NGÔN NGỮ

& đời sống

TẠP CHÍ CỦA
HỘI NGÔN NGỮ HỌC VIỆT NAM
SỐ 13 (267) 2017

ISSN 0868 - 3409

SỐ ĐẦU TIÊN XUẤT BẢN BẰNG TIẾNG ANH

NGÔN NGỮ & ĐỜI SỐNG
SỐ 13 (267) - 2017

NGÔN NGỮ & đời sống

LANGUAGE AND LIFE

TẠP CHÍ CỦA HỘI NGÔN NGỮ HỌC VIỆT NAM

A JOURNAL OF LINGUISTICS SOCIETY OF VIETNAM

NĂM THỨ 23
23rd YEARS
Mỗi tháng một số
MONTHLY
Số 13 (267) - 2017

TỔNG BIÊN TẬP	**Editor-in-Chief**
GS.TS. Nguyễn Văn Khang	Prof. Dr. NGUYEN VAN KHANG
PHÓ TỔNG BIÊN TẬP	**Vice Editor-in-Chief**
PGS.TS. Phạm Văn Hảo	Assoc.Prof. Dr. PHAM VAN HAO
PGS.TS. Phan Văn Quế	Assoc.Prof. Dr. PHAN VAN QUE

Trụ sở/Office:

Tầng 1, nhà C, ngõ 301 đường Nguyễn Trãi, Thanh Xuân, Hà Nội

Floor 1, House C, Lane 301 Nguyen Trai Street, Thanh Xuan District, Hanoi

Điện thoại/Tel: (84) (024) 3.7624212; E-mail: ngonnguvadoisong@gmail.com

Giấy phép xuất bản/Licence: 244/GP-BTTT (7-8-2014)

Chỉ số/Index: ISSN 0868 - 3409

In tại /Printed at: Nxb Chính trị Quốc gia/National Political Publishing House

LANGUAGE & LIFE

Vol. 13(267)-2017

CONTENTS

NGÔN NGỮ & ĐỜI SỐNG
Số 13(267)-2017

MỤC LỤC

LINGUISTICS AND VIETNAMESE LINGUISTICS

MORPHOLOGICAL EFFECTS IN READING ALOUD VIETNAMESE COMPOUNDS(1)

PHẠM HIỂN*, BENJAMIN TUCKER & HARALD BAAYEN*****

ABSTRACT: This mega-study reports a word naming experiment addressing the production of Vietnamese compounds. Instead of considering as response variable only the naming latency, we also investigate as response variable the acoustic duration of the speech produced. Effects of compound frequency, word length, and constituent family size were present in both latencies and durations, but effects of constituent frequency were absent. This sets Vietnamese apart from languages such as English and Dutch, for which constituent frequency effects are well attested. We attribute the absence of constituent frequency effects on bisyllabic structures constituting the basic, unmarked phonological form of the Vietnamese word. Our data also challenge models of speech production holding that the onset of speech production would not be affected by the properties of glottal stop initial syllables, as we observed naming latencies to be co-determined by the family size, tone, and syllable type of the second syllable. A remarkable convergence of the effects of frequency and family size for response latencies and acoustic durations validates the word naming task combining these two response variables as an experimental paradigm for the study of phonological encoding in speech production.

KEY WORDS: phonotactical; lexical naming; frequency; lexical density; acoustic duration; Vietnamese, mega-study; speech production.

RECEIVED: NOVEMBER 10th, 2017 EDITED: DECEMBER 2nd, 2017

1. Introduction

Vietnamese is a syllable-timed tone language of the Mon-Khmer branch of the Austro-Asiatic language family, well known in the linguistic typology literature for being a morphologically isolating language. Vietnamese has no inflection nor derivation, but compounding is ubiquitous. Most Vietnamese words consist of two so-called syllabemes, phonotactically highly restricted syllables that simultaneously function as the basic phonological and the basic morphological units. Because many syllablemes can also be used as independent onomasiological units, albeit typically infrequently (Calculations on the Vietnamese contemporary dictionary (Hoàng Phê, 2000) indicate that although 81.5% of syllabemes can be used as monosyllabic words, only 15% of all words consist of one syllabeme only. The remainder of the lexicon is comprised of 71% disyllabic words, 13% three-to-four syllabic words, and 1% five-syllabic words.), two-syllabeme words pose similar questions about morphological processing as do compounds in Germanic languages such as English, Dutch, and German. Is lexical processing in Vietnamese de-compositional, in the sense that access in comprehension and production is, in some sense, mediated by the constituent syllabemes? Or is lexical processing in this language more holistic, and is the pivotal role in lexical processing reserved for the unmarked form of a Vietnamese word, the two-syllabeme compound?

Although Vietnamese has a rich history of linguistic research (e.g., Trần et al., 1941; Thompson, 1965; Nguyễn, 1981, 1985, 1997), there are no experimental studies of lexical processing in this language, with the exception of Pham and Baayen (2015). These authors report a large-scale visual lexical decision experiment. This experiment revealed opposite effects for compound frequency and constituent frequency and family size (Schreuder and Baayen, 1997; Moscoso del Prado Martín et al., 2004) counts. Higher frequency compounds elicited shorter responses, but compounds with a high constituent frequency and/or family size elicited elongated responses. This pattern of results is very different from that typically reported in regression studies for languages such as English and Dutch (Kuperman et al., 2008, 2009; Baayen et al., 2010), where especially the left constituent's frequency typically predicts shorter

**Dr; Institute of Linguistics, VASS; Email: phamhieniol@gmail.com*

** *Assoc. Prof. Dr.; University of Alberta, Canada*

*** *Prof. Dr; University of Tübingen, Germany*

response latencies. The results obtained for Vietnamese are consistent with an interpretation of the bi-syllabeme word being the unmarked, phonologically preferred form of the Vietnamese word.

The present study aims to provide further clarification of the role of the compound and its constituents in lexical processing, but targeting speech production rather than language comprehension. To test this, we opted for a word naming task. Word naming requires an initial comprehension process, the reading of the visually presented input, which is followed by processes involved in preparing for and executing articulation. Word naming is vulnerable to the criticism that as a hybrid of comprehension and production, it provides little insight into the details of speech production. To assess this criticism, we collected not only naming latencies, but also the acoustic durations of the words read aloud. A central question, then, is whether naming latencies and acoustic durations reveal similar or dissimilar functional dependencies on lexical-distributional properties such as frequency and morphological family size on the one hand, and on intrinsic properties such as lexical tone and syllable structure on the other hand. The more the functional dependencies for latencies and durations diverge, the more likely it is that the naming latencies reflect comprehension, and the acoustic durations reflect speech production. The more similar the two are, the more the naming task is validated as a production task. To see this, note that the acoustic durations reflect the speed with which articulation took place. This speed is likely to be determined primarily by the processes preparing for and executing articulation. Naming latencies provide information only on the processes preparing for articulation. The more these latencies functionally approximate the durations, the more likely it is that the naming latencies are not dominated by initial comprehension processes.

