (language-specific) phonotactic restrictions might be beneficial for encoding and retaining phonemes' serial order. As this LTM knowledge is likely to be better developed in monolingual children, these are likely to outperform bilingual children on the retention of phonemes' serial order as well.

Participants in our monolingual group speak Dutch as their native language. Children in the bilingual group speak Turkish as their native language and acquired Dutch as a second language. Even though the role of phonotactic knowledge must be considered when two such different languages are involved, this problem was ruled out by the specific stimulus set for the NRT. Despite some differences between the phoneme inventories of Turkish and Dutch, especially in the vowel set, our nonwords, constructed according to the phonological rules in Dutch, happened to respect the phonotactic restrictions in the bilinguals' native language (see *Materials and design* section). For instance, in Turkish, consonant clusters, a hallmark of phonotactics, rarely consist of more than two consonants. In addition, they never occur at the

onset of native words (Topbaş & Kopkalli-Yavuz, 2008). None of these phonotactic properties of the bilinguals' native language were violated in our item set. Hence, we do not expect a difference between the monolingual and bilingual groups with respect to serial order recall, as the participant groups are not expected to differ with respect to their STM sequencing memory (the literacy development of all children being normal) and the items' phonotactic constraints are not expected to be more difficult for the bilingual group.

A second objective of our study is to evaluate the impact of prior language knowledge – more particularly, expressive vocabulary knowledge – to the overall performance on a NRT, i.e., the number of correct responses. If pre-existing language knowledge (LTM knowledge) is critical to overall NRT performance, participants with better expressive vocabulary knowledge in Dutch (monolingual group) should outperform participants with a smaller vocabulary size in Dutch (bilingual group).

<HA>Method

<HB>Participants

Monolingual and bilingual participants were recruited in two primary schools located in the central region of Belgium². Eighteen Turkish–Dutch bilinguals participated in the study (mean age: 8 years 11 months, SD = 5 months, range 8 years 3 months – 9 years 6 months). All participants spoke Turkish as L1 and had acquired Dutch as L2 during their early school years, with a mean age of acquisition of 3 years 8 months (SD = 7 months, range 2 years 8 months – 4 years 2 months). All bilingual participants acquired their home language first and were exposed to Dutch upon entering Dutch day care, preschool, or kindergarten settings. At

the time of testing, they communicated primarily in Dutch with siblings and peers at school and used their native language in their home environment. Twenty Dutch-speaking monolinguals participated in this study (mean age: 8 years 9 months, SD = 3 months, range 8 years 4 months–9 years 3 months). The monolingual participants had not been exposed to any other foreign language, except Dutch. The groups did not differ in age ($p = .11$). All children were in third grade and developed normally according to parent reports and teacher interviews and reports. They had no speech, language, or hearing problems, or other serious concerns in any developmental area.

<HB> *Materials and design*

Children were tested individually by a native speaker of Dutch. They had to perform a vocabulary test and a NRT.

<HC> *Vocabulary test*

The Dutch standardized version of the CELF-4-NL (Semel, Wiig & Secord, 2008) was administered. The subtest ‘Expressive vocabulary’ was used to evaluate the participants’ productive lexical knowledge. This task requires participants to name an object, person, or activity portrayed in an illustration.

<HC> *Nonword Repetition Task (NRT)*

We used the Flemish version of the NRT (Boets, 2006). This NRT task consists of 36 nonwords, complying with Dutch phonotactics and varying in length from three to five syllables. As mentioned in the introduction all items respected the phonotactic properties of

Turkish as well. Stimuli are listed in Appendix A. All test items were prerecorded on a CD by a female native speaker of Dutch and presented once to the participant through headphones. Responses were recorded via a head-mounted microphone-for subsequent offline scoring of repetition accuracy. Corrective feedback was provided for two practice items. The child was then presented with the 36 test items and asked to repeat the presented nonword. The test items were blocked on the basis of their syllable length, starting with three-syllable items and systematically moving up to five-syllable items.

Each nonword repetition response was phonetically transcribed by the experimenter using SAMPA (Speech Assessment Methods Phonetic Alphabet) and scored with respect to retention at the nonword level (global analysis) and the phoneme level (in-depth analysis). For the purpose of reliability checking, all responses were transcribed and scored by a second judge, who was blind to the original transcriptions and scoring. For the transcription, an inter-judge agreement of 92.3% was found, based on the ratio of the number of identically transcribed phonemes on the total number of transcribed phonemes. For the scoring, there was also high agreement between the scores at each linguistic level, i.e., 98.6% at the nonword level-and 93.9 % at the phoneme level. In case of disagreement, the experimenter and second judge decided together how the responses were scored.

<HB> *Scoring technique*

In a first set of analyses, we evaluated each response in three different ways. First, we checked whether the item was correctly recalled or not (nonword level score). Second, we scored the response with respect to the two phoneme related-variables: (1) phoneme identity and (2) serial order (phoneme level scores). Hence, we used the following scoring scheme:

<HC> *Nonword level score*

This score reflects the overall correctness of the response, i.e., nonword level performance. A response was counted as correct and given a binary score of 1 if the entire nonword was correctly recalled, i.e., no errors were made at the stress level, the syllable level, or the phoneme level. For example, for the target nonword /*pig.dúl.mek*/ the response /*pig.dúl.mek*/, would receive a score of 1. At this level, any other response was scored as 0, i.e., an error response.

<HC> *Phoneme level scores*

To evaluate children's ability to retain phonemes' IDENTITY and SERIAL ORDER when repeating a nonword, each response was given a binary score on these two phoneme-related variables. As mentioned earlier nonword repetition requires the recall of both the phonemes themselves and their serial order. Given our binary coding of each response on these two variables four different response types can be distinguished: (a) a correctly repeated nonword was given a binary score of 1 on both phoneme-related variables; (b) an incorrect response containing only one or more phoneme identity errors was given a binary score of 0 on phoneme identity retention and a score of 1 on serial order retention; (c) an incorrect response containing only one or more serial order errors was given a binary score of 0 on serial order retention and a score of 1 on phoneme identity retention; (d) an incorrect response containing both phoneme identity and serial order errors was given a binary score of 0 on both phoneme-related variables.

