Bilingualism: Language and Cognition

cambridge.org/bil

Phonological transfer effects in novice learners: A learner's brain detects grammar errors only if the language sounds familiar

Sabine Gosselke Berthelsen[1], Merle Horne[1], Yury Shtyrov[2,3,4] and Mikael Roll[1]

[1]Centre for Languages and Literature, Lund University, Lund, Sweden; [2]Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark; [3]Institute of Cognitive Neuroscience, HSE University, Moscow, Russia and [4]Sirius University of Science and Technology, Sochi, Russia

Research Article

Cite this article: Gosselke Berthelsen S, Horne M, Shtyrov Y, Roll M (2021). Phonological transfer effects in novice learners: A learner's brain detects grammar errors only if the language sounds familiar. *Bilingualism: Language and Cognition* **24**, 656–669. https://doi.org/10.1017/S1366728921000134

Received: 30 June 2020
Revised: 18 November 2020
Accepted: 8 March 2021
First published online: 12 April 2021

Keywords:
rapid second language acquisition (SLA); L2 error processing; transfer; grammar; ERP; N400; pictures; morphological tone

Address for correspondence:
Sabine Gosselke Berthelsen, Center for Languages and Literature, Lund University, Box 201, 22100 Lund, Sweden
sabine.gosselke_berthelsen@ling.lu.se

Abstract

Many aspects of a new language, including grammar rules, can be acquired and accessed within minutes. In the present study, we investigate how initial learners respond when the rules of a novel language are not adhered to. Through spoken word-picture association-learning, tonal and non-tonal speakers were taught artificial words. Along with lexicosemantic content expressed by consonants, the words contained grammatical properties embedded in vowels and tones. Pictures that were mismatched with any of the words' phonological cues elicited an N400 in tonal learners. Non-tonal learners only produced an N400 when the mismatch was based on a word's vowel or consonants, not the tone. The emergence of the N400 might indicate that error processing in L2 learners (unlike canonical processing) does not initially differentiate between grammar and semantics. Importantly, only errors based on familiar phonological cues evoked a mismatch-related response, highlighting the importance of phonological transfer in initial second language acquisition.

Introduction

Second language learners can acquire many aspects of a new language (L2) at surprisingly fast rates. In both naturalistic and artificial acquisition settings, novice learners show phonological (e.g., Gullberg, Roberts, Dimroth, Veroude & Indefrey, 2010), lexicosemantic (e.g., Dittinger, Scherer, Jäncke, Besson & Elmer, 2019; Gullberg et al., 2010; Kimppa, Kujala, Leminen, Vainio & Shtyrov, 2015), and even grammatical L2 knowledge (e.g., Cunillera et al., 2009; de Diego-Balaguer, Toro, Rodríguez-Fornells & Bachoud-Lévi, 2007; Gosselke Berthelsen, Horne, Shtyrov & Roll, 2020) after mere minutes of language contact or training. An important facilitating factor during such early stages of second language acquisition (SLA) is transfer. To this effect, initial learners often acquire words with native language (L1) morphology and phonology more accurately and quickly than words with unfamiliar morphological and phonological features (Havas, Taylor, Vaquero, de Diego-Balaguer, Rodríguez-Fornells & Davis, 2018; McKean, Letts & Howard, 2013). This is likely due to a more comprehensive neural processing of L1-like novel words which can draw on fine-tuned L1 networks. This claim finds support in electroencephalographic (EEG) studies which show that only words with familiar phonology induce rapid automatic word assessment and other pre-attentive processing mechanisms in initial learners (Gosselke Berthelsen et al., 2020; Kimppa et al., 2015). In the present paper, we extend the investigation of both initial L2 processing and the role of transfer by examining how novice learners initially process visually presented referents that mismatch lexicosemantic or morphosyntactic properties of acoustically presented words in a new language context. We investigate whether the learners respond differently to mismatches that are based on familiar and unfamiliar phonological cues.

Lexicosemantic and grammatical errors in L1 and L2

When a native speaker encounters a lexicosemantic or grammatical mismatch in the incoming language stream, their brain reacts to these errors with characteristic responses that can easily be observed using EEG measurements. A lexicosemantic error, for instance, will increase the amplitude of the N400 component. The N400 is a natural response to meaningful stimuli measured over the posterior part of the scalp. Its amplitude is in an inverse relation to the expectancy of an encountered stimulus with respect to context and world knowledge (Ganis, Kutas & Sereno, 1996; Kutas & Federmeier, 2014; Kutas & Hillyard, 1980). Hence, the component is particularly enhanced for lexicosemantic errors and can serve as an indicator of lexical and semantic accuracy. Sometimes, the N400 has even been observed for morphosyntactic manipulations (Barber & Carreiras, 2005; Guarjardo & Wicha, 2014). In those cases, the morphosyntactic irregularities are either embedded in contexts that also include prominent semantic

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

manipulations (Guarjardo & Wicha, 2014) or constructions with a strong semantic load (i.e., N400 for noun-postnominal adjective agreement vs. left anterior negativity [LAN] for determiner-noun agreement, Barber & Carreiras, 2005). Most often, however, morphosyntactic processing is distinct from semantic processing. It typically evokes one or several of a range of sequential event-related potentials (ERPs): early left anterior negativity (ELAN), LAN, and/or P600. These components are enhanced and thus most easily elicited in contexts where specific parts of morphosyntactic processing are manipulated so as to fail, i.e., in error contexts. The suggested functions of the different morphosyntactic components are as follows: the ELAN, with a latency of ~150 ms, is believed to signal the assessment of a word's grammatical category, presumably based on the initial activation of its morphosyntactic features (e.g., Friederici, 2002; Gosselke Berthelsen et al., 2020; Neville, Nicol, Barss, Forster & Garret, 1991). The LAN, subsequently, is similar to the lexicosemantic N400 in timing, but not distribution. It is largest at frontal electrodes, often left-lateralised, and is likely reflective of assessment of, for instance, gender or number relations (Molinaro, Barber, Caffarara & Carreiras, 2014; Roll, Gosselke, Lindgren & Horne, 2013; Schremm, Novén, Horne & Roll, 2019). This assessment is likely made possible by a rule-based integration of morphosyntactic and semantic content (Gosselke Berthelsen et al., 2020; Krott & Lebib, 2013). The P600, finally, is a positive response strongest at posterior electrodes at around 600 ms. It reflects utterance-level integration processes and is enhanced in contexts of encountered errors (morphosyntactic and sometimes semantic) due to a need of revision or repair (Kim & Osterhout, 2005; Osterhout & Holcomb, 1992; Roll, Horne & Lindgren, 2007). Given their different but often complementary assumed functions, the three major morphosyntactic components can occur together or separately. For the most part, they stand in clear contrast to the lexicosemantic N400.