We opted for a single-subject mega-study (for multiple-subject mega-studies, see, e.g., Balota et al., 2007; Yap et al., 2010) of Vietnamese, given logistic constraints on running Vietnamese subjects in Canada on the one hand, and good experiences with a single subject mega-study vis-à-vis a smaller-scale multiple-subject design (Pham and Baayen, 2015). In our experience, a broad coverage of the lexicon is essential for a proper assessment of the interplay of the different factors interacting during lexical processing.

For the analysis of the data harvested with this single-subject regression design, we make use of generalized additive mixed models (GAMMS). Our regression design allows us to address a wide range of issues with comprehensive statistical modeling. By investigating whole-word versus constituent effects, the role of decompositional processes in speech production can be addressed. The model of Levelt et al. (1999) predicts frequency effects of a complex word’s constituents, and the absence of whole-word frequency effects. For Dutch, these predictions appear to be correct, although there are hints of whole-word effects in some experiments (e.g., Bien et al., 2005, 2011). However, whole-word frequency effects in speech production using the picture naming task have also been reported (Janssen et al., 2008). Given the results previously obtained for Vietnamese using the lexical decision task, it seems unlikely that we will find strong constituent frequency effects for Vietnamese on speech production.

With the present mega-study, we targeted two further issues in the speech production literature. The first issue concerns the role of phonological neighbors in speech production. The WEAVER^{++} model of Levelt et al. (1999) predicts the absence of neighborhood effects in speech production. However, neighborhood effects in speech production have been reported, although results have been some what inconsistent (Vitevitch, 2002; Vitevitch and Stamer, 2006; Gahl et al., 2012; Sadat et al., 2014). Given the highly restricted phonotactics of Vietnamese syllabemes, neighborhoods are expected to be dense, which leads us to expect delayed selection of the proper syllabemes in speech production. In the present study, we operationalized neighborhood density by means of the Levenshtein distance calculated over the IPA transcriptions of the compounds.

Tone			
Onset	Rhyme		
	Prevocalic	Nucleus	Coda

Figure 1: Vietnamese syllable structure

The second issue concerns the degree to which speech production is modularized. According to the model of Levelt et al. (1999), the phonological encoding of morphemes proceeds sequentially, beginning with the first morpheme. The model makes the strong prediction that the planning of the initial morpheme is not influenced by subsequent morphemes. The implicit priming task shows, for instance, that knowledge of non-initial segments or morphemes does not facilitate production. Co-articulation processes across syllable boundaries (Bell-Berti and Harris, 1979; Mok, 2010), however, challenge such highly modularized procedures. In the case of Vietnamese, speakers have to realize complex F_0 trajectories for the tones of the syllabemes, and it is conceivable that the planning of these trajectories and possibly tone sandhi processes cannot be postponed (see e.g., Nguyễn and Ingram, 2007), and instead require advanced planning. Given the simple phonological structure of Vietnamese syllabemes, as shown in Figure 1, it might even be the case that the syllabeme structure of the second syllable is already being planned during the planning of the first syllabeme. Specifically, the Vietnamese syllable structure follows the following scheme:

$$C1(w)V(C2)+T \tag{1}$$

where *C1* is the initial onset consonant; *w* is the bilabial prevocalic /w/; *V* is the vowel nucleus; *C2* is the final coda consonant; *T* is the superimposed tone. A Vietnamese syllable consists of three obligatory elements: an onset, a vowel and a tone.

2. Experiment

2.1. Method

Materials

This study examined Vietnamese disyllabic words. Almost all stimuli were taken from the *Từ điển tiếng Việt* 'Vietnamese Dictionary' (Hoàng Phê, 2000). We first included all the disyllabic words in this dictionary, resulting in a total of 16,883 compound words. From this list, we removed all reduplicative words. The resulting set of target words, which comprises of 13,999 compound words, covered all parts of speech. For each word in the list, we obtained a wide range of lexical-distributional variables, including word length, the frequency of occurrence and dispersion of the compounds in a newspaper corpus, the number of phonological neighbors, and the frequency, dispersion, and family size counts (Moscoso del Prado Martín et al., 2004; Baayen, 2010) for the first and the second syllabeme. In addition, the tone realized on the first and second syllabeme, as well as the syllable types of the syllabemes were added as factorial predictors. As controls for voicekey artifacts, we included manner and place of articulation of the first segment, as well a factor indicating whether the first segment was voiced or voiceless. Predictors pertaining to the compound graph such as shortest path lengths, which Pham and Baayen (2015) found to be predictive for visual lexical decision in Vietnamese, did not reach significance in the present study, and will therefore not be discussed further. The analyses reported below are based on the 8875 words for which all lexical distributional statistics were available to us, for which the voicekey was triggered properly, and for which the speech was successfully recorded.

Participants

The first author of this article was the only participant in this experiment. It took him 48 hours to complete the experiment, at his own pace, over a 6-week period.

Apparatus

All the stimuli were merged into one list. A script was written to randomly select 400 stimuli from the list, then merge those stimuli into a template script for DMDX in which 10 practice trials were filled at the beginning of the experiment for accommodating the participant and also for checking that the recording system works properly. By doing so, the participant, who was also the experiment designer, remained completely uninformed about the actual contents of each sub-experiment. The list for a given sub-experiment was also selected randomly from the set of lists.

Stimuli were presented on a 21-in. Dell computer screen with a refresh rate of 60 Hz and a resolution of 1440 x 900 pixels; the computer was controlled by an Intel 3.6GHz processor. Stimuli were presented in lowercase 26-point Courier New font, and they appeared as black characters on a grey background. Stimuli were presented and responses collected with the DMDX software (Forster and Forster, 2003).