Our classification of responses in these four categories was based on the following criteria:

(1) NO IDENTITY OR SERIAL ORDER ERRORS (correct responses): all target phonemes appeared in the same serial order in the response, i.e., target: */pig.dúl.mek/*, response: */pig.dúl.mek/*.

(2) PHONEME IDENTITY ERRORS. (a) PHONEME ADDITION: one or more phonemes were added in the response compared to the original nonword. For example, target: */pig.dúl.mek/*, response: */pig.dúls.mek/*. (b) PHONEME SUBSTITUTION: one or more phonemes in the target were replaced by other phonemes in the response, e.g., target: */pig.dúl.mek/*, response: */pig.dól.mek/*. Here, the target phoneme */u/* was replaced by the phoneme */o/*. (c) PHONEME OMISSION: one or more phonemes in the target were omitted in the response. For example, target: */pig.dúl.mek/*, response: */pig.dú.mek/*, the phoneme */l/* was omitted in the response. (c) PHONEME SUBSTITUTION IN COMBINATION WITH PHONEME OMISSION, for example, target: */pig.dúl.mek/*, response: */pis.dú.mek/*. In the response, the target phoneme */g/* was replaced by the phoneme */s/* whereas the target phoneme */l/* was omitted.

Phoneme identity errors do not by themselves disturb the serial order of the correctly recalled target phonemes, i.e., their relative position in the nonword stimulus is preserved (for the distinction between relative position vs. absolute position, see further). They can obviously co-occur with serial order errors (see below).

(3) SERIAL ORDER ERRORS. Note that it is obvious, both from a logical and from an analytical perspective, that the retention of a phoneme's serial order can only be determined if the phoneme itself has been correctly recalled. Indeed, it makes no sense to attribute positional properties to phonemes that are absent in the output (omissions) or did not appear

in the input (additions, substitutions). Moreover, if each phoneme identity error also counted as a serial order error, the former errors would be a subset of the latter (potentially a large subset, depending on the number of responses with serial order errors and no phoneme identity errors). This would be at odds with our intention to separate these two retention variables.

To determine whether the serial order of the correctly recalled phonemes was respected, we used the procedure suggested by McKelvie (1987). Using this approach, input phoneme strings of at least 2 phonemes that appear in the response are considered correct irrespective of their absolute position when counting from the beginning or the end of the nonword sequence. Next, all remaining single phonemes that appear in the target are handled. The serial position of a single phoneme is correct if it appears in the correct serial order when counting either from the beginning or the end of the target nonword, but not both. For example, given the nonword */pig.dúl.mek/* and the response */pil.dúg.mek/*, the response would be checked as follows. With respect to phoneme identity all 9 phonemes are correctly identified. With respect to serial order the phonemes */pi/*, */du/*, */mek/* are correctly ordered when using the McKelvie method. However, the single phonemes */g/* and */l/* are swapped between the third to sixth positions, violating the rule that single phonemes have to be in the correct absolute serial order when counting either from the beginning or the end of the target nonword. Such PHONEME SHIFTS were the only error type containing a serial order error without an identity error: each target phoneme appeared in the response, but some phonemes were recalled at another (i.e., incorrect) serial position,

(4) COMBINATION OF IDENTITY AND SERIAL ORDER ERRORS. (a) PHONEME SHIFT IN COMBINATION WITH PHONEME SUBSTITUTION, e.g., target: */pig.dúl.mek/*, response:

/pil.dúg.nek/, target phonemes /g/ and /l/ are swapped (giving rise to a serial order error) whereas target phoneme /m/ is replaced by /n/ (giving rise to an identity error). (b) PHONEME SHIFT IN COMBINATION WITH PHONEME OMISSION, e.g., target: */pig.dúl.mek/*, response: */pug.dí.mek/*, target phonemes /i/ and /u/ are swapped (giving rise to a serial order error) and target phoneme /l/ is omitted (giving rise to an identity error). (c) PHONEME SHIFT IN COMBINATION WITH A PHONEME SUBSTITUTION AND A PHONEME OMISSION: for example, target: */pig.dúl.mek/*, response: */pug.dí.nek/*, target phonemes /i/ and /u/ are swapped (giving rise to a serial order error), target phoneme /m/ is replaced by /n/ and target phoneme /l/ is omitted (both giving rise to identity errors). (d) ALL PREVIOUS ERRORS IN COMBINATION WITH A PHONEME ADDITION. Table 1 provides an overview of the different response types and their binary values on both retention variables at the phoneme level.

< Insert Table 1 about here >

Phoneme Identity and Serial Order were treated as binary variables because we wanted to find out whether participants were able to repeat the nonword without making any phoneme identity error and/or without making any serial order error. This scoring technique is quite similar to the way responses are scored at the nonword level, i.e., have ALL phonemes been repeated in their CORRECT serial position (hence, a correct response) or not? As a matter of fact, this question is split up in two separate questions, which makes it possible to probe into the phoneme level: (a) have all phonemes been repeated? (phoneme identity retention) and (b) have all correctly repeated phonemes been placed in their correct serial position? (serial order retention). The use of a binary coding system is not new in STM research. Note

that most STM tasks and NRT tasks make use of binary scores (see for instance Martens & de Jong, 2006; Grainger, 1990). Recently, Schraeyen et al. (in press) used the binary coding scheme described above to investigate the relationship between young children's literacy skill and their retention performance at the phoneme level. By using binary scores at the phoneme level we address the following question at the conceptual level: has the participant been able to make a response without any failure of the identity and/or serial order retention mechanism? As mentioned above we hypothesize that this ability will be stronger in monolinguals than in bilinguals for the retention of phonemes' identity but not for their serial order. It might be tempting to suggest the use of proportional scores when analyzing the retention of phonemes' identity and serial order, e.g., the proportion of phonemes from the nonword that has been correctly repeated. However, proportions create a non-trivial problem when the data are analyzed with a linear mixed effects model, i.e., a model in which both the theoretically relevant predictors and several random variables (here: participants and items) are included. First, it is well-known that proportions often give rise to marginal residuals that are not normally distributed, whereas this is a basic assumption of such models. Marginal residuals are the difference scores between the observed values and the values that are predicted by the fitted model. Second, proportions have the undesirable property that predicted values are sometimes larger than the theoretical maximum of 1. Obviously, it is impossible to observe a proportion that is larger than 1. These two problems actually occurred when scoring the responses in the form of proportional data. Importantly, it can be shown that these two disadvantages do not occur for binary data, which are analyzed with a generalized linear mixed effects model with a logit link (see for instance Cohen, Cohen, West & Aiken, 2003). Hence, the only way to study the responses at the phoneme level in a statistically

sound way is by scoring these responses with respect to two binary variables. As argued above, these analyses run quite parallel to the way responses are analyzed at the nonword level and, moreover, can be translated in sensible questions at the conceptual level.