Error processing has also been studied in second language learners. With respect to lexicosemantic mismatches in the L2, learners resemble native speakers and produce an N400 already at low proficiency levels. With respect to morphosyntactic processing, however, learners differ profoundly from native speakers. Rather than an ELAN, LAN, or P600 for grammatical irregularities, learners often produce an N400, especially at low proficiency levels (McLaughlin et al., 2010; Tanner, McLaughlin, Herschensohn & Osterhout, 2012). The N400 initially acts as a second language learner's default response to any linguistic inconsistency, (morpho)syntax included. During continued acquisition, language-specific regularities gradually become grammaticalised in the learner's brain (Steinhauer, 2014): at intermediate proficiencies, learners produce P600 responses (e.g., Tanner et al., 2012; Tokowicz & MacWhinney, 2005), followed by the LAN at high proficiency levels (e.g., Bowden, Steinhauer, Sanz & Ullman, 2013; Gillon Dowens, Vergara, Barber & Carreiras, 2010; Ojima, Nakata & Kakigi, 2005), and, very rarely, an ELAN (e.g., Hanna, Shtyrov, Williams & Pulvermüller, 2016; Hed, Schremm, Horne & Roll, 2019; Rossi, Gugler, Friederici & Hahne, 2006). Evidently, however, language processing can be accelerated dramatically in targeted experimental settings. Here, learners frequently show both lexicosemantic and morphosyntactic ERP effects with minimal instruction or exposure. The order in which the ERP components emerge stays the same. Yet, the P600 emerges already after hours or even minutes of training in experimental settings (Batterink & Neville, 2013; Davidson & Indefrey, 2009; de Diego-Balaguer et al., 2007; Friederici, Steinhauer & Pfeifer, 2002; Grey, Sanz, Morgan-Short & Ullman, 2018; Havas, Laine & Rodríguez-Fornells, 2017; Mueller, Hahne, Fujii & Friederici, 2005). The same holds true for the LAN and possibly the ELAN which, however, seem to depend more on intermitted consolidation periods (Friederici et al., 2002; Mueller, Hirotani & Friederici, 2007; Morgan-Short, Steinhauer, Sanz & Ullman, 2012). While a recent artificial learning study found an ELAN and a LAN without consolidation, this may well be explained by the fact that this study analysed canonical grammar processing rather than error processing (Gosselke Berthelsen et al., 2020). Collectively, the findings for error processing in L2 learners suggest that lexicosemantic error processing appears to precede grammatical processing and that the latter proceeds gradually. Learners can process L2 words like native speakers but it either takes time or specific, experimental focus for the error-related ERP responses to emerge. In experimental settings, language processing can emerge rapidly, even in initial learners and often, in fact, before behavioural performance transcends chance levels (cf., McLaughlin, Osterhout & Kim, 2004; Hed et al., 2019). However, some grammar-related ERP components (ELAN/LAN) have thus far only been observed after consolidation periods, at least in violation paradigms.

Transfer in the early stages of SLA and word-level tone

One factor that appears critical for how rapidly and naturally novel items and rules can be acquired and processed is transfer from the learner's native language on the basis of L1-L2 similarity. Transfer is relevant on many linguistic levels, including syntax (e.g., Foucart & Frenck-Mestre, 2011; 2012) and phonology (e.g., Gosselke Berthelsen et al., 2020; Kimppa et al., 2015). Transfer tends to affect early, pre-conscious processing more than late, post-conscious processing or offline responses (e.g., Andersson, Sayehli & Gullberg, 2019; Gillon Dowens et al., 2010; Gillon Dowens, Guo, Guo, Barber & Carreiras, 2011). With respect to phonological similarity, only novel words that are in accordance with the learners' L1 rules for sounds and sound structure are assessed and processed pre-attentively by initial learners. This has been found for a number of different pre-attentive responses: an early lexicosemantic gating component at ~50 ms, for instance, shows rapid lexical trace formation only for words that have L1-like phonotactic (Kimppa et al., 2015) or prosodic structure (Gosselke Berthelsen et al., 2020). For words where L1 transfer is not possible, the lexicosemantic gating component is unaffected. Similarly, the fronto-central N1, strongly involved in automatic auditory processing, is significantly reduced for phonologically illegal pseudowords but not for legal pseudowords (Silva, Vigário, Leone Fernandez, Jerónimo, Alter & Frota, 2019). This may suggest that novel words with illegal phonology are not processed as words, but as unusual sound patterns. Finally, even pre-conscious grammar processing is affected by the possibility for phonological transfer. In a study on artificial word acquisition, Gosselke Berthelsen et al. (2020) found an ELAN response for pre-attentive morphosyntax activation only in learners for whom L1-L2 similarity was high. Together, these studies thus indicate that, at least initially, only words with L1-like phonology can be processed outside of consciousness. This is likely a result of listeners being able to extend their neural subsystems, fine-tuned to the automatic processing of native sounds, to native-sounding foreign words, in line with the concept of L1-L2 transfer. An exception to the illustrated limitations in pre-attentive processing has been observed in children

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

(Partanen et al., 2017). Children produced changes in early components for non-native as well as native-sounding words. This is likely due to larger neural plasticity in the developing brain. However, the children still displayed clear hemispheric differences in the processing of phonologically native and non-native novel words, thus indicating that phonological similarity and transfer have an important impact on pre-attentive processing in all learners, even children. Presumably as a consequence of transfer and the possibility for pre-attentive processing, unfamiliar phonology also sometimes negatively affects behavioural learning outcomes (Havas et al., 2018; McKean et al., 2013) and impedes predictive processing (Lozano-Argüelles, Sagarra & Casillas, 2019; Sagarra & Casillas, 2018).

A grammar feature that lends itself well to the study of transfer in initial L2 acquisition is the so-called grammatical tone. Grammatical tone is a suprasegmental feature that is added onto the segments (vowels, consonants) of a word to make grammatical distinctions (e.g., case, gender, or number). Since the tone itself as well as its grammatical content are independent, additive features, the word's lexical and non-tonal grammatical meaning (e.g., suffixes) can be accessed irrespective of the tone. The tone's grammatical meaning, on the other hand, can only be accessed through the tone itself. If words with grammatical tone are taught to (beginner) learners from tonal and non-tonal language backgrounds, it is possible to directly compare general acquisition abilities (segmental features) to potentially transfer-affected acquisition (suprasegmental features) within the same words[1]. Importantly for the purpose of the current study, this can be extended to a direct comparison of initial learners' error processing (i.e., errors across segmental features) and error processing in light of potential transfer (errors in the suprasegmental tone dimension). An important premise, of course, is that L1-L2 transfer is possible for tone. This has, indeed, been observed behaviourally, at least when the L2 tone is less complex than the L1 tone and similar in form and function (e.g., Braun & Johnson, 2011; Hallé, Chang & Best, 2004; So & Best, 2010; Wayland & Li, 2008). Neurophysiological accounts of transfer effects in L2 learners of tone languages, on the other hand, are rare. However, studies using mismatch negativity (MMN) paradigms have observed transfer based modulations of L2 tone perception. To this effect, only learners with a tonal L1 showed modulations of MMN effects based on tone functionality and not only pitch height (Shen & Froud, 2019; Yu, et al., 2019). For post-perceptive processing, no N400 for lexical tone mismatches or LAN/P600 for mismatches in grammar-related tone is found in natural L2 learners with a non-tonal L1 (Gosselke Berthelsen et al., 2020; Pelzl, Lau, Guo & DeKeyser, 2019). In addition to these null results, the clearest indication of transfer was found for artificial, grammatical tone in Gosselke Berthelsen et al. (2020): a rapidly emerging left anterior negativity (LAN), a P600, as well as early, pre-attentive processing (i.e., word recognition effect and ELAN) emerged in learners with a tonal L1. Non-tonal learners, in contrast, showed no evidence of pre-attentive processing and a later LAN onset. Thus, we see distinct transfer effects for L2 tone acquisition and processing. We, therefore, suggest that grammatical tone is well suited for examining how learners, differentially affected by transfer, differ in their processing of L2 errors.

[1]It is possible, of course, that the mere presence of tone, despite its additive nature, poses an acquisition-hindrance for non-tonal learners. This would result in impoverished performance and processing in all conditions and would become apparent in the analysis of the segmental features.

Word-picture association learning and picture processing

While transfer is an important and often deliberately studied feature in initial L2 acquisition, studies attempt to limit the EXPLICIT presence of the learners' L1 in initial learning contexts to avoid a conscious mediation of the L2 through the L1. A useful method for circumventing overt L1 exposure during initial L2 learning is word-picture association learning where new L2 words are taught with the help of pictures. Interestingly, studies using this paradigm have uniformly found that picture associations can install rapid lexicosemantic (Dittinger et al., 2019; François, Cunillera, Garcia, Laine & Rodríguez-Fornells, 2017; Havas et al., 2017; Yang & Li, 2019) or grammatical (Gosselke Berthelsen et al., 2020; Havas et al., 2017) processing in novel words. For this to occur, learners must be aware of the lexical and grammatical content of the meaning-assigning pictures. Relatively little, however, is known about how this affects the processing of grammatically relevant features (e.g., number or gender) in meaning-carrying image referents themselves. We know from early N400 literature, that at least the N400 appears to be largely amodal and increases for inconsistent sentence endings regardless of whether they are expressed through words or pictures (e.g., Ganis et al., 1996; Kutas & van Petten, 1990; Nigam, Hoffman & Simons, 1992). In fact, even pictures primed with single auditory words show language-like N400 modulations (Pratarelli, 1994). Processing of grammatical properties in pictures, on the other hand, has been studied considerably less. One study that did investigate this found a LAN (but no P600) for gender mismatched pictures in sentential contexts (Wicha, Moreno & Kutas, 2003). This suggests that at least some lexicosemantic and morphosyntactic processes can be activated by pictures. Based on these findings, we assume that lexicosemantic and grammatical errors in the picture of a word-picture association paradigm should evoke the corresponding ERP responses (e.g., N400, LAN) and therefore constitute a useful tool for testing error processing in learners at the very beginning of the acquisition process.