The experiment was run in a double-walled sound attenuated booth. Speech was recorded by DMDX via an Alesis Multimix8 USB 2.0 system, which acted as an external sound card and provided phantom power for a Countryman E6 Earset microphone. During the experiment, the microphone was worn on the participant's ear and a boom extending the capsule to the participant's mouth. The microphone's capsule was placed at approximately 2 cm from the corner of the participant's mouth. The participant sat at an approximate distance of 60 cm from the monitor.

Procedure

In an experimental session, the first author read aloud 400 words presented in random order. A session took about 60 minutes (including breaks) and was subdivided into four blocks of 100 stimuli each. Between each block, the first author pressed the space bar to continue. These interruptions provided the participant with information about his progress through the session. The participant completed a maximum of two sessions per day.

The first author read as quickly and as accurately as possible the word presented on the screen. A trial started with a centered fixation point '+' that was presented for 500 msec, followed by the target letter string, which remained on the screen until the participant responded or until two seconds had elapsed. A voice trigger was used to obtain naming latencies. Furthermore, all responses were recorded and processed with Praat (Boersma and Weenink, 2012) using a script implementing an intensity analysis to extract onset latency and the acoustic duration of the word read out.

2.2. Results

Mispronunciations, responses contaminated with coughs, and responses lost due to equipment failure, as well as words that are missing lexical distributional statistics are removed from the dataset. The resulting dataset contained 8875 compound words. We analyzed both the naming latencies and the acoustic durations of the words read out by the participant.

Table 1. *Generalized additive mixed model fitted to the naming latencies*

A. parametric coefficients	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
Intercept	6.3301	0.0212	298.1473	<0.0001
Phonological Neighborhood Size	0.0033	0.0014	2.3448	0.0191
B. smooth terms	**edf**	**Ref.df**	***F*-value**	***p*-value**
tensor Compound Frequency by Word	4.4913	5.2448	34.3124	<0.0001
tensor Left Family Size by Right Family	5.3862	6.1132	15.0522	< 0.0001
by-session random smooths for Trial	159.2157	215.0000	12.6853	< 0.0001
random intercepts tone first syllabeme	4.7843	5.0000	21.9638	< 0.0001
random intercepts tone second	3.4996	5.0000	1.8328	0.0449

random intercepts second syllabeme	2.3402	7.0000	1.5703	0.0037
random intercepts first place	4.6478	5.0000	23.5155	0.0259
random intercepts first manner	2.9848	3.0000	718.2719	< 0.0001

Naming latencies

We included three predictors characterizing the acoustic properties of the first segment to account for variation caused by these properties in the voice key measures (see e.g., Treiman et al., 1995): Phonation (with levels voiced and voiceless), Place of articulation (levels: alveolar, bilabial, glottal, labiodental, palatal, velar), and Manner of articulation (levels: approximant, fricative, nasal, plosive). The latter two factors were entered as random-effect factors into the model specification, in order to obtain a model that not only accounts for potential differences across factor levels, but that is also parsimonious in the number of parameters. The same considerations motivated bringing the tones (six in all) for the first syllabeme, and the tones for the second syllable (again six levels) into the model specification as random-effect factors. Because the way tone is written indeed adds orthographic complexity to the written word (tone *ngang* does not have a tone mark in the orthography, tone *nặng* has a visually smaller diacritic than tone *ngã* or tone *sắc*) we include tone as random-effect factors. Details about tone can be seen in the Discussion section. Naming latencies were log-transformed in order to remove most of the rightward skew in their distribution, thereby avoiding potential adverse effects of overly influential outliers.

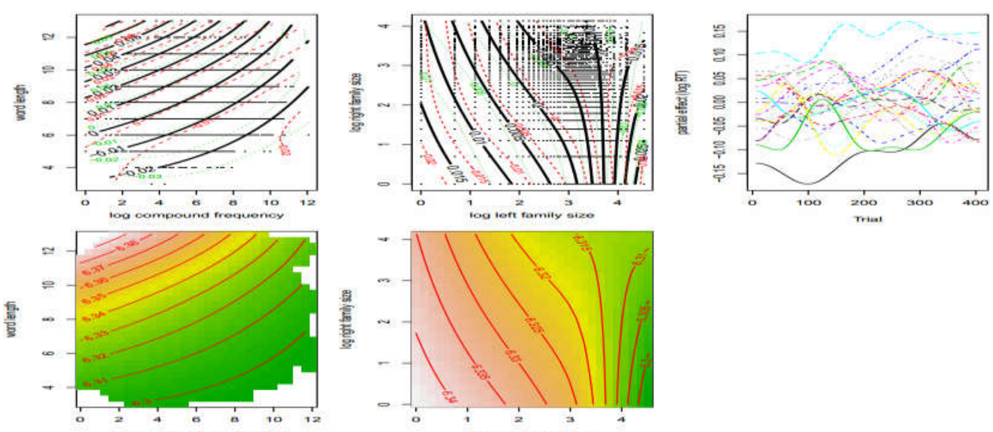

Figure 2: The GAMM for the naming latencies. Left panels: the tensor smooth for compound frequency by word length; central panels: the tensor smooth for left family size by right family size; right upper panel: the by-session random curves for trial; The upper left and central panels depict the partial effects with 1 standard error confidence regions around the contour lines. The corresponding lower panels show the corresponding fitted surface (predicted from all regressors in the model at their most typical values).

We fitted a generalized additive mixed model (GAMM) to the data, using the MGCV package (Wood, 2006, 2011) in the R programming environment (R Core Team, 2013). Generalized additive mixed models offer the analyst better possibilities for capturing autocorrelational structures in the data that arise in response latencies (and, as we shall see, in acoustic durations) across the sequences of trials in an experiment. Standard linear mixed-effects models (e.g., Pinheiro and Bates, 2000; Bates et al., 2013) allow the analyst to model subject-specific trends in the time series of responses by means of subject-specific intercepts and slopes. However, the actual trends tend not to be linear, but curvilinear. GAMMS provide factor smooths for curvilinear trends that are shrunk towards zero, effectively creating 'wiggly' trends over experimental time (Trial) that are truly random effects. GAMMS offer the researcher the additional possibility of bringing into the model specification that the errors are autocorrelated following a simple ARIMA (AR) process:

$$e_{t+1} = \rho e_t + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma), \quad -1 \leq \rho \leq 1. \tag{2}$$

A further advantage of generalized additive mixed models is that they offer improved means for modeling interactions between numeric predictors. The multiplicative interaction of the standard linear model assumes the fitted surface is part of a hyperbolic plane, which for many actual data is unrealistic. We therefore opted for tensor products to model the interplay of lexical predictors.