In a second set of analyses we wanted to study the effect of language background (monolingual vs. bilingual) on the distribution of the three error types. To this end we contrasted pure identity errors (i.e., responses with one or more failures to retain a phoneme's identity but without serial order errors) with all other errors, (2) pure serial order errors with all other errors and (3) combined errors with all other errors.

<HB>*Data analysis*

At first, Pearson correlation coefficients were calculated between vocabulary scores and the three dependent variables: nonword level performance, phoneme identity performance, and serial order performance. These correlations enabled us to first explore the linear association between the different measures used in this study. For this purpose, we used total sum scores, i.e., the total number of correct responses (with respect to each dependent variable) per participant, aggregated across all items.

Next, we performed a more fine-grained analysis in which the effect of the independent variables was examined at the level of the INDIVIDUAL ITEM RESPONSES. To this end, the data were analyzed with generalized linear mixed-effects models (GLMMs). A logistic link function was used because the dependent variables were binary. GLMMs make it possible to simultaneously assess the variance explained by the random effect of participants and the random effect of items. Inclusion of the latter is necessary to avoid the language-as-a-fixed-effect fallacy (Clark, 1973). In all initial models, we included Language Group (coded

as bilingual = 0, monolingual = 1), Syllable Length (3, 4, and 5 syllables, coded with 2 dummies and syllable length = 3 as the reference category), and the Language Group x Syllable Length interaction as fixed effects. In addition, Serial Order performance was used as a covariate in the analysis of Phoneme Identity performance (and vice versa) to statistically control for the collinearity between these factors (see Table 3). It is common practice to use such an approach (see Verbeke & Molenberghs, 2000). During the model-building phase, it was examined whether removing the interaction term significantly decreased the model fit. All model parameters were estimated using the Maximum Likelihood method, which has the advantage that likelihood-ratio tests can be conducted to formally compare the fit of the different (nested) models (e.g., the model with the interaction term and the model without this term). Once the final model was obtained, it was refitted using the Restricted Maximum Likelihood approach because the latter approach yields more accurate estimates of the variance components in smaller samples. For the mathematical details of GLMMs the interested reader is referred to Molenberghs and Verbeke (2005). All models were fitted using R (packages lme4, Bates & Maechler, 2009; language R, Baayen, 2008). Throughout the paper, the significance level was set at $p \leq .05$.

<HA>Results

<HB>Vocabulary performance and NRT accuracy

Table 2 presents the means and standard deviations for the expressive vocabulary measure and the overall accuracy scores on the NRT measure. Independent-samples t-tests for each measure separately revealed that children in the monolingual group had higher expressive vocabulary scores than those in the bilingual group, $t(36) = 8.48$, $p < .0001$. In addition, the

monolingual group had higher overall accuracy NRT-scores than the bilingual group, $t(36) = 2.75, p = .01$.

<Insert Table 2 about here >

<HB> *Correlation analyses*

As can be seen in Table 3, we found high positive (Pearson) correlations between the participant totals in all pair-wise combinations of the three dependent variables: nonword retention, phoneme identity retention, and serial order retention (all $r_s \geq .86$, all $p_s < .0001$). These correlations reveal that children with higher (or lower) levels of nonword level performance also tend to have better (or poorer) phoneme identity performance and better (or poorer) serial order performance. In other words, children who are better in recalling entire nonwords make more responses in which all phonemes are recalled correctly (correct responses but also responses with serial order errors) and more responses in which the phonemes that have been correctly recalled are all placed in their correct (relative) serial order (i.e., correct responses but also responses with identity errors). Obviously, these high correlations are in part due to the fact that correct nonword repetitions are scored as correct responses with respect to all three dependent variables.

< Insert Table 3 about here>

The correlations between Language Group and the three dependent variables were substantially lower. Language Group was significantly positively correlated with nonword level performance ($r = .41, p < .01$) and phoneme identity performance ($r = .39, p < .01$), but

not with serial order performance ($r = .15, p > .05$). These correlations are in line with our hypotheses: monolinguals were predicted to outperform bilinguals on the overall accuracy of their responses and on the retention of phonemes' identity but not on the retention of their serial order. Below we will report the results of the GLMMs in which the effect of Language Group on each of our three dependent measures was analyzed (based on item-level scores).

<BC>Nonword level performance in monolinguals and bilinguals

The results of the GLMM with overall nonword performance as the dependent variable are summarized in Table 4. It turned out that the two-way interaction Language Group x Syllable Length was not significant, $p > .05$. Therefore, in the final model, this interaction term was removed.

< Insert Table 4 about here>

As expected, a main effect of Language Group was observed, $p = .005$, demonstrating that the probability of a correct response (i.e., a correct nonword repetition) was higher for the monolingual group than for the bilingual group. We also observed a main effect of Syllable Length, indicating that the probability of a correct response was higher for three-syllable nonwords than for four-syllable nonwords, $p = .0007$, and was higher for three-syllable nonwords than for five-syllable nonwords, $p < .0001$. In addition, the probability of a correct response was higher for four-syllable nonwords than for five-syllable nonwords, $\beta = -1.76$ for syllable length 5, $p = .0004$ (this β estimate is obtained by fitting the same model as shown in Table 4 but using syllable length 4 as the reference level; data not shown).