The current study

Addressing a thus far understudied issue in initial second language acquisition, we investigated how error processing is realised during the initial acquisition of L2 words. As mentioned, learners at first process all L2 errors lexicosemantically. Grammatical components for error processing emerge later in both natural and artificial SLA. In artificial SLA, this process can, however, be sped up and grammatical processing can be found with minimal exposure. In the present study, we drew on a word-picture association paradigm that had previously produced rapid grammatical processing in novel words (Gosselke Berthelsen et al., 2020). The words in the present study contained grammatical tone and were tested on learners with and without a tonal L1. This allowed for a fine-grained study of phonological transfer effects. As a control condition, the words also contained a grammatical vowel change which both learner groups were equally familiar with from their L1. Importantly, we included mismatches in the word-picture pairs that became transparent at picture onset and analysed the learners' behavioural and neural responses to the resulting errors. Previous studies have illustrated that pictures in linguistic contexts

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

can evoke both lexicosemantic and grammatical responses. Moreover, we were aware that word-picture association paradigms can successfully be used to teach grammar; a fact which entails that the linguistic content of pictures in this paradigm must be evident to learners. Therefore, pictures were considered a suitable, straightforward way of studying the processing of morphosyntactic mismatches in novice learners. Based on previous results and due to the strong focus on grammar in this paradigm, we expected grammatical responses (e.g., LAN/P600) to errors in the initial learners. We further anticipated error processing to be facilitated by L1-L2 transfer. To test for this, we elicited errors in a phonological dimension that one learner group was unfamiliar with (i.e., tone) or in dimensions that both learner groups knew from their L1 (i.e., vowel and consonant). We hypothesised that transfer effects in the form of reduced or missing mismatch responses would only appear in the unfamiliar tone condition, since the tone was an additive feature and the remainder of the word could be acquired independently of the tone.

Methods

Participants

For the present study, we recruited two groups of right-handed participants (assessed by revised Edinburgh Handedness Inventory; Williams, 2013): 24 participants with a native language with grammatically relevant tone, Swedish[2], henceforth 'tonal L1s' (TL1s), as well as 24 participants with a closely-related non-tonal native language, German, henceforth 'non-tonal L1s' (NTL1s). The groups were matched for gender (TL1 = 10 males, NTL1 = 11 males), age (M_{TL1} = 23.7 years, SD = 2.6; M_{NTL1} = 23.7 years, SD = 1.6), socioeconomic status (Hollingshead, 1975), working memory span (tested via an automated operation span task, cf. Unsworth, Heitz, Schrock & Engle, 2005), as well as the number of languages spoken at above-elementary proficiency (M_{TL1} = 2, SD = 1; M_{NTL1} = 2, SD = 1). Furthermore, all participants had normal hearing defined as pure-tone hearing thresholds ≤20 dB Hearing Level (ISO, 2004) and could correctly discern tested auditory variations in linguistic input (i.e., vowel length, M_{TL1} = 97.7%, SD = 3.0; M_{NTL1} = 96.5%, SD = 4.0) as well as non-linguistic input (i.e., pitch, M_{TL1} = 91.2%, SD = 9.0; M_{NTL1} = 93.8%, SD = 6.2).

All experiments were conducted in agreement with the ethical guidelines for experiments in the Declaration of Helsinki and carried out in the Lund University Humanities Lab. All NTL1 participants were exchange students at Lund University. The NTL1s had no extensive knowledge of Swedish (highest self-assessed level of proficiency was B1 [=intermediate] of the Common European Framework of Reference, mean proficiency was A1 [=beginner]). Despite having lived in Sweden for some months (M = 23 weeks, SD = 16) and having studied Swedish to some extent (M = 10 weeks, SD = 7), they only engaged with Swedish actively for on average 5 hours per week (i.e., studying, conversation, listening, SD = 7) and passively for 10 hours (i.e., Swedish spoken in the background, SD = 8). Only one of the NTL1 participants reported having heard of Swedish word accents and none were aware of their ties to grammar nor did they include them in their vocabulary acquisition routines. Additionally, we tested the NTL1s on their perception of Swedish word accents and both behavioural and neuropsychological data suggested that they processed them as linguistically irrelevant.

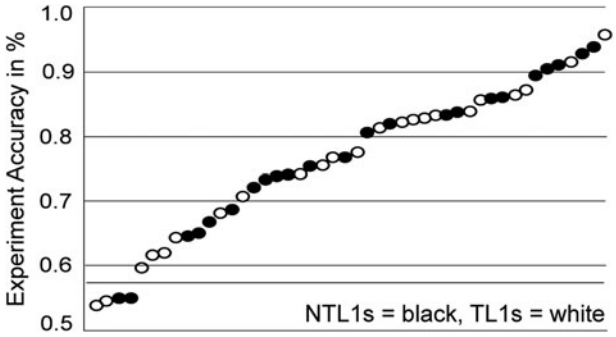


Fig. 1. Distribution of participants' accuracy data during the first session, from low to high. Participant group indicated by black (TL1) and white (NTL1) dots.

Keeping with participant sizes in previous studies with related paradigms (e.g., 23: de Diego-Balaguer, Rodríguez-Fornells & Bacoud-Lévi, 2015; 19: Havas, Laine & Rodríguez-Fornells, 2017; Leminen, Kimppa, Leminen, Lehtonen, Mäkelä & Shtyrov, 2016), we originally recorded 24 participants per group. As each participant had their own set of stimuli, multiples of eight allowed us to counterbalance the distribution of the four grammatical features over vowels and tones. Prior to analyses, we excluded one participant from each group, due to rudimentary knowledge of a tone language (Chinese) and experienced discomfort with the EEG equipment, respectively. Closely inspecting the behavioural data, we subsequently found that some participants performed poorly on the acquisition task of the main study and remained at chance level for error detection throughout the first session of the experiment (see section 2.3 below). Closer inspection of the distribution of the accuracy data (Figure 1) indicated a discontinuity in accuracy between 55 and 60% that we interpreted as a natural cut-off point. We therefore excluded the four participants (2 NTL1s, 2TL1s) whose accuracy during the first session stayed below 55%, resulting in final group sizes of 21 participants.

Stimuli

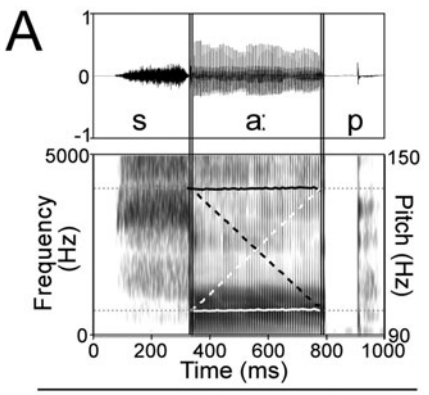
We used a learning paradigm that has previously been reported to be efficient in learning novel lexical and morphosyntactic features with tonal and non-tonal distinctions (Gosselke Berthelsen et al., 2018). Participants learned new word-picture associations by hearing an auditorily presented novel word followed by a meaning-assigning picture, as described in more detail below.