Table 1 summarizes the GAMM fitted to the naming latencies. The upper part of this table presents the parametric part of the model, with an intercept and a positive slope for phonological neighborhood size. Naming latencies increased for words with more phonological neighbors. There was no significant effect of First Phonation, thus voicing did not affect latencies measured.

The second half of the table presents the smooth terms in the model, including the effects of random-effect factors. An interaction of compound frequency by word length (in letters) was modeled with 4.6 effective degrees of freedom (*edf*) using a tensor product smooth. (Greater values for *edf* indicate greater degrees of wiggliness.) The upper left panel of Figure 2 shows the partial effect of these two predictors, with 1 SE confidence intervals around the contour lines. The corresponding lower panel presents the fitted surface with all other regressors in the model taken into account. As word length and compound frequency do not interact with other predictors, this surface is the same as the partial surface, except for an upward shift along the Z-axis that is mostly due to the intercept having been added to the predicted values. The tensor surface shows that word length has an inhibitory effect that is strongest for low-frequency compounds, and substantially reduced for high-frequency compounds. Conversely, (log-transformed) compound frequency is in general facilitatory, but its effect is very weak for short words, whereas its effect is most pronounced for long words.

The central panels of Figure 2 present the tensor smooth for the interaction of the family size of the left and right syllabemes. Across the full range of right syllable family sizes, we observe an effect of left syllable family size. However, although there is a facilitatory effect of right syllabeme family size for lower values of left syllable family size, this facilitation disappears for words with large left syllabeme family sizes.

The upper right panel of Figure 2 illustrates the intertrial dependencies at the different sessions during which data were collected, modeled with shrunk factor smooths. These factor smooths also take into account the modulations of the intercept across the different sessions - note that the session curves reach the vertical axis at different heights.

The last five rows of Table 1 clarify that the tones of the first and second syllabeme contribute to the fit of the model, and that the same holds for the controls for the voicekey, manner and place of articulation. An alternative to entering the tones of the two syllabemes as separate predictors is to include as random effect the pairs of tones realized on the compound's syllables. The resulting model is very similar both with respect to overall goodness of fit as with respect to the effects of the other predictors. Finally, the syllable type of the second (but not the first) syllable (CV, CVC, CwV, CwVC) reached significance. With a more fine-grained partition of syllable types that distinguishes between syllabemes with and without the first consonant (CV, V, CVC, VC, CwV, wV, CwVC, wVC), the random effect for the type of the first syllabeme also reaches significance. The token frequencies of the first and second syllabeme did not reach significance.

Inspection of the autocorrelation function of the errors revealed the errors were still autocorrelated even though the effect of trial was explicitly modeled. We therefore used an AR1 model for the residuals, with $\rho = 0.3$, thereby avoiding anti-conservative p-values. We checked for the presence of overly influential outliers by removing data points with absolute scaled residuals smaller than 2.5 SE. was inspected. As results remained highly similar, we reported the untrimmed full model.

Acoustic durations

Trials eliciting extreme word durations (less than 200 ms or exceeding 800 ms) were removed from the dataset (a loss of about 100 data points). The total number of words analysed is 8795.

Table 2 and Figure 3 summarize the GAMM fitted to the acoustic durations of the words read aloud. The upper part of Table 2 indicates that words beginning with a voiceless segment had a shorter duration by approximately 14 milliseconds. This is likely due to the difficulty in identifying the onset of the stop closure.

A second predictor with a linear effect is the frequency of the second syllable. Durations increased with second syllable frequency. The frequency of the first syllabeme did not reach significance, nor did the two frequency measures enter into an interaction.

Table 2. *Generalized additive mixed model fitted to the acoustic durations*

A. parametric coefficients	**Estimate**	**Std. Error**	***t*-value**	***p*-value**
Intercept	564.9062	28.3273	19.9421	< 0.0001
Left Syllabeme: Voiceless Initial Segment	-13.8162	1.8336	-7.5348	< 0.0001
Frequency Second Syllabeme	2.9582	1.0859	2.7243	0.0065
B. smooth terms	**edf**	**Ref.df**	***F*-value**	***p*-value**
tensor Comp. Freq by Word Length	7.2248	9.1284	25.3215	< 0.0001
smooth Phonological Neighborhood Size	3.6078	4.2956	9.5478	< 0.0001
tensor left Fam by right Fam Size	5.8543	7.3901	6.4407	< 0.0001
by-session random smooths for Trial	75.6504	215.0000	4.2266	< 0.0001
random intercepts tone left syllabeme	4.7723	5.0000	35.6649	< 0.0001
random intercepts tone right syllabeme	4.9941	5.0000	1778.4508	< 0.0001
random intercepts syllable type left syllb.	2.1840	3.0000	4.4002	0.0526
random intercepts syllable type right syllb.	2.9633	3.0000	871.7621	< 0.0001
random intercepts left place	4.8771	5.0000	410.7771	< 0.0001
random intercepts left manner	2.9414	3.0000	1373.8096	< 0.0001

The first tensor product in the sub-table of smooth terms concerns the interaction of compound frequency and word length. The partial effect and the corresponding fitted surface are shown in the left panels of Figure 3. For low-frequency compounds, the effect of word length is strong, but for high-frequency words, its effect is minimal. The effect of word frequency is facilitatory across the board, but most clearly so for the longer words, and least for short words.

The family size measures for the left and right syllabeme entered into a non-linear interaction visualized in the second column of Figure 3. Durations are longest when both family sizes are small, and shortest when both are large. The lower right panel of Figure 3 illustrates that words with more phonological neighbors are realized with longer acoustic durations. The effect is virtually non-existent for the lower counts, but manifests itself clearly for higher neighborhood sizes. The upper right panel shows how acoustic durations change as the speaker proceeded through a session. As for the response latencies, we observe substantial local consistency, and global variation, as a session unfolds.