<HB> *Phoneme level performance in monolinguals and bilinguals*

At the level of phoneme performance, we distinguished between the retention of (a) phoneme identity and (b) serial order. In Table 5, the mean error rates for the different response types are presented as a function of both language group and syllable length.

<Insert Table 5 about here>

Note that in the GLMM analyses below, all responses were included, as each response (i.e., correct or incorrect) could be scored with respect to the two phoneme-related variables. In effect, each response was a demonstration of the child's ability (a) to recall all phoneme identities correctly or not and (b) to recall their serial order correctly or not.

<HC> *Retention of phoneme Identity*

Table 6 summarizes the results of the final model for the analysis of children's phoneme identity performance. The interaction Language Group x Syllable Length was not significant and, therefore, removed in the final model. The final GLMM revealed a main effect of Language Group after statistically controlling for serial order performance, $p = .003$.

Monolingual children made significantly more responses in which all phonemes were correctly recalled. The effect of serial order performance on the retention of phoneme identity was significant ($p < .0001$), which is due to the strong correlation between the two phoneme-related variables (see also Table 3). Importantly, the effect of Language Group on the retention of phoneme identity was significant when this correlation was brought under statistical control. As the effect of Language Group could not be contaminated by this

collinearity it only accounts for the variance in phoneme identity retention that could not also be explained by serial order retention. In addition, we observed a main effect of Syllable Length for both monolinguals and bilinguals (no interaction with Language Group). Phoneme identity performance was higher in three-syllable nonwords compared to four-syllable nonwords, $p = .002$, and compared to five-syllable nonwords, $p < .0001$. Similarly, phoneme identity performance was higher in four-syllable nonwords compared to five-syllable nonwords, $\beta = -1.24$ for syllable length 5, $p = .0002$ (this β estimate is obtained by fitting the same model as shown in Table 6 but using syllable length 4 as the reference level; data not shown).

< Insert Table 6 about here >

<HC> *Retention of serial order*

Table 7 summarizes the results of the final model for the analysis of serial order performance. The interaction Language Group x Syllable Length was not significant and, therefore, removed in the final model. The final GLMM revealed no main effect of Language Group, $p = .63$, indicating that there was no significant influence of Language Group on the retention of serial order. We observed a significant main effect of Syllable Length (all $p < .0001$). This indicates that serial order retention decreased with increasing syllable length. For both monolinguals and bilinguals (no interaction with Language Group), serial order performance was higher in three-syllable nonwords than in four-syllable nonwords, $p < .0001$, and five-syllable nonwords, $p < .0001$. Similarly, serial order performance was higher in four-syllable nonwords than in five-syllable nonwords, $\beta = -.96$ for syllable length 5, $p < .0001$ (this β

estimate is obtained by fitting the same model as shown in Table 7 but using syllable length 4 as the reference level; data not shown). Finally, children's phoneme identity performance significantly decreased as their serial order decreased ($p < .0001$, see also the high correlation value in Table 3). As the GLMM brings this correlation under statistical control the effect of Language Group on serial order performance is uncontaminated by this collinearity in the data.

< Insert Table 7 about here >

<HB>Distribution of error types in monolinguals and bilinguals

All preceding analyses were performed on all data and, hence, included the four type response types: correct responses, pure identity errors, pure serial order errors, and combined errors. In the analysis below we will look at the effect of Language Group on the distribution of the error types only. When there is a retention failure, the two groups may differ in the association strength between the two retention mechanisms, i.e., the probability that, if one error type is made, the other one is also made. For instance, one group might make more pure identity errors than the other group, whereas the opposite pattern might occur for combined errors. This would reveal that the two retention mechanisms are more strongly associated in one group than in the other. Table 8 provides an overview of the number (and percentage) of error responses per error category.

< Insert Table 8 about here >

Clearly, pure serial order errors, i.e., a problem with the retention of phonemes' serial order when the retention of all phonemes' identity is perfect, occur very seldom, both in the monolingual group (4.8 %) and bilingual group (4.1 %). Not surprisingly, a GLMM in which the probability of such an error was predicted by Language Group revealed no effect of this predictor, $\beta = .21$ (SE $\beta = .18$), $p = .66$. Considering these facts, we decided to remove this error type from the analysis.

Hence, the purpose of the analysis is to investigate the effect of Language Group on the distribution of pure identity errors and combined errors, both representing more than 40% of the total set of error responses in the two groups. Interestingly, these two error types differ only in the presence or absence of a serial order error (combined errors vs. pure identity errors, respectively). Hence, the distribution of the two error types is determined by the likelihood of making a serial order error IF A PHONEME IDENTITY ERROR HAS BEEN MADE. The more serial errors are made, the stronger the co-occurrence of the two error types. If Language Group has a significant effect on the serial order scores in this subset of the data (the identity score being constant), this will reveal a different association strength between identity errors and serial order errors. Table 9 summarizes the results of the final model for the analysis of these two error types (pure identity vs. combined errors).

< Insert Table 9 about here >

The interaction Language Group x Syllable Length was not significant and, therefore, removed in the final model. The final GLMM revealed no main effect of Language Group, $p = .97$, indicating that the probability of making a serial order error, given that an identity

error is made, is the same for the monolingual and bilingual groups. Hence, the association strength between the two types of retention errors is the same in both language groups. In addition, for both groups, the probability of making a serial order error, given that an identity error is made, increases with increasing syllable length.

The above contrast between two response types amounted to an investigation of the effect of Language Group on a conditional probability, i.e., the probability of making a serial order error if an identity error has been made. It would have been interesting to make the complimentary analysis: the effect of Language Group on the retention of phoneme identity if a serial error has been made. However, the small percentage of pure serial order errors makes this a hazardous exercise in a statistical test (GLMM) that takes the individual response as the unit of analysis. However, the more robust chi-square test that compares the two language groups with respect to the distribution of the two error types, i.e., pure serial order errors and combined errors, reveals that this distribution does not differ between the two groups ($\chi^2 = .43, p = .51$). Incidentally, the complimentary chi-square test, assessing the effect of language group on the effect of serial order if an identity error has been made was non-significant as well, confirming the outcome of the above GLMM ($\chi^2 = 1.03, p = .31$). Together, these results suggest that the association strength between retention errors at the level of phoneme identity and at the level of serial order was the same for the monolingual and bilingual children. There was a strong correlation between the two error types (Table 3) but the conditional probability of making one phoneme-related error if an error on the other phoneme-related variable had been made did not differ between groups.