Auditory stimuli

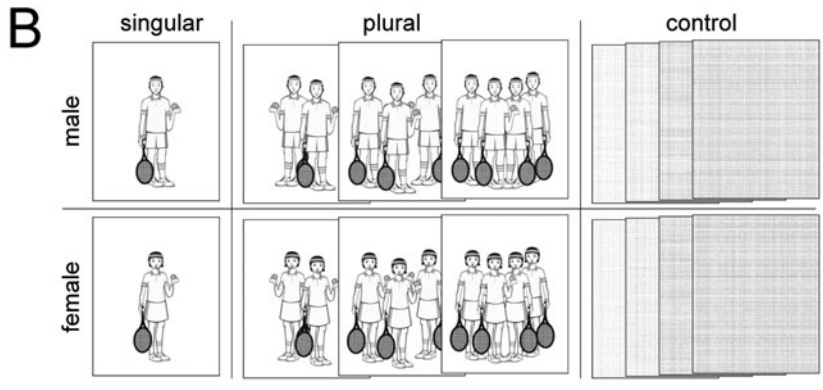
For the sound stimuli, we constructed pseudowords from consonants and vowels, which were produced by a native speaker of Russian (rather than German or Swedish to prevent differential carry-over effects) and chosen on the basis of being phonologically close in all three languages. Using Praat (Boersma, 2001), consonant and vowel durations were standardised. Subsequently, consonants and vowels were spliced into 24 simple CVC pseudowords and four different tones were added to the words with the help of pitch-manipulation. See Figure 2 for an overview of all words and an example stimulus.

[2]Swedish has tones (word accents) that are strongly associated with grammatical suffixes. A low tone (accent 1) on the stem of the word *fisk* ("fish"), for instance, can be followed by the definite singular suffix *-en,* (*fisk$_1$-en*, "fish-the") but not by the plural suffix *-ar* (**fisk$_1$-ar*, "fish-PL"). The plural instead requires a high tone (accent 2) on the word stem: *fisk$_2$-ar*.

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press



		Vowels a	ɛ	i	u
Initial consonants	d	dap	dɛp	dif	duf
	f	fap	fɛp	fif	fuf
	k	kaf	kɛf	kip	kup
	l	lap	lɛp	lir	lur
	s	sap	sɛp	sis	sus
	t	taf	tɛf	tip	tup



Professions			
doctor	teacher	chemist	photographer
boxer	painter	race driver	basketball player
fisher	waiter	journalist	fire fighter
tailor	gardener	tennis player	flight attendant
cook	flautist	violinist	hairdresser
singer	cleaner	ballet dancer	police officer

Fig. 2. A. Auditory word stimuli: Acoustic waveform, spectrogram and the four fundamental frequency (F0) curves for an example stimulus (above) and a list of all test words (below). B. Visual stimuli: An example set of picture stimuli for one profession (above) and a list of all professions in the experiment (below).

Picture stimuli

We chose images of humans portrayed as 24 different professions for the pictorial stimuli. For each profession, we constructed eight different images: one picture each with one, two, three, or four female or male workers, respectively. This way, in addition to the lexical information (profession), we were able to visually implement two grammatical categories: gender (masculine-feminine) and number (singular-plural). For the control condition, non-meaningful pictures were constructed by scrambling all pixels from the profession pictures. See Figure 2 for an overview of all professions and an example set of picture stimuli. See the supplementary material for an overview of all used picture stimuli.

Experimental procedure

All participants took part in two acquisition sessions, 24 hours apart. They sat on a chair one meter from a computer screen and kept their index fingers on a response box that stood on a table in front of them. The experiment was controlled by E-Prime 2 stimulation software (Psychology Software Tools Inc., Sharpsburg, PA). All auditory stimuli were routed through a GSI 16 Audiometer (Grason & Stadler Inc., Eden Prairie, MN) and presented at 70 dB SPL through a pair of circumaural earphones (California Headphone Company, Danville, CA). The presentation level was verified using a Brüel and Kjaer 2231 sound level meter with a 4134 microphone in a 4153 Artificial Ear. After a short instruction and training session (only in session 1), participants were asked to learn 24 words consisting of six professions that could be masculine or feminine (gender) and singular or plural (number). Professions were always expressed through the initial and final consonant (consonant frame). Gender and number were always expressed by vowel or tone, equally distributed across participants. Each participant had a unique set of words (i.e., list of six professions and combination of vowels, tones and their respective meaning). All vowels and tones were used equally often as part of target and control words. For an example set of stimuli, see Table 1.

Table 1. An example set of stimulus words.

	Target words									
	Masculine				Feminine				Control words (no meaning)	
	Singular		*Plural*		*Singular*		*Plural*			
violinist	d*i*f	**fall**	d*u*f	**fall**	d*i*f	**high**	d*u*f	**high**	dap / dɛp	rise/low
cleaner	f*i*f	**fall**	f*u*f	**fall**	f*i*f	**high**	f*u*f	**high**	fap / fɛp	rise/low
tennis player	k*i*p	**fall**	k*u*p	**fall**	k*i*p	**high**	k*u*p	**high**	kaf / kɛf	rise/low
painter	l*i*r	**fall**	l*u*r	**fall**	l*i*r	**high**	l*u*r	**high**	lap / lɛp	rise/low
gardener	s*i*s	**fall**	s*u*s	**fall**	s*i*s	**high**	s*u*s	**high**	sap / sɛp	rise/low
race driver	t*i*p	**fall**	t*u*p	**fall**	t*i*p	**high**	t*u*p	**high**	taf / tɛf	rise/low

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

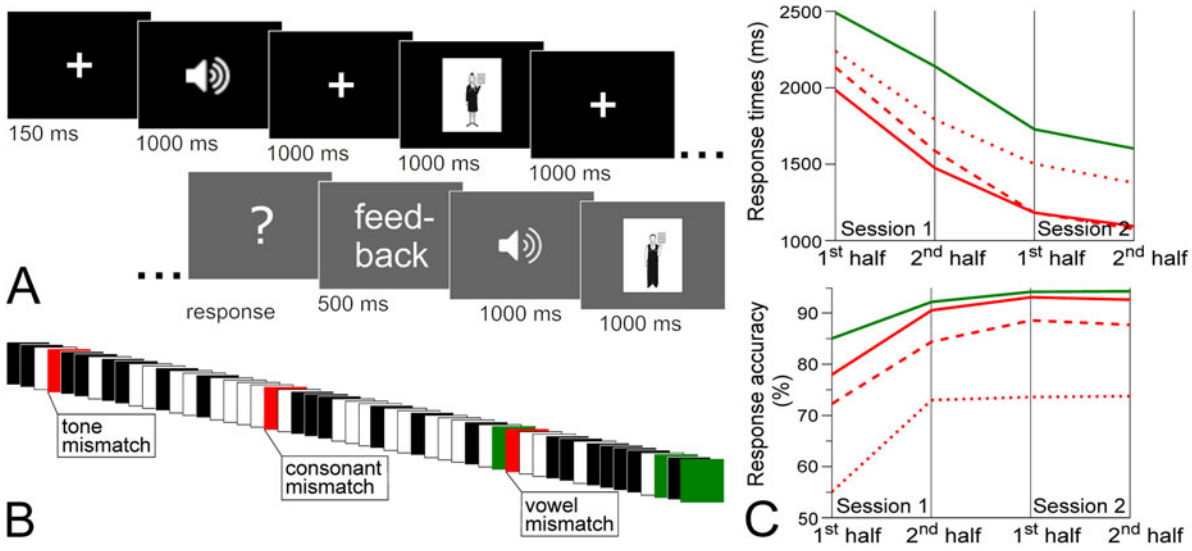


Fig. 3. A. Experiment procedure: Black slides show test and control trials, grey slides are added for question trials. B. Example cycle from the experiment highlighting the random distribution of test trials (black), control trials (white), matched question trials (green), and mismatch trials (red). C. Behavioural results for the question trials: Matched question trials in green, mismatched trials in red. Consonant mismatch in solid lines, vowel mismatch in dashed lines, and tone mismatch in dotted lines.

Using the procedure shown in Figure 3, all 48 word-picture combinations from each participant's target and control set were presented 30 times in each of the two sessions. In every full repetition of all 48 experiment items (cycle), three pseudorandomly selected meaningful word-picture trials were followed by overt questions pertaining to the correctness of the previous word-picture pair. Every cycle also included three additional question trials where there was a mismatch between word and picture related to one of the learned categories: number, gender (i.e., based on tone or vowel) or profession (based on consonants). Participants responded to all question trials by pressing a button on the response box and overt feedback on their response was given. After every 10 cycles (~40 minutes), participants were offered a break.