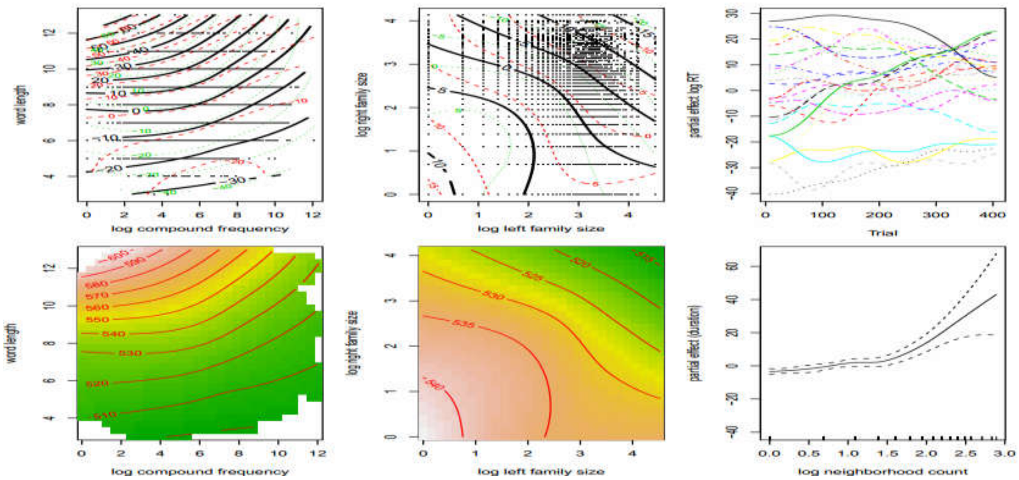

Figure 3: Non-linear effects in the GAMM fitted to the acoustic durations. Left panels: the interaction of frequency by word length; center panels: the interaction of left and right family size; upper right panel: the by-session random curves for trial, lower right panel: the effect of phonological neighborhood size.

Inspection of the autocorrelation function of the errors revealed no further autocorrelation in the residuals. A model with potentially overly influential outliers (with absolute scaled residuals smaller than 2.5 SE) was inspected. As results were virtually identical, the original untrimmed model is reported here.

3. Discussion

For both onset latencies and acoustic durations, we observed an effect of compound frequency in interaction with word length. The interaction of the two predictors was somewhat stronger for the acoustic durations. Nevertheless, for both response variables, long low-frequency words emerged with the longest naming latencies as well as the longest acoustic durations. The facilitatory effect of compound frequency for the naming latencies fits well with other studies on compound processing (e.g., Baayen, 2010; De Jong et al., 2002; Juhasz et al., 2003; Kuperman et al., 2008, 2009), and has been interpreted as evidence for whole-word lexical representations.

Some recent studies (see e.g., Adelman et al., 2006; Brysbaert and New, 2009; Plummer et al., 2013) reported words' contextual diversity (known as dispersion in lexical statistics) to be an important factor for gauging how well words are entrenched in memory. Adelman et al. (2006) observed for English that when frequency is residualized on contextual diversity, frequency is no longer a significant predictor. We failed to replicate this effect for Vietnamese. Adding dispersion as a predictor does not improve the fit of the model to the data. Frequency can be replaced by dispersion, but this does not lead to a better fit either - in fact, the quality of the fit decreases slightly. It is conceivable that measures more sensitive to contextual co-occurrence (MacDonald and Shillcock, 2002; Baayen, 2010) may provide superior predictivity.

Surprisingly, we were not able to document any reliable frequency effects tied to the constituent syllabeme for the response latencies. Replacing syllabeme frequency by syllabeme dispersion led to inferior model fits, with no support whatsoever for a frequency effect for the second syllabeme, and a non-significant trend for facilitation from the first syllable's dispersion ($p = 0.0885$).

A syllabeme frequency effect is present for the acoustic durations, but only for the second syllabeme, such that as the frequency of the second syllabeme increased so did the duration. For English, a decrease in acoustic duration has been reported to go hand in hand with increasing lexical frequency (Aylett and Turk, 2004). This has led to the hypothesis that informationally redundant words (i.e., high-frequency words) would be shortened to maintain a uniform flow of information per time unit (the smooth signal hypothesis). Counterevidence for this hypothesis has been reported for Dutch interfixes by Kuperman et al. (2006) and for the German high front vowel by Tomaschek et al. (2013). The present study adds further evidence that the relationship between frequency and acoustic duration may not be straightforward.

However, it is possible that second syllabeme frequency is a sub-optimal measure. One might think a dispersion based variant would perform better. Unfortunately, the correlation of second syllabeme frequency and second syllabeme dispersion is extremely high ($r = 0.96$). When added to the model reported above, it helps increase the goodness of fit, but with a sign opposite to that of second syllabeme frequency, indicating suppression (Friedman and Wall, 2005). Furthermore, second syllabeme frequency remains significant with a substantially smaller standard error compared to the dispersion measure. Thus, it is unlikely that changing from frequency to dispersion is at all helpful.

The frequency of the second syllabeme is also well correlated with second syllabeme family size ($r = 0.69$). Importantly, when the family size effects are witheld from the model specification, the second syllabeme frequency is no longer significant. When second syllabeme frequency is witheld, most of its effect is absorbed by the right syllabeme family size, especially for low first syllabeme family size values, with a minimal decrease in goodness of fit. We therefore refrain from interpreting the second syllabeme frequency effect.

For both naming latencies and acoustic durations, solid syllable effects are present in the form of family size effects, thereby extending the evidence for the relevance of this lexical distributional measure from Indo-European, Semitic, and Finno-Ugric (Moscoso del Prado Martín et al., 2004) to the Austro-Asiatic language family. In terms of network science (Jungnickel, 2007; Baayen, 2010), the family size of the first syllabeme is its out-degree, and the family size of the second syllabeme is its in-degree in the Vietnamese directed compound graph. The main pattern is that a greater degree (evidence for a more important status as a 'hub' in the network) affords both shorter naming latencies and shorter acoustic durations. For the present data, the centrality of a syllabeme as a hub in the lexical network appears to be a much more solid predictor than its bare frequency of occurrence.

However, it is conceivable that a cumulative frequency measure for syllables, counting not how often a syllable occurs as an independent word, but how often it is used as part of a compound, would capture at least part of the variance currently explained by the family size measures. Earlier work on Dutch (Schreuder and Baayen, 1997) suggested the cumulative frequency of a constituent across compounds to have little independent predictivity, but we need to establish that this holds as well for Vietnamese. We calculated the left and right cumulative syllabeme frequency for all of the compounds in our data set, aggregating for each syllabeme the frequency with which it occurs in a compound. When we added the log-transformed cumulative syllabeme frequency counts as predictors to the model, they failed to reach significance, whereas the family size measures remained fully significant. This supports the conclusions reached by Schreuder and Baayen (1997) that for morphologically related words, it is type counts that are predictive, and not token counts.