As mentioned above, the technique of varying the binary value on one phoneme-related variable while keeping the value on the other variable constant could not be used for

pairwise comparisons involving the minority response group of pure serial order errors. However, we used the technique to assess the effect of language group on the comparison between correct responses and pure identity errors (both representing a considerable percentage of the data), i.e., a contrast in which all responses were correct with respect to the retention of serial order but differed with respect to the retention of phonemes' identity. This analysis makes it possible to study the effect of Language Group on identity retention if NO serial error has been made. In this analysis the effect of language group on identity retention can be studied without the need to statistically control for its correlation with serial order retention (as in the earlier reported analysis on all responses), as the latter variable is restricted to the response subset in which no serial order errors are made.

Table 10 summarizes the results of the final model for the analysis of phoneme identity performance in the absence of serial order errors. The interaction Language Group x Syllable Length was not significant and, therefore, removed in the final model. The final GLMM revealed a main effect of Language Group, $p = .0005$. Monolingual children made significantly more responses in which all phonemes were correctly recalled. In addition, we observed a main effect of Syllable Length for both monolinguals and bilinguals (no interaction with Language Group). Phoneme identity performance was higher in three-syllable nonwords than in four-syllable nonwords, $p = .006$, and five-syllable nonwords, $p < .0001$. Similarly, phoneme identity performance was higher in four-syllable nonwords than in five-syllable nonwords, $\beta = -1.22$ for syllable length 5, $p = .003$ (this β estimate is obtained by fitting the same model as shown in Table 10 but using syllable length 4 as the reference level; data not shown).

< Insert Table 10 about here >

<HA>Discussion

The present study was designed to further understand earlier reported overall accuracy differences in NRT-performance between bilinguals and monolinguals by means of an in-depth analysis of their nonword repetition responses. By taking retention ability at the PHONEME LEVEL into account, we collected more detailed information on the determinants of typically developing bilinguals' STM capacity for the encoding and retrieval of non-native phonological information than has been revealed in previous studies.

In bilingual language assessment, a STM task such as a NRT has been proposed as a valuable component of a language assessment protocol. However, so far, NRT performance has been mainly assessed by determining the percentage of correctly repeated nonwords (number of correct responses on the total number of items in the test) or the percentage of correctly recalled phonemes in the participant's responses (Gutiérrez-Clellen & Simon-Cereijido, 2010), except for some recent studies, which, for instance, considered onsets versus codas (Sorenson Duncan & Paradis, 2016), or consonants versus vowels (Topbaş et al., 2014). In addition, Girbau and Schwartz (2008) examined NRT performance in bilingual children with and without language impairment and analyzed substitution, omission, and addition errors. Even though these studies made an important step towards a more fine-grained analysis of NRT responses in bilinguals, none of them probed into the constituent retention processes that are required to perform the NRT. That was the goal of the current study.

In this study we differentiated between participants' retention of phonemes' identity (phoneme identity performance) and the retention of phonemes' serial order (serial order

performance). This was based on the idea that correct nonword repetition depends on the correct retention of the item's constituent phonemes and the serial order of these phonemes. These two types of retention ability can fail independently of each other. Hence, studying these two retention skills separately makes it possible to formulate more precise claims on the processes that underlie a participant's overall NRT performance. Note that, by taking these phoneme-related variables into account, error responses in a NRT also become an important piece of information. Indeed, such responses also provide valuable information, as they can be incorrect with respect to one phoneme-related variable but correct with respect to the other. In other words, the introduction of phoneme-related variables makes it possible to dissect the retention of nonwords into component retention processes, each of which sheds light on a distinct memory process that is required for adequate task performance.

Earlier findings suggest that retention of item information depends on language knowledge stored in LTM, whereas retention of serial order information is related to the quality of a STM device specifically dedicated to the storage of serial order information (Gupta, 2003). We expected bilinguals to have more problems than monolinguals with the retention of the phonemes in a nonword in their L2, as the result of a smaller vocabulary and the fewer possibilities to extract well-defined phonemic representations. In contrast, we did not expect the bilinguals to have problems with the retention of phonemes' serial order as the result of a deficit in their STM serial ordering device (as has been reported for dyslexics, for instance, see Hachmann, Bogaerts, Szmalec, Woumans, Duyck & Job, 2014). Indeed, none of the bilingual (or monolingual) children in our study had developmental problems. However, when the serial order of PHONEMES in nonwords must be retained, group differences might also originate in different degrees of phonotactic knowledge, which is stored in LTM. These

are likely to result from differences in the familiarity with the combinatorial possibilities of the phonemes in the target language, i.e., the language's phonotactic restrictions. Despite differences between Dutch and Turkish, the bilinguals' native language, we did not expect phonotactic knowledge to cause any differences in the serial retention scores of the two language groups. Indeed, the phonotactic restrictions that characterized the items in our NRT materials also occurred in the bilinguals' native language, making them useful for the bilingual children as a familiar encoding scheme for the phonemes in the nonword stimulus. Hence, we did not predict differences between the two language groups with respect to serial order retention.

At first, we investigated whether we could find the pattern that is typically reported in previous NRT research, that is, an association between NRT performance and prior language knowledge. We found a significant positive correlation between language group and overall nonword level performance ($r = .41, p < .01$, see Table 3), i.e., the effect of language group on the total number of nonword responses per participant. The same effect was found when analyzing the data at the level of individual responses, i.e., by means of a GLMM. The probability of a correct response was higher in monolinguals compared to bilinguals. This finding is in line with previous studies, indicating that monolinguals outperform bilinguals on a NRT-task in bilinguals' second language (e.g., Windsor et al., 2010; Messer et al., 2010; Kohnert, Windsor & Yim, 2006; Lanfranchi & Swanson, 2005) and confirms the well-established idea that vocabulary knowledge, which significantly differed between the two groups (see Table 2), is associated with STM performance (e.g., Michas & Henry, 1994). In addition, we found that the number of syllables in a nonword predicted the probability of a correct nonword repetition, an effect that did not differ between groups. This outcome is also

consistent with a body of literature, indicating better recall of shorter nonwords as compared to longer nonwords in both monolinguals (Gathercole, 2006) and bilinguals (Girbau & Schwartz, 2008). This replication is not surprising, as it seems obvious that longer nonwords require more STM resources than short ones.