Electroencephalography (EEG)

Throughout the experiment, the participants' brain activity was recorded using 64 Ag-AgCl EEG electrodes mounted in an electrode cap (EASYCAP GmbH, Herrsching, Germany), a SynAmps^2 EEG amplifier (Compumedics Neuroscan, Victoria, Australia), and Curry Neuroimaging Suite 7 software (Compumedics Neuroscan). To monitor eye movements, horizontal and vertical bipolar electrooculogram channels (EOG) were added. Impedances for the scalp channels were kept below 3 kΩ and below 10 kΩ for the eye channels. The left mastoid (M1) was used as online reference and the frontocentral electrode AFz as ground. EEG was recorded with a 500 Hz sampling rate using DC mode and an online anti-aliasing low pass filter at 200 Hz.

The recorded EEG data was then re-referenced offline to average reference, and subsequently filtered with a 0.01 Hz high pass and a 30 Hz low pass filter. ERP epochs of 1200 ms including a 200-ms baseline were extracted for word and picture stimuli at word disambiguation point and picture onset, respectively. The extracted epochs included both learning and question trials. Independent component analysis (ICA) (Jung, Makeig, Humphries, Lee, McKeown, Iragui & Sejnowski, 2000) was conducted on all epochs. ICA components representing eye artefacts and single bad channels were removed. Finally, all epochs still exceeding ±100 μV were excluded. Only the epochs pertaining to picture stimuli were considered for the analysis.

Statistical analysis

Behavioural data analysis was conducted separately for the two experimental factors 'Response Time' (RT) and 'Response Accuracy' which were recorded for question trials. RT measures were log-transformed to normalise the distribution and submitted to a mixed Analysis of Variance (ANOVA) in IBM SPSS Statistics 25 (International Business Machines Corp., Armonk, NY, United States) with within-subject factor 'Type' (consonant-related word-picture mismatch [henceforth, consonant mismatch], vowel-related word-picture mismatch [henceforth, vowel mismatch], and tone-related word-picture mismatch [henceforth, tone mismatch], and correctly matched word-picture trials [henceforth, matched pictures]), temporal factors 'Session' (session 1 vs. session 2) and 'Half' (first vs. second half of a session), and the between-subject factor 'Learner Group' (TL1s vs. NTL1s). For the response accuracy data, d′ scores were calculated for each participant, condition and time window by comparing z transforms of hit rates (correct acceptance of matched trials relative to the total number of matched trials) and false alarm rates (incorrect acceptance of mismatched trials compared to the total number of mismatched trials - by mismatch type). Log-linear corrections (Hautus, 1995) were applied to extreme values (i.e., 0 and 1). D′ scores were submitted to a mixed ANOVA with the factors 'Mismatch Type' (consonant, vowel, tone), the temporal factors 'Session' and 'Half' and the between-subject factor 'Learner Group'. Greenhouse-Geisser correction was used when applicable. Main effects and interactions were considered significant at a p-value of < 0.05. For pairwise comparisons, False Discovery Rate (FDR) corrections (Benjamini & Hochberg, 1995) were used.

To find effects of interest in the ERPs, global root mean squares (gRMS) of the data were used to investigate when the

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

learner groups' neural activity was maximal (cf. Lehmann & Skrandies, 1980). When there were peak latency differences between the learner groups, we submitted the participants' unweighted mean gRMS peaks to a 2-tailed independent samples t-test. If there were significant between-group differences for gRMS peak latency, we established different time windows for the ERP analyses for the respective peak and analysed the groups separately. Subsequent to the gRMS analysis, we inspected the ERPs at gRMS peak latencies to find the ERP effects which the peaks were related to. The gRMS curve reaches a maximum when there is a CHANGE in neural activity and gRMS peaks therefore often mark effect onset rather than effect peak (cf. e.g., Roll, Söderström, Mannfolk, Shtyrov, Johansson, van Westen & Horne, 2015). Thus, using gRMS maxima as effect onsets, we selected time windows of 70 ms for the semi-early peaks (150–300 ms) and 200 ms for late peaks (above 300 ms) for the ERP analysis.

Testing for ERP effects at the gRMS peak latencies, mean amplitudes over the selected time windows for each electrode and condition were submitted to a cluster-based permutation test to find electrode groups that differed significantly between conditions. Each permutation test included one of the mismatched conditions and the match condition. We also tested for interactions with the between-subject factor 'Learner Group' as well as the temporal factors 'Session' and 'Half' and combinations thereof (e.g., interaction with 'Half' in session 1) in all conditions. All permutation analyses were carried out with help of the nonparametric cluster-based permutation approach implemented in Fieldtrip (Maris & Oostenveld, 2007). Using the Monte Carlo method to account for large data sets, we ran 1000 random permutations of the data. Clusters of three or more electrodes that had a p-value of < 0.05 were considered significant.

Finally, we conducted two-tailed Pearson correlations between response times and effect amplitudes (i.e., difference between mismatch and match amplitudes for electrodes in mismatch clusters) for all emerging error-related clusters in order to examine possible relationships between error processing in novice learners and subsequent behavioural responses.

Results

Behaviour

For the behavioural variable Response Accuracy, the mixed ANOVA produced a main effect of Mismatch Type, see Table 2. FDR corrected pairwise comparisons revealed that participants were significantly more accurate for vowel or consonant conditions than for tones. Further, mismatches in consonants were more accurately detected than those involving vowels. Learners improved throughout and between sessions, particularly from the beginning to the end of session 1. The non-tonal learners' tone detection ability was initially so poor that it differed clearly from vowel and consonant mismatches but increased over time and approached vowel mismatch levels towards the end of the learning sessions.

For Response Times, we found a main effect of Type (see Table 3). FDR corrected pairwise comparisons revealed that participants were faster at detecting consonant mismatches and vowel

Table 2. Results of the mixed Analysis of Variance (ANOVA) analysis for Response Accuracy (d′ values) as well as means and standard variations of the raw data. (Tone Mm = tone mismatch, Vowel Mm = vowel mismatch, Cons. Mm = consonant mismatch, S1 = session 1, S2 = session 2, H1 = first half, H2 = second half)

Main effects, Interactions, Pairwise Comparisons	F	p	Means in % (SD)			
Mismatch Type	19.12	<.001				
Pairwise comparisons: Tone Mm vs Vowel Mm		<.001	Tone Mismatch			69.07 (26)
vs Cons. Mm		<.001	Vowel Mismatch			83.50 (14)
Vowel Mm vs Cons. Mm		.004	Cons. Mismatch			88.77 (9)
Session	96.65	<.001	S1	79.91 (9)	S2	87.46 (12)
Half	122.58	<.001	H1	80.19 (10)	H2	86.30 (11)
Session * Half	24.33	<.001				
Session 1: Half	145.44	<.001	H1	72.75 (11)	H2	85.28 (10)
Session 2: Half	43.61	<.001	H1	87.61 (12)	H2	87.31 (13)
Mismatch Type * Half * Learner Group	3.12	.050				
TL1s: Type * Half	0.34	.694				
NTL1s: Type * Half	3.51	.044				
Half 1: Mismatch Type	16.62	<.001				
Pairwise comp.: Tone Mm vs Vowel Mm		<.001	Tone Mismatch			58.10 (25)
vs Cons. Mm		<.001	Vowel Mismatch			80.32 (19)
			Cons. Mismatch			84.44 (13)
Half 2: Mismatch Type	5.31	.017				
Pairwise comp.: Tone Mm vs Cons. Mm		.045	Tone Mismatch			70.48 (31)
			Vowel Mismatch			82.54 (18)
			Cons. Mismatch			89.68 (10)

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

Table 3. Results of the mixed Analysis of Variance (ANOVA) analysis for Response Times (log-transformations) as well as means and standard variations of the raw data. (Tone Mm = tone mismatch, Vowel Mm = vowel mismatch, Cons. Mm = consonant mismatch, S1 = session 1, S2 = session 2, H1 = first half, H2 = second half)