What the present data indicate is that for word naming latencies, and for acoustic durations as well, the frequencies of a compound's constituent syllabemes are at best weakly predictive. The absence of solid support for syllable frequency effects in naming latencies contrasts markedly with the strong constituent frequency effects in word naming reported for the first constituent of compounds in English (Baayen et al., 2010). Thus, the role of the syllabeme as a constituent appears to be substantially reduced in Vietnamese compared to English. Further evidence for such a difference can be gleaned from a prior study on lexical processing in Vietnamese, using the visual lexical decision task. Pham and Baayen (2015) reported a principal components regression analysis indicating that a latent variable representing both syllable frequency and family size predicted longer response latencies. To disentangle the distinct effect of syllable family size and frequency, we conducted a supplementary analysis of their data, which suggested strong inhibition for syllable frequency and mild facilitation for syllable family size. This inhibitory constituent frequency effect in visual lexical decision also stands in marked contrast to the facilitation typically observed for English (Baayen et al., 2010) and Dutch (Kuperman et al., 2008, 2009). This result plus the whole compound frequency effect then supports a processing model where these compounds are being processed as wholes.

Why is constituent frequency inhibitory in Vietnamese lexical decision and non-predictive in Vietnamese word naming, whereas in English and Dutch, facilitatory effects of constituent frequency are well attested (see, e.g., Baayen et al., 2010)? We think that a marked cross-linguistic difference in the relative frequency of compound and constituent frequencies is at issue here. For the current set of 8875 Vietnamese compounds, 72% have a compound frequency exceeding the frequency of the left syllabeme. The corresponding percentage for the right syllabeme is 73%. By contrast, using the data set of 1252 English compounds studied by Baayen (2010), the corresponding percentages are an order of magnitude smaller at a mere 0.9% and 0.8%. What we observe here is a fundamental difference in the markedness of compounding. In English, simple words are the central, unmarked, onomasiological units, whereas compounds present the marked case. In Vietnamese, two-syllabeme compounds constitute the unmarked word formation pattern, whereas single syllabeme words instantiate the marked, less frequently used pattern. Given that syllabemes are both syllables and lexemes (in the sense of Aronoff, 1994), we can equivalently describe the preferential phonological form of a Vietnamese word as bi-syllabic.

In the visual lexical decision task, single syllabeme words are low-frequency competitors of the compounds containing them, and as such slow lexical decisions. The reason why constituent frequency effects disappear in word naming is likely to be due to the nature of the word naming task, which is less sensitive to meaning compared to the lexical decision task (see, e.g., Baayen et al., 2006). This explanation implies that the family size effect for compounds' constituents in Vietnamese, which has also been observed for English word naming (Baayen et al., 2010), must reflect familiarity with the articulation of these constituents in speech production, rather than semantic co-activation of family members. This interpretation of the family size effect fits well with the analysis of the acoustic durations. Words with larger morphological families were pronounced shorter, whereas words with many phonological neighbors were realized with longer acoustic durations.

Table 3. *Posterior modes for tone by syllabeme and task*

	1st Syllb: Duration	1st Syllb: Latency	2nd Syllb: Duration	2nd Syllb: Latency
ngang	-1.2073	0.0105	47.6179	-0.0013
huyền	9.4485	-0.0027	61.8613	0.0008
hỏi	-5.4276	-0.0159	13.1098	-0.0040
ngã	12.5321	-0.0085	-23.9459	0.0041
sắc	-8.3506	0.0096	-7.3964	-0.0006
nặng	-6.9951	0.0070	-91.2466	0.0009

Table 3 summarizes the effects of the six tones of northern Vietnamese (see e.g., Đoàn, 1977; Nguyễn and Edmondson, 1998, for further details): (1) *ngang* mid level, (2) *huyền* low falling (breathy), (3) *hỏi* mid falling(-rising), harsh, and (4) *ngã* mid rising, glottalized, (5) *sắc* mid rising, tense, and (6) *nặng* mid falling, glottalized, short. The table presents a cross-classification by response variable (latency versus duration) and position (first versus second syllabeme). There are substantial changes in duration of the same tone depending on whether it is realized on the first or on the second syllabeme. For the second syllabeme, we find more extensive lengthening (see, e.g., tone 2) as well as more extensive shortening (see, e.g., tone 6). This suggests a tight relation between vowel duration and tone (cf. Cao, 1962). Thus, tone *nặng* has a steep fall that is executed quickly and is often glottalized or followed by a glottal stop, resulting in a short vowel duration. The greater effect on duration for the second syllabeme is supported by a paired *t*-test on the absolute magnitudes of the effects ($p = 0.055$).

Interestingly, the effects of the tones on the latencies present a mirror image. Here, effects are large for the initial syllabeme, and an order of magnitude smaller for the second syllabeme ($p = 0.005$, paired *t*-test). These results provide an important constraint on models of speech production. According to the WEAVER^{++} model of Levelt et al. (1999), the production of a word would proceed morpheme by morpheme, and within a word, segment by segment. Various experiments (e.g., Cholin et al., 2004, 2006) using the implicit priming paradigm carried out on Dutch suggested that knowledge of units later on in a sequence does not afford shorter production onsets. The present experiment on a non Indo-European language provides unambiguous evidence that the onset of articulation is co-determined by the tone to be realized on the second syllabeme. The effects are small, but sufficiently well-supported by the GAMM. This indicates that these compounds are processed as a whole.

This effect fits well with the facilitating effect of the family size of the second syllabeme on the naming latency. As can be seen in Figure 2, the magnitude of the family size of the left constituent is larger than that of the right family size (which is even entirely absent for words with large left syllable families). As the initial syllabeme has to be articulated first, a strong effect of its family size effect is unsurprising. When the initial syllabeme is used less often in compounds, we do find a small effect of the family size of the second syllabeme, consistent with the small effect of the tone of the second syllabeme.

The importance of properties of the second syllabeme for the onset of articulation of the compound receives further support from the effect of the syllable type of the second syllabeme. As can be seen in Table 4, second syllabemes ending in a consonant give rise to elongated naming latencies. For word durations, the magnitudes of the effects of the syllable type are much larger for the second syllable than for the first. Table 4 illustrates that acoustic durations are longer for vowel-final second syllables, and reduced for second syllable ending in a consonant. It is likely that part of this effect is due to the fact that in Vietnamese final stops are unreleased. As a consequence, there is little information in the acoustic signal to be used to determine the end of the closure gesture, thus the end of the word would be marked at the onset of the closure. The segmenter will take the beginning of the closure as the end of articulation, whereas in reality, the gesture for the closure may extend considerably in time before a (burstless) release takes place. Therefore, the durational differences observed for the second syllable are confounded with inevitable inaccuracies of automatic word segmentation of speech files. Note that for the initial syllable, syllable final consonants do not present a problem for the segmenter, as the initial consonant in the onset of the second syllable provides sufficient information to properly delimit the end of the preceding consonant.