Secondly, we were interested in finding out whether the probability of making no phoneme identity errors, i.e., flawless phoneme identity performance, and the probability of making no serial order errors on the correctly recalled phonemes, i.e., flawless serial order performance, is predicted by participants' language group (monolingual vs. bilingual). The correlation analyses on the basis of participants' total number of correct nonword repetitions revealed a significant correlation between language group and phoneme identity performance ($r = .39, p < .0001$, an outcome that is equivalent to an independent samples t-test). In contrast, we found no significant correlation between language group and serial order performance ($r = .15, p > .05$). Thus, a rough, exploratory analysis suggests that the ability to retain phonemes' identity seems to be dependent on participants' language background, but the ability to retain phonemes' serial order does not.

However, the fact that there was a strong positive correlation between participants' phoneme identity performance and their serial order performance ($r = .86, p < .0001$) makes it difficult to interpret these correlations with language group. To assess the effect of language group on each phoneme-related variable, the collinearity between these two variables must be statistically removed from the variance that language group has to predict. This makes it possible to assess the unique effect of language group on each phoneme-related variable. Therefore, we opted for GLMM analyses in which one type of phoneme performance (phoneme identity or serial order performance) was included as a covariate when predicting

performance on the other phoneme-related variable. For instance, we predicted phoneme identity performance on individual responses from participants' language group and nonword syllable length while statistically controlling for serial order performance (see also Staels & Van den Broeck, 2014). The GLMM in which PHONEME IDENTITY PERFORMANCE was analyzed revealed an effect of language group after controlling for serial order performance ($p = .003$). Note that this is a strong effect. Even when statistically controlling for the very strong correlation between the two phoneme-related variables, the effect of language group is still significant. In other words, the difference between monolinguals and bilinguals still significantly predicts the variance in the retention of phoneme identity that cannot be explained by serial order retention. Bilingual children performed at a level below that of their monolingual age peers in three-, four- and five-syllable nonwords with respect to the retention of phonemes. The interaction between language group and syllable length was non-significant. However, the GLMM in which both monolinguals and bilinguals were compared on their SERIAL ORDER PERFORMANCE did not yield an effect of language group after removal of the variability on serial order retention that was caused by phoneme identity retention. Bilinguals were as successful as their monolingual peers in retaining the correct serial order of correctly recalled phonemes when repeating a nonword.

Thirdly, we focused on participants' error responses to check whether the distribution of error types (pure phoneme identity errors, pure serial order errors, or a combination of both errors) was similar in both groups. We removed the pure serial order errors from the data, i.e., responses on which only serial order errors were made, as these accounted for a very small minority of the data (about 5% of the responses in both groups). A GLMM assessing the effect of language group on the distribution of the remaining two error types revealed that the

distribution of these two main error types was the same for both groups. Importantly, as pure identity errors and combined errors only differ in the absence or presence of a serial order error, respectively, (a phoneme identity error being made in both) the failure to find an effect of language group in these error data is in line with the outcome of the global analysis (including all responses: correct and incorrect ones). Language group does not appear to affect children's serial order retention ability. Despite this convergence between the two analyses, there is a crucial difference between them, besides the different response types that are included in each analysis. The overall analysis investigates the effect of language group after the collinearity between the two phoneme variables has been statistically removed from the variability in the serial error scores. In contrast, the analysis of the effect of language group on serial order retention in the above set of error responses investigates whether language group can account for the variability in the actual serial order data. Indeed, the constant value for the retention of phonemes' identity (i.e., retention failure) removes the need to statistically control for this type of retention. Hence, the latter analysis shows that the failure to find an effect of language group on serial order performance in the global GLMM was not due to a reduction in the variability of the data after statistical controlling for the correlation between the two phoneme variables.

The analysis also demonstrates that the associative strength between the mechanisms that are responsible for identity retention and for serial order retention is the same in both groups, at least in one direction. More particularly, the conditional probability of making a serial order retention error if a phoneme identity error has been made is the same for both language groups. The chi-square analysis comparing the distribution of pure serial errors and combined errors in both language groups (a GLMM being unfit for this purpose, given the

small number of pure serial order errors) suggested that the associative strength between the two error types is the same for both groups in the opposite direction as well. The non-significant interaction between the error type variable and language group in this analysis indicates that the probability of making a phoneme identity retention error if a serial error is made is the same for both groups.

Finally, we generalized the rationale of keeping the value on one phoneme variable constant while varying the other. Given the small percentage of pure serial order errors, the only remaining contrast of this type that is amenable to a GLMM analysis involved the correct responses and the pure identity errors (two sizeable response sets in both language groups). These response types share the property that no serial order error has been made. The outcome of this analysis, too, was in line with the analysis on all responses: if no serial order error was made, bilinguals made significantly more identity errors than monolinguals.

In sum, the most important conclusion is that (a) a failure of the retention mechanism for phoneme identity (i.e., phoneme identity performance), operationally defined as a failure to reproduce all phonemes from the nonword stimulus, occurred more often in bilinguals than in monolinguals, and that (b) a failure of the retention mechanism for serial order occurred equally often in both groups. Thus, the results confirm our hypothesis that phoneme identity retention but not serial order retention is a crucial factor for describing the widely reported differences in overall NRT-performance between typically developing monolingual and bilingual children (with the proviso that the nonwords' phonotactic constraints are in line with the phonotactic restrictions in the bilinguals' native language).

As this type of analyses in a single NRT task are novel, our results cannot directly be compared with previous research in BILINGUAL children. However, it is interesting to consider

our results in the light of the recent conceptualization of item and order retention in current STM models.