Main effects, Interactions, Pairwise Comparisons		F	p	Means in ms (SD)			
Type		39.71	<.001				
Pairwise comparisons:	Tone Mm vs Vowel Mm		<.001	Tone Mismatch			1731 (794)
	vs Cons. Mm		<.001	Vowel Mismatch			1497 (705)
	vs Match		<.001	Cons. Mismatch			1437 (661)
	Vowel Mm vs Match		<.001	Match			1991 (701)
	Cons. Mm vs Match		<.001				
Session		49.50	<.001	S1	1982 (773)	S2	1346 (713)
Half		39.07	<.001	H1	1807 (670)	H2	1521 (720)
Session * Half		11.83	<.001				
Session 1: Half		34.69	<.001	H1	2214 (807)	H2	1751 (865)
Session 2: Half		3.73	.013	H1	1400 (746)	H2	1291 (710)
Session * Type		9.07	<.001				
Session 1: Type		26.20	<.001				
Pairwise comparisons:	Tone Mm vs Cons. Mm		.007	Tone Mismatch			2019 (894)
	vs Match		<.001	Vowel Mismatch			861 (835)
	Vowel Mm vs Match		<.001	Cons. Mismatch			1733 (823)
	Cons. Mm vs Match		<.001	Match			2316 (781)
Session 2: Type		37.96	<.001				
Pairwise comparisons:	Tone Mm vs Vowel Mm		<.001	Tone Mismatch			1444 (846)
	vs Cons. Mm		<.001	Vowel Mismatch			1132 (750)
	vs Match		.001	Cons. Mismatch			1140 (676)
	Vowel Mm vs Match		<.001	Match			1666 (744)
	Cons. Mm vs Match		<.001				

mismatches than tone mismatches or matched pictures (for means and p-values, refer to Table 3). Furthermore, Response Times improved throughout the sessions and between sessions. The strongest improvement took place within session 1. Vowel mismatch detection was not significantly faster than tone mismatch detection until session 2. There were no significant interactions with Learner Group.

Electrophysiology

Results for gRMS peak latency

Two gRMS peaks emerged in the data for both participant groups: ~180 ms (visual N1 latency) and ~370 ms (N400/LAN latency) after the picture onset. In the group average, the first peak was delayed for the TL1 group compared to the NTL1 group, cf. Figure 4. A 2-tailed independent samples t-test showed significant differences in average gRMS peak latency between the learner groups, $t(40) = -2.22$, $p = .032$. The TL1 group's first gRMS peak ($M = 185$ ms, $SD = 12$) was significantly delayed compared to the NTL1s' ($M = 176$ ms, $SD = 13$). Accordingly, the time windows for the first gRMS peak were 185-255 ms for the TL1 group and 176-246 ms for the NTL1 group (i.e., peak +70 ms). For the second gRMS peak ($M_{TL1} = 373$ ms, $SD = 49$, $M_{NTL1} = 379$ ms, $SD = 44$), there were no significant differences in timing. Hence, the time window submitted for the second gRMS peak was defined as 370-570 ms (i.e., ~peak +200 ms).

Results for the ERP data

At N1 latency, the permutation analysis did not find any electrode clusters with different amplitudes for any mismatch condition or interaction. At N400/LAN latency, the permutation analysis produced a negative, centroposterior cluster (CP3, CP1, CP2, CP4, P1, P2), $p = .005$, indicative of an N400 effect, as well two positive, anterior clusters: Fpz, F7, F5, F3, F4, $p = .021$, and F1, Fz, FC1, $p = .034$ (cf. Figure 5B) for the difference between consonant mismatch and matched pictures. The permutation test also identified a difference between vowel mismatch and matched pictures which manifested as a negative cluster at centroposterior electrodes (C6, CP3, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P8, P07, P08, Oz), $p = .001$, again indicating an N400 response, as well as three positive clusters at anterior electrodes: AF4, F1, Fz, F2, FC3, FC1, FC2, C1, $p = .002$, F3, F4, Fpz, AF7, AF8, F5, p = .004, and Fp1, Fp2, F6, $p = .035$ (cf. Figure 5B). Finally, in the comparison of tone mismatched and matched picture conditions, the permutation analysis did not find any differences for both groups collectively. The interaction with Learner Group for tone mismatch compared to matched pictures produced an N400-related centroposterior cluster (P1, Pz, P2), $p = .043$, (cf. Figure 5B) where amplitudes

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

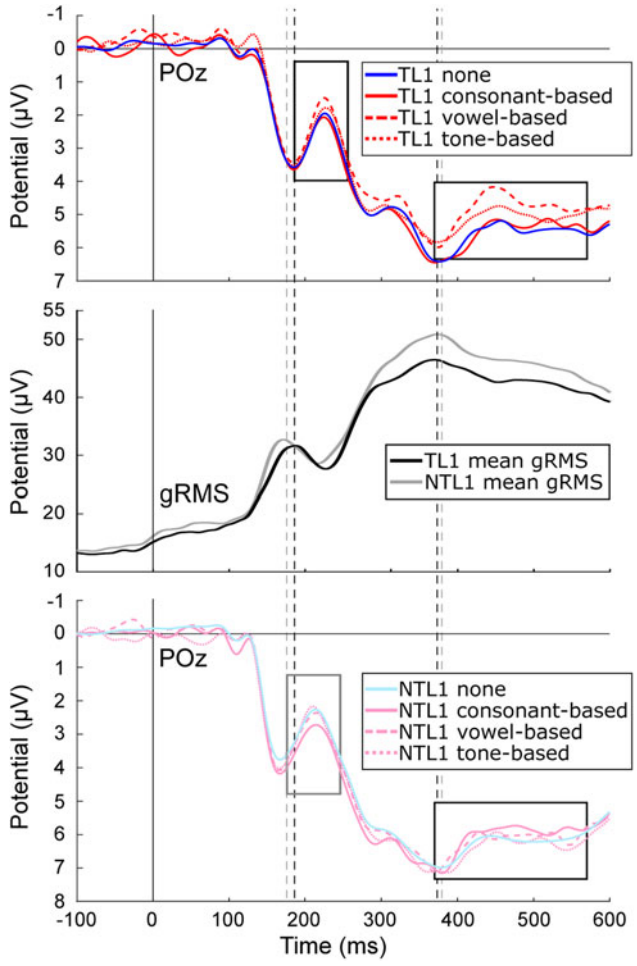


Fig. 4. Middle: Latency of gRMS peaks for both learner groups, both sessions and all conditions averaged. Top: ERPs averaged over both sessions separately for all four experimental conditions in the tonal learner group at posterior electrode POz. Bottom: ERPs averaged over both sessions separately for all four experimental conditions in the non-tonal learner group at POz. Vertical dashed lines show the group's average gRMS peaks calculated from the subjects' individual average peak in each condition. Black = TL1, grey = NTL1.

were significantly more negative in the TL1 group than the NTL1 group. There were no interactions with Learner Group in the other conditions. Interactions with the temporal factors Session and Half or combinations thereof did not produce any significant clusters for either of the mismatch types.

Correlation behavioural and neurophysiological results

When testing for correlations between response times and ERP effects for error processing (i.e., N400), we found no significant correlations in the consonant condition. In the vowel condition, there was a significant positive correlation between vowel N400 and response times to vowel mismatches, $r(40) = .370$, $p = .016$: the larger the N400 effect for vowel, the faster the response times to vowel mismatches (cf. Figure 5C). For the tone N400 in the TL1 group, we similarly observed a positive correlation between tone N400 and response times to tone mismatches, $r(19) = .452$, $p = .040$. (For the non-tonal learners, in comparison, neural responses to tone mismatches and response times were not correlated: $r(19) = .033$, $p = .889$).