Table 4. *Posterior modes for syllable type by syllabeme and task*

	2nd Syllabeme: Latency	1st Syllabeme: Duration	2nd Syllabeme: Duration
CV	-0.0000	2.7996	14.6455
CVC	0.0009	3.0161	-26.1235
CwV	0.0001	-4.4389	27.3200
CwVC	-0.0043	-1.3768	-15.8421

Both the analysis of the response latencies and the analysis of the acoustic durations provide support for an effect of the count of phonological neighbors of the compound. The effect of neighborhood density on the durations is substantial (on the order of 40 ms, see Figure 3), whereas the effect on the naming latencies is tiny (on the order of 5 ms). The inhibitory effect of neighborhood density in the response latencies is consistent with the inhibitory effect reported for this measure for picture naming by Sadat et al. (2014). The effect of neighborhood density on the acoustic durations, however, is different from the effect of neighborhood density reported by Gahl et al. (2012) for English. These authors show for English that words with many neighbors are easier to produce, and hence afford phonetic reduction. It is unclear whether similar considerations would pertain to Vietnamese, which, unlike English, is not a stress-timed language but a syllable-timed language (see e.g., Roach, 1982; Cummins et al., 1999; Romano et al., 2011) in which reduction phenomena linked to morphologically determined syllable structures in stress-timed languages such as Dutch or English (Kemps et al., 2005a,b) are unlikely to occur. Our results indicate that in Vietnamese, a large neighborhood density gives rise to substantial durational enhancement instead of durational shortening.

In a lexical decision task, with a superset of the present words and the same participant (Pham and Baayen, 2015), the neighborhood count had a facilitatory effect ($\hat{\beta} = -0.0192$, $p < 0.0001$).

The joint evidence of the two experiments indicates that facilitation arises in lexical decision due to neighbors providing evidence for lexicality, whereas in tasks in which a given target has to be selected for articulation, the same neighbors become competitors, delaying onset of articulation, and leading to prolonged acoustic durations. One reason why in speech production large neighborhoods give rise to a processing disadvantage could be that there are strong phonotactic restrictions on syllable structure. The four syllable types discussed above (CV, CVC, CwV, CwVC), in combination with a phoneme inventory with twenty-two consonant onsets, thirteen vowels, three diphthongs, one glide, and a restricted set of eight consonants for the coda position, allow for only a relatively limited number of syllables, and hence a dense lexical space. Preparing an articulatory pathway through this dense space, and maintaining that pathway against the alternative pathways of lexical neighbors, is likely to be time-costly, leading to both delayed onset of production, and to longer acoustic durations.

Given the analyses of the acoustic durations and the response latencies, the question remains to what extent durations and latencies are correlated. It turns out that there is a small but significant negative correlation between the two ($r = -0.05$, $t(8784) = -5.13$, $p < 0.0001$). The more time spent on preparation for articulation, the faster the execution of the articulatory programs can take place.

4. Concluding remarks

For the study of speech production, the picture naming task has the advantage of using a stimulus (a picture) that is non-linguistic in nature. By contrast, the word naming task combines a linguistic comprehension task (reading a word) with a speech production task (subsequently saying the word out. A disadvantage of the picture naming task is that many words are not, or not easily, depictable. The word naming task does not suffer from this disadvantage. For gauging speech production processes, however, the naming latency captures only the time required for preparing articulation. In addition to the naming latency response variable, this study considers an important factor as a second response variable, which has often been disregarded in naming studies, namely the acoustic duration of the speech produced.

With a coverage of nearly 9,000 words, we observed some remarkable similarities, as well as some subtle differences with findings in studies on English and Dutch. Words with larger phonological neighborhoods required longer response preparation times, and were articulated with longer acoustic durations. Lower-frequency longer words also demand longer preparation, and come with elongated durations. Furthermore, the greater the family sizes of the left and right syllabemes are, the shorter the time required for preparing for articulation, and the shorter a compound's acoustic duration can be. The convergence of the GAMMS for latency and duration indicates that the word naming task captures important aspects of the later stages of the speech production process with remarkable fidelity.

Yet, some subtle differences are worth pointing out. The effect of phonological neighbors appears to be driven more by higher numbers of neighbors when it comes to acoustic durations, as evidenced by the non-linear positive accelerating functional form of its partial effect. The effect of the second syllable's family size is much stronger for the acoustic durations than for the response latencies, indicating that the response latencies are dominated by processes preparing the articulation of the initial syllable. In addition, the effects of length and compound frequency (and their interaction) appear to be somewhat stronger for the acoustic durations. It is conceivable that the importance of these effects for the production process is underestimated when querying response latencies, due to confounding with the preceding comprehension process. Finally, for reasons that are as yet unclear to us, the effect of the tone realized on the first syllable is large for the response latency and small for acoustic duration, whereas for the tone realized on the second syllabeme, the reverse holds. Further research will have to clarify whether there are asymmetries in these effects with respect to the acoustic durations of the constituent syllabemes. Irrespective of how the effects of tone are to be understood, it is clear that even during the planning of the articulation of the first syllabeme, the properties of the second syllabeme are being taken into consideration, challenging the speech production model of Levelt et al. (1999), but consistent with the literature on co-articulatory effects in speech (Nittrouer and Studdert-Kennedy, 1987; Fowler and Saltzman, 1993).

(1) *This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 602.10-2016.05.*

REFERENCES

1. Adelman, J. S., Brown, G. D. A., and Quesada, J. F. (2006), *Contextual diversity, not word frequency, determines word-naming and lexical decision times.* Psychological Science, 17(9):814–823.

2. Aylett, M. and Turk, A. (2004), *The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech.* Language and Speech, 47:31–56.

3. Baayen, R. H. (2010), *The directed compound graph of English. An exploration of lexical connectivity and its processing consequences*. In Olsen, S., editor, New impulses in word-formation (Linguistische Berichte Sonderheft 17), pages 383-402. Buske, Hamburg.