Given the observation that the Turkish–Dutch bilingual children in our study showed significantly lower expressive vocabulary skills in Dutch, it might not be surprising that phoneme identity performance was better in monolinguals compared to bilinguals. Indeed, a possible explanation, which is in line with previous research, is that bilinguals are less acquainted with phonemes in their second language, because they know fewer words in that language, and, hence, have fewer opportunities for inducing the phonemes in the language. As a result, they might have more problems to flawlessly segment the non-familiar phoneme input of a nonword. This could be the consequence of less well-developed or less accessible phoneme representations, both possibilities being plausible outcomes of the less frequent exposure to the phonemes of the language from which the nonwords are derived (see also Metsala & Walley, 1998). As a consequence, the presented nonwords would not as readily ‘fall apart’ into their constituent phonemes as is the case for monolinguals. Although previous STM research did not directly distinguish item and serial order retention in a single NRT task, our results are compatible with the idea proposed by current STM models that storage of ITEM information (i.e., phoneme identity performance in our study) strongly depends on the richness of the language knowledge system (e.g., Cumming, Page & Norris, 2003; Gupta et al., 2005; Majerus et al., 2008).

In these STM models the retention of SERIAL ORDER information is linked to a specialized STM system, so called ‘sequence memory’ (Gupta, 2003; Gupta et al., 2005), which makes it possible to store the succession of novel sequences. The existence of such a mechanism would not predict problems associated with the retention of phoneme sequences,

i.e., phonemes' serial order, in our typically developing (monolingual and bilingual) children, who have no history of language problems or phonologically processing problems. However, as serial order was defined over phonemes in the current study, participants' knowledge of the ordering possibilities among phonemes may have played a role in the retention of the nonword items. More particularly, LTM knowledge of phonotactic constraints on the phonemic 'items' is likely to have been a determinant of the retention of phonemes' serial order. However, as the nonwords did not violate phonotactic constraints in Turkish (the bilinguals' native language) this would not cause the bilinguals to perform worse on serial order retention than the monolinguals either. Our results supported this hypothesis.

Note that the conclusion with respect to serial order retention is contingent on the convergence of Dutch and Turkish phonotactic constraints in the specific set of nonword items in our experiment. It is quite possible (even plausible) that monolingual children will have a better serial order performance than their bilingual peers when the nonwords do not respect the phonotactic restrictions of the bilinguals' native language. Indeed, it is expected that one's familiarity with the words of a language will not only entail a sensitivity to the phonemes in the language but also to these phonemes' combinatorial possibilities, i.e., the phonotactic regularities in the language, which determine legal phoneme sequences (serial order of phonemes). This sensitivity will increase as vocabulary size increases, which is likely to have a beneficial effect on participants' ability to encode, retain, and recall sequences of phonemes that do not form a real word (i.e., items in the NRT). Further experiments will have to put this hypothesis to the experimental test.

<HA>Conclusion

The present study examined monolingual and bilingual children's retention abilities for phonemes' identities and their serial order, in an attempt to better understand the mechanisms that underlie the often observed differences in the overall NRT performance in both groups. To this end, we developed measures for the retention of the phonemes themselves and the retention of their serial order. This method has not been used by previous researchers, who always focused on the proportion of correct responses or the proportion of correctly recalled phonemes. In line with previous research, the primary finding confirms that overall NRT performance is influenced by the amount of prior exposure to the target language: the proportion of correct responses was higher for monolinguals than for bilinguals. More importantly, the analyses at the phoneme level shed light on factors that may explain this difference. Bilinguals turned out to be less sensitive to retain phonemes than monolinguals, presumably as a consequence of less familiarity with the words in the target language, which we inferred from bilinguals' significantly smaller expressive vocabulary knowledge. A smaller vocabulary size is likely to cause more poorly developed or less accessible phoneme representations, which are required to adequately encode novel phoneme strings. In addition, bilinguals were equally sensitive to phonemes' serial order as monolinguals when repeating nonwords. The latter finding, being a null effect, does not by itself confirm the presence of a special sequence memory in current STM-models, but is compatible with such a mechanism. Indeed, as all children developed normally, no problems with a sequence memory in STM were expected in the monolingual and bilingual groups. However, as the retention of nonwords integrates the retention of item and serial order information, serial order information for phonemes is also likely related to participants' familiarity with the way in which phonemes can be serially ordered in the target language, i.e., to LTM language

knowledge. In contrast to previous studies, serial order memory was not tested by having participants recall the random order of familiar items but by having them recall the non-random, i.e., phonotactically constrained, order of (familiar) items, i.e., phonemes. Despite the differences between Dutch and Turkish at the phonological level, the phonotactic principles of both languages happened to coincide in our set of nonwords, which is likely to account for the finding that bilinguals did not perform worse on the serial order measure than monolinguals.

In sum, the findings of the present study enhance our understanding of mechanisms that underlie inferior NRT performance in bilinguals' L2. No doubt, more research needs to be conducted on the relationship between NRT performance and L2 language exposure. Based on the current findings, some recommendations for future studies can be made. For instance, it would be interesting to study the supportive role of phonotactic knowledge in the retention of serial order in children's L2. This can be achieved by designing a NRT that consists of nonwords whose phonotactic patterns are dominant in bilinguals' second language but are absent in their first language.

<HA>References

- Bates, D., & Maechler, M. (2009). *lme4: Linear mixed-effects models using Eigen and Eigen* [Computer software manual]. Available: <http://CRAN.R-project.org/package=lme4>.
- Boets, B. (2006). *Early literacy development in children at risk for dyslexia. A longitudinal study of the general magnocellular theory* (Unpublished doctoral dissertation). University of Leuven, Belgium.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: a network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581.

- Burgess, N. & Hitch, G.J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55, 627-652.
- Cheung, H. (1996). Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology*, 32, 867-873.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Coady, J.A., & Evans, J.L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communicative Disorders*, 43, 1-40.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences*, 3rd edition. Lawrence Erlbaum Associates.
- Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language and literacy: developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry*, 48, 147-156.
- Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, 11, 43-63.
- Edwards, J., Beckman, M.E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47, 421-436.
- Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory & Language*, 29, 336-360.

- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. J. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, 45A, 1-30.
- Gathercole, S. E. (2006). Complexities and constraints in nonword repetition and word learning. *Applied Psycholinguistics*, 27, 599-613.
- Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish-English bilingual children with and without Specific Language Impairment. *Journal of Communication Disorders*, 41, 124-145.
- Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology*, 56A, 1213-1236.
- Gupta, P., Lipinski, J., Abbs, B., & Lin, P-H. (2005). Serial Position Effects in Nonword Repetition. *Journal of Memory and Language*, 53, 141-162.
- Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (eds.), *Interactions between short-term and long-term memory in the verbal domain*, pp. 108–135. Hove, England: Psychology Press.
- Gutiérrez-Clellen, V., & Simon-Cereijido, G. (2010). Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English-Speaking Children: Does the Language of Assessment Matter? *Learning Disabilities Research & Practice*, 25, 48-58.
- Hachmann, W., Bogaerts, L., Szmalec A., Woumans, E., Duyck, W., & Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of dyslexia*, 64, 121-136.

- Hitch, G.J., Fastame, M.C., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory*, 3-4, 247-258.
- Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30, 685-701.
- Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice*, 21, 19–29.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
- Lanfranchi, S., & Swanson, H.L. (2005). Short-term memory and working memory in children as a function of language-specific knowledge in English and Spanish. *Learning and Individual Differences*, 15, 299-319.
- Lee, S.A.S., & Gorman, B.K. (2012). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism*, 17, 479-495.
- Majerus, S., & Cowan, N. (2016). The nature of verbal short-term impairment in dyslexia: The importance of serial order. *Frontiers in Psychology*, 7, 1-8. doi: 10.3389/fpsyg.2016.01522
- Masoura, E. & Gathercole, S. E. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology*, 34, 383-388.
- McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology*, 114, 75–88.

- Messer, M.H., Leseman, P.P.M., Mayo, A.Y., & Boom, J. (2010). Phonotactic probability effects in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, 105, 306-323.
- Michas, I.C., & Henry, L.A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology*, 12, 147–163.
- Munson, B., Kurtz, B.A., & Windsor, J. (2005). The Influence of Vocabulary Size, Phonotactic Probability, and Wordlikeness on Nonword Repetitions of Children with and without Language Impairments. *Journal of Speech, Language, and Hearing Research*, 48, 1033-1047.
- Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language*, 50, 113-133.
- Page, M.P.A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761-781.
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, 50, 408-412.
- Schraeyen, K., Geudens, A., Ghesquire, P., Van der Elst, W., & Sandra, D. (in press). Poor performance on the retention of phonemes' serial order in short-term memory reflects young children's poor reading skills. *The Mental Lexicon*.
- Semel, E., Wiig, E.H., & Secord, W.A. (2008). *Clinical Evaluation of Language Fundamentals (CELF-4NL): handleiding*. Amsterdam: Pearson.
- Sorenson Duncan, T., & Paradis, J. (2016). English Language Learners' Nonword Repetition Performance: The Influence of Age, L2 Vocabulary Size, Length of L2 Exposure, and L1 Phonology. *Journal of speech, language, and hearing research*, 59, 39-48.

- Staels, E., & Van den Broeck, W. (2014). Order short-term memory is not impaired in dyslexia and does not affect orthographic learning. *Frontiers in Human Neuroscience*, 8, 1-16.
- Summers, C., Bohman, T.M., Gillam, R.B., Pena, E.D., & Bedore, L.M. (2010). Bilingual performance on NWR in Spanish and English. *International Journal of Language and Communication Disorders*, 45, 480-493.
- Thordardottir, E., Kehayia, E., Lessard, N., Sutton, A. & Trudeau, N. (2010). Typical performance on tests of language knowledge and language processing of French-speaking 5-year-olds. *Canadian Journal of Speech Language Pathology and Audiology*, 34, 5-16.
- Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46, 1-16.
- Thorn, A.S.C., & Gathercole, S.E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology*, 52A, 303–324.
- Topbaş, S., & Kopkalli-Yavuz, H. (2008). Reviewing sonority for word-final sonorant+obstruent consonant cluster development in Turkish. *Clinical Linguistics and Phonetics*, 22, 871–880.
- Topbaş, S., Kaçar-Kütükçü, D., Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical linguistics & phonetics*, 28, 602-16.
- Windsor, J., Kohnert, K., Lobitz, K., & Pham, G. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*,

19, 298-310.

Footnotes

¹ In the literature the term ‘nonwords’ is used to refer to items in a NRT-task. However, note that all NRT-items in this study were pseudowords, i.e., items respecting the phonological constraints of Dutch.

² In the description of these two groups all children who repeated their year have been removed.

Appendix A. Target Items for the Nonword Repetition Task (NRT)

Three-syllable items	Four-syllable items	Five-syllable items
/pi.pó.ket/	/sir.peg.wót.nal/	/bi.ko.ta.dú.kep/
/ves.róf.sif/	/kep.dar.bít.puk/	/gam.bis.kef.lún.jor/
/lem.rós.pag/	/lu.mo.gá.pes/	/wog.tum.dis.jáf.kel/
/pir.dáp.ket/	/ni.ne.lú.mar/	/bog.nup.sar.líf.tek/
/ku.mí.gar/	/jaf.ker.túm.sil/	/ga.go.li.sú.gef/
/po.sá.lin/	/be.fo.tí.ral/	/na.mo.ni.fú.nem/
/ni.mú.naf/	/zos.gef.zíl.vas/	/vug.zas.gof.líf.sef/
/nom.lún.fam/	/hi.so.ré.fum/	/pip.taf.bip.két.duk/
/pig.dúl.mek/	/no.si.gé.fas/	/so.ra.mi.ké.tul/
/har.lón.wig/	/keg.jol.bíf.mas/	/lu.re.fa.ní.pos/
/gi.ká.lom/	/pe.da.tí.tup/	/hu.sa.li.mó.gep/
/sa.vé.fus/	/mif.nem.lúm.zan/	/mun.fom.lin.zém.bam/