Discussion

The present study investigated the processing of lexicosemantic and morphosyntactic errors in initial second language acquisition and the role of transfer in this process. In a word-picture association paradigm, learners acquired novel words with grammatical tone and, as a control, grammatical vowel change. All participants were familiar with systematic, grammatically distinctive vowel alterations from their L1. With respect to grammatical tone, however, the learners differed, such that they either were (TL1s) or were not (NTL1s) familiar with grammatically meaningful tone from their L1. This created the possibility of differential L1-L2 phonological transfer in the novel words. We occasionally presented pictures that were mismatched with the different phonological features of the preceding auditory words in the word-picture association paradigm. By doing so, we introduced lexicosemantic and morphosyntactic errors during the first four hours of acquisition. We expected to find neurophysiological responses related to lexicosemantic (N400) or morphosyntactic (E/LAN, P600) processing, respectively, for all phonological mismatch conditions in the TL1 learners. We anticipated grammatical ERP components in this learner group even though grammatical processing usually emerges relatively late in L2 learners. Acquisition conditions in the present study were optimised by allowing for complete L1-L2 transfer and through a strong focus on grammar. Furthermore, as mentioned in the introduction, we knew from previous studies that learners can deduce the grammatical content of the pictures in the word-picture association paradigm. We therefore assumed that grammatical ERP components, most likely a P600 and possibly a LAN or ELAN, would emerge for grammatically mismatched pictures in the tonal learners. For the NTL1 group, on the other hand, we expected reduced or missing ERP effects and possibly poorer behavioural performance for errors which were based on mismatches across the tone dimension but similar processing to the TL1 group for the other mismatch conditions.

Differential processing demands in beginner learners: Posterior N1

Somewhat unexpectedly, an early gRMS peak pointed at an important role of the visual perception-related posterior N1. The N1 (or N170), with a posterior distribution for visual input, is an exogenous response which is sensitive to task demands (Callaway & Halliday, 1982; Nash & Williams, 1982). Like other ERP components, the N1 is usually analysed with respect to amplitude differences between conditions or groups. The posterior N1 in the present study, however, showed timing rather than amplitude differences: the TL1 group's N1 was delayed in relation to the NTL1 group's N1. Such differences in N1 timing, although rare, have been attested in the literature and are generally associated with differentially effortful processing. As an example from the auditory domain, for instance, word recognition tasks in children with specific language impairment have elicited a delayed N1 compared to the N1 response from a matched control group of normally developing children (Malins, Desroches, Robertson, Newman, Archibald & Joanisse, 2013). With visual input, similar results have been reported when stimuli were processed under differential task demands. Children diagnosed with ADHD, for example, had a delayed N1 response to light flashes during an active detection task compared to a passive viewing (Callaway & Halliday, 1982). Similarly,

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

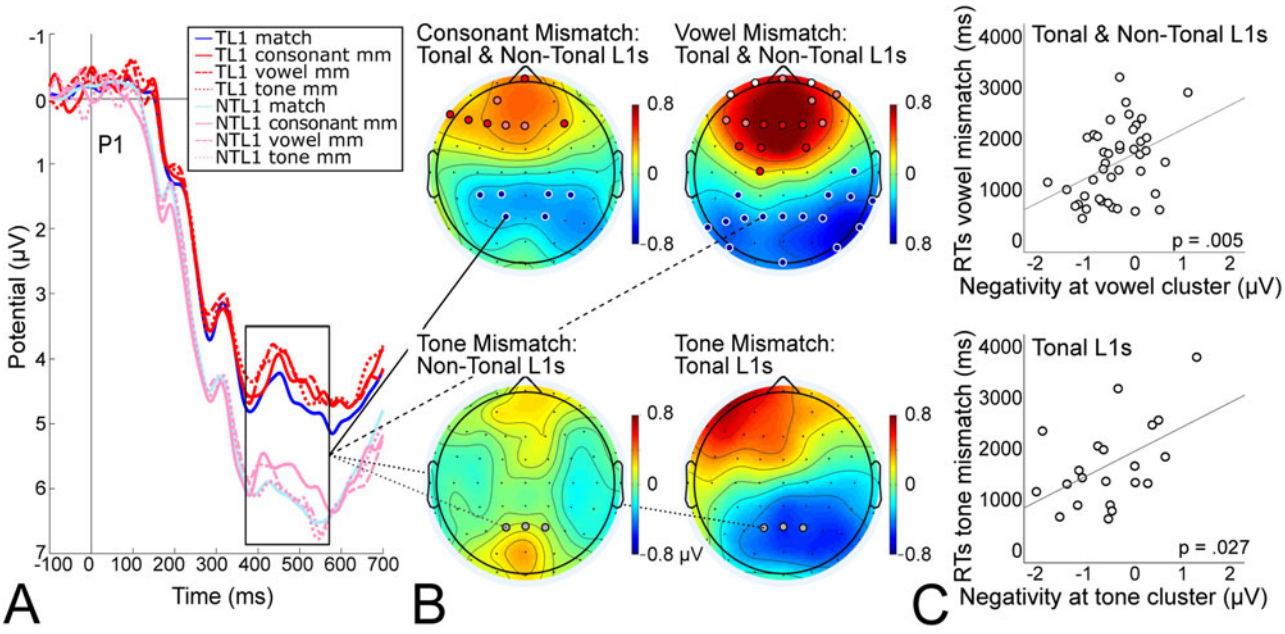


Fig. 5. A. ERPs for the two participant groups and all conditions at posterior electrode P1. B. Subtraction topographies for the different mismatch-match conditions. Significant negative clusters indicated with blue dots, positive clusters range from dark red (most significant) to white (least significant), and interaction clusters with grey dots. C. Scatterplots for the correlation of negativity at N400 clusters and the response times for vowel mismatch in both groups, as well as negativity at N400 clusters and RTs for tone mismatch in the TL1 group.

healthy adults have shown a delayed N1 when tasked with stimulus identification rather than simple stimulus detection for moving visual stimuli (Fort, Besle, Giard & Pernier, 2005). In line with those studies, we believe that the N1 latency in the present study is related to differential cognitive demands for picture perception and identification in the two learner groups and that this is based on top-down influences. From the word-picture association, the tonal learners rapidly internalise the words' three phonological features (tone, vowel, and consonant) and their respective meaning. The activation of these features during word processing likely pre-activates the associated pictorial properties (number, gender, and profession) in the TL1 learners prior to the emergence of the picture. Thus, when the picture referent appears in the input, the TL1 learners are faced with the difficult task of pre-consciously checking and identifying the image against the three pre-activated categories. The NTL1 learners, on the other hand, have presumably only internalised two phonological categories, i.e., vowel and consonant, and their corresponding meaning, which influence their pre-conscious perception and the picture identification task. We believe that the automatic access to a greater number of pictorial properties, through pre-activation, increased automatic picture identification difficulty in the non-tonal learners which manifested in a delayed posterior N1. Interestingly, there were no differences in either N1 latency or amplitude between mismatching or matching pictures in either participant group, suggesting that validity does not affect processing at this latency. Importantly, the N1 results show that the absence or presence of positive L1-L2 transfer in auditory words has a strong influence on the pre-attentive processing of upcoming pictorial referents. Hence, transfer not only affects the processing of the transfer-facilitated words themselves but also allows for pre-activation of closely related items in the immediate context via top-down processes, which influences word perception. This finding calls for more detailed investigations into the generalisability and scope of transfer-facilitated processing effects. Can a comparable N1 latency effect also be found for written word form identification and possibly even in the auditory modality? If so, how close to each other do the associated elements need to be? Can positive transfer effects on the first element in a non-adjacent dependency, for instance, influence the pre-attentive perception and identification of the second element further down the language stream? Addressing these and related questions could provide important insights for a more comprehensive understanding of the N1 in the perceptive processing of pre-activated elements in language contexts.