4. Baayen, R. H., Kuperman, V., and Bertram, R. (2010), *Frequency effects in compound processing*. In Scalise, S. and Vogel, I., editors, Cross-Disciplinary Issues in Compounding, pages 257–270. Benjamins, Amsterdam/Philadelphia.

5. Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007), *The English Lexicon Project*. Behavior Research Methods, 39(3):445-459.

6. Bates, D., Bolker, B., Maechler, M., and Walker, S. (2013), *lme4: Linear mixed-effects models using Eigen and S4*. Cran R Project.

7. Bell-Berti, F. and Harris, K. S. (1979), *Anticipatory coarticulation: Some implications from a study of lip rounding*. The Journal of the Acoustical Society of America, 65(5):1268–1270.

8. Bien, H., Baayen, R. H., and Levelt, W. M. J. (2011), *Frequency effects in the production of Dutch deverbal adjectives and inflected verbs*. Language and Cognitive Processes, 27:683–715.

9. Cao, X. H. (1962), *Bàn về giải thuyết âm vị học một số vận mẫu có nguyên âm ngắn trong tiếng Việt [Discussion on the phonological hypothesis on some rimes containing short vowels in Vietnamese*]. Thông báo khoa học Đại học tổng hợp Hà Nội, 2:146-154.

10. Cholin, J., Levelt, W. J., and Schiller, N. O. (2006), *Effects of syllable frequency in speech production*. Cognition, 99(2):205 – 235.

11. Đoàn, T. T. (1977), *Ngữ âm tiếng Việt [Vietnamese phonetics]*. Đại học và Trung học chuyên nghiệp, Hà Nội.

12. Forster, K. I. and Forster, J. C. (2003), *DMDX: A windows display program with millisecond accuracy.* Behaviour Research Methods, Instruments, and Computers, 35(1):116–124.

13. Friedman, L. and Wall, M. (2005), *Graphical views of suppression and multicollinearity in multiple regression*. The American Statistician, 59:127-136.

14. Gahl, S., Yao, Y., and Johnson, K. (2012). *Why reduce? phonological neighborhood density and phonetic reduction inspontaneous speech*. Journal of Memory and Language, 66(4):789-806.

15. Janssen, N., Bi, Y., and Caramazza, A. (2008), *A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds.* Language Cognitive Processes, 23(7):1191–1223.

16. Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003), *The effects of morphology on the processing of compound words: Evidence from lexical decision, naming, and eye fixations*. British Journal of Psychology, 94:223-244.

17. Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2006*), Morphological predictability and acoustic salience of interfixes in Dutch compounds*. *JASA*, 122:2018-2024.

18. Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999), *A theory of lexical access in speech production*. Behavioral and Brain Sciences, 22(1):1-38; discussion 38-75.

19. Mok, P. K. (2010), *Language-specific realizations of syllable structure and vowel-to-vowel coarticulation*. The Journal of the Acoustical Society of America, 128(3):1346-1356.

20. Moscoso del Prado Martín, F., et al. (2004), *Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew*. Journal of Experimental Psychology: Learning, Memory and Cognition, 30:1271-1278.

21. Pham, H. and Baayen, H. R. (2015), *Vietnamese compounds show an anti-frequency effect in visual lexical decision*. Language, Cognition & Neuroscience, 30(9):1077-1095.

22. Plummer, P., Perea, M., and Rayner, K. (2014), *The influence of contextual diversity on eye movements in reading*. Journal of experimental psychology*,* 40(1):275-83.

23. Sadat, J., et al. (2014), *Reconciling phonological neighborhood effects in speech production through single trial analysis*. Cognitive Psychology, 68(0):33-58.

24. Schreuder, R. and Baayen, R. H. (1997), *How complex simplex words can be*. Journal of Memory and Language, 37(1):118-139.

25. Tomaschek, F., Wieling, M., Arnold, D., and Baayen, H. (2013), *Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience*. Language and Speech. To appear in Interspeech 2013.

26. Treiman, R., Mullennix, J., Bijeljac-Babic, R., and Richmond-Welty, E. D. (1995), *The special role of rimes in the description, use, and acquisition of English orthography*. Journal of Experimental Psychology: General, 124(2):107-136.

Hiệu ứng hình thái học trong việc đọc thành lời từ ghép tiếng Việt

Tóm tắt: Nghiên cứu này báo cáo kết quả thực nghiệm ghi nhận quá trình tạo sản ngôn ngữ qua việc đọc từ ghép tiếng Việt. Thay vì chỉ khảo sát biến hồi đáp là độ trễ đọc từ, chúng tôi còn khảo sát biến hồi đáp là trường độ âm học của lời nói được phát ra. Các hiệu ứng tần số của từ ghép, độ dài của từ, và kích cỡ nhóm thành tố cấu tạo từ đều được ghi nhận trong cả độ trễ và trường độ, tuy nhiên, các hiệu ứng tần số thành tố cấu tạo từ thì không được ghi nhận. Điều này cho thấy tiếng Việt khác biệt so với tiếng Anh và tiếng Hà Lan, trong hai ngôn ngữ đó các hiệu ứng tần số của thành tố đã được kiểm chứng. Chúng tôi cho rằng sự vắng mặt hiệu ứng tần số của thành tố cấu tạo từ là do các cấu trúc song tiết tạo nên dạng thức âm vị học không đánh dấu cơ bản của từ tiếng Việt. Dữ liệu của chúng tôi cũng đòi hỏi các mô hình sản sinh lời nói phải chấp nhận rằng khởi âm của việc tạo sản lời nói có thể không bị ảnh hưởng bởi các thuộc tính của các âm tiết có âm đầu là âm tắc thanh hầu, vì chúng tôi đã quan sát thấy độ trễ trong việc đọc từ được xác định qua độ lớn của nhóm, thanh điệu và kiểu loại âm tiết của âm tiết thứ hai. Sự hội tụ đáng kể của các hiệu ứng tần số và độ lớn của nhóm đối với độ trễ hồi đáp và trường độ âm học xác nhận tác nghiệm đọc từ bao gồm hai biến hồi đáp này như là một hệ hình thực nghiệm xuất sắc cho việc nghiên cứu mã hoá âm vị học trong tạo sản lời nói.

Từ khoá: đọc từ vựng; hiệu ứng tần số; mật độ từ vựng; trường độ âm học; tiếng Việt; nghiên cứu dữ liệu lớn; tạo sản lời nói.