Mismatch responses in beginner learners: N400 as a response to all error types

For mismatch processing, we anticipated differential results depending on mismatch type. For semantic errors, which were contained in mismatches between the words' consonants and the profession depicted in the associated picture, we expected an N400 response. For morphosyntactic errors, which were elicited by mismatches of picture properties related to the words' vowels or tones, on the other hand, we expected to see morphosyntactic ERP components. The vowels and tones had clear and easily abstractable morphosyntactic content and L1-L2 transfer was possible. For comparable circumstances, previous studies on artificial acquisition rapidly elicited at the very least a P600 (e.g., Batterink & Neville, 2013; Havas et al., 2017) and sometimes even additional early frontal negativities (de Diego-Balaguer et al., 2007). However, we find no evidence of a P600-related late positivity in response to the morphosyntactic mismatches in the present study. Although surprising, the lack of a P600 might be explained by the fact that mismatches were between novel spoken words and pictorial referents. While N400 and LAN effects have previously been observed for pictures in linguistic contexts (e.g., Federmeier & Kutas, 2001; Wicha et al., 2003), a P600 was never reported in those studies. It is thus possible that the absence

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

of the P600 in the present study is simply due to the mismatch format. Instead of a P600, we found an earlier negativity, as indicated by a prominent gRMS peak at ~370 ms. This negativity was related to an N400 response and was found consistently for all mismatch types, but only when L1-L2 transfer was possible. The emergence of an N400 rather than the expected LAN may be an unintended result of the use of visual referents. The pictorial form of the mismatches may automatically have induced semantic processing. We assume that learners are aware of 'gender' and 'number' categories in a word-picture association paradigm, as the related visual cues in the pictures can install grammatical processing in the words (Gosselke Berthelsen et al., 2020). However, this does not necessarily entail that the same pictorial features could not in theory be processed semantically when they are presented as part of the pictures, especially in L2 contexts. That is to say, the abstracted, morphosyntactic label 'feminine' for the words could be processed concretely as the semantic feature 'female' during picture processing using visual features. For number this would correspond to a grammatical vs numeric number distinction between words and pictures such as 'plural' vs 'three'. The pictorial form of the mismatches is thus an entirely possible explanation of the absence of P600 and LAN effects.

An alternative explanation for the lack of morphosyntactic responses in the present data might be categorical differences between canonical processing and error processing in L2 learners. In a direct comparison of results elicited with the present word-picture association learning paradigm, we find an ELAN, LAN and P600 for canonical words but a N400 for mismatched pictures. This shows that learners can use the present paradigm to produce a full range of grammatical components for newly acquired words. These are elicited after just 20 minutes of acquisition in the comparison of learned words and meaningless pseudowords and thus for canonical processing (Gosselke Berthelsen et al., 2020). For errors in the word-picture pairs, in contrast, we do not observe morphosyntactic processing although this should in theory have been possible. Previous literature does find quickly emerging morphosyntactic components even in error contexts (Batterink & Neville, 2013; de Diego-Balaguer et al., 2007; Havas et al., 2017). Those studies, however, either had a much stronger, solitary focus on syntax (Batterink & Neville, 2013; de Diego-Balaguer et al., 2007), or a considerably less complex morphosyntactic system (Havas et al., 2017). The learners in the present study, instead, produced an initial L2 default response to all types of L2 errors: an N400 (Osterhout, Poliakov, Inoue, McLaughlin, Valentine, Pitkanen, Frenck-Mestre & Hirschensohn, 2008). This default response suggests that differentiations of mismatches as lexicosemantic or morphosyntactic do not take place. This differentiation might generally be of little importance to beginner learners. Error detection arguably takes precedence over error classification. In fact, this might be an important factor to keep in mind for our interpretation of grammar processing in second language learners. Traditionally, grammar processing is measured with the help of violation paradigms. The present results hint that this very tradition might potentially skew the picture. Learners might not in fact be incapable of grammar processing per se but instead simply be indifferent to different error types. They might then respond only to the unexpectedness of an erroneous item and the effect that this has on utterance-level meaning, something which would elicit an N400. Therefore, the late emergence of full-fledged grammatical processing of errors in both natural (e.g., Gillon Dowens et al., 2010; Tanner et al., 2012) and artificial second language acquisition (cf. e.g., Tanner et al., 2012; Morgan-Short et al., 2012) might be related to the fact that the learners' L2 processing resources are initially exhausted by canonical processing and the detection but not classification of errors. In summary, the present findings suggest a need for carefully evaluating the use of violation paradigms as measurements of L2 proficiency in beginner learners and, more generally, grammar processing in second language contexts. Future studies should more closely investigate potential categorical differences between canonical processing and error processing in L2 learners. Also, to untangle the contribution of error processing and picture processing from the current findings, it is essential that future studies test the processing of grammar errors contained in auditory words rather than pictures. This could even be achieved using the same word-picture association paradigm as in the present study but including intermitted reverse test pairs (i.e., picture-word) after every learning cycle where matches and mismatches are then contained in the auditory words.

The second main finding in the mismatch processing responses is the different processing of tone-mismatched picture referents between learner groups. This emphasises once more the importance of L1-L2 similarity in initial L2 acquisition: only mismatches based on familiar phonological features elicited a mismatch response in the learners. This is presumably related to the inability to pre-consciously process unfamiliar phonological elements; a realisation that is well-attested in the rapid learning literature (e.g., Gosselke Berthelsen et al., 2020; Kimppa et al., 2015; Silva et al., 2019). The non-tonal learners cannot pre-attentively access the word's tone or its content, which in turn likely inhibits automatic pre-activation of the tone-related picture features. As there are no pre-activated features for the tone condition for the NTL1s, the appearance of a mismatch in the pictures will not lead to an expectancy-based mismatch, hence preventing the emergence of an N400. Notice, however, that the lack of an expectancy-related mismatch response in the NTL1 group does not entail that mismatches were not detected. Behavioural results show no significant differences between the tonal and non-tonal learners in tone mismatch detection accuracy ($M_{TL1} = 68.9\%$; $M_{NTL1} = 62.6\%$) or response times ($M_{TL1} = 1724$ ms; $M_{NTL1} = 1853$ ms), replicating findings which show that transfer affects online but not offline processing in L2 learners (e.g., Andersson et al., 2019). For features where transfer is not possible, learners presumably rely on later, more consciously evaluative processing to detect mismatches. This contrasts with pre-activation-induced mismatch detection for transferable features, which is likely initiated during pre-conscious processing of the feature and its associated meaning. Interestingly, the additive nature of the grammatical tone entailed that the non-tonal learners could process the remainder of the word, i.e., the consonant and vowel content, independently of the tone. We could, therefore, observe an N400 for errors based on mismatches across the vowel and consonant dimension even in the NTL1 group. The transfer-related effect of the words' phonological features on the processing of mismatches in pictures again shows that transfer affects not only the word itself but also the processing of tightly associated referents, demonstrating the overarching influence of L1-L2 similarity in SLA. As argued for the N1, above, further studies are needed to determine the potential reach of such transfer-based top-down effects on the processing of upcoming words. It would be useful to determine, for instance, whether only the immediately following element can be top-down mediated or whether this effect extends to any

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

closely associated element, possibly with a number of intervening constituents.

Finally, in accordance with previous findings on the rapidness of initial L2 acquisition (e.g., Cunillera et al., 2009; de Diego-Balaguer et al., 2007; Dittinger et al., 2019; Gullberg et al., 2010; Gosselke Berthelsen et al., 2020; Kimppa et al., 2015), the N400 for mismatched pictures manifested already within 1 hour, i.e., within 15 learning cycles and 15 mismatch trials per condition. At this point, mismatch processing had manifested to such a degree that further changes over time were undetectable. Indeed, the effect was likely present already after 30 minutes. We see great improvements in the behavioural responses within the first session and know from previous literature that neurophysiological changes tend to precede behavioural ones (e.g., McLaughlin et al., 2010). A targeted investigation into the onset of error processing in L2 acquisition would require a reliable way of studying ERP responses at smaller time intervals: for instance, by means of testing a very large number of participants. Alternatively, future studies could use an acquisition paradigm with a considerably larger number of target words so that the number of error trials increases as a function of the total number of trials.

Behavioural responses to errors in initial learners

The two learner groups in the present study were virtually indistinguishable in their behavioural performance on mismatch detection. Although mean accuracy for tone mismatches was slightly lower in the non-tonal learner group and mean response times longer, between-group differences were not significant. The observed findings tie in nicely with previous literature showing that learners' offline responses are often unaffected by transfer (e.g., Andersson et al., 2019). Instead, facilitation effects based on L1-L2 similarity tend to manifest in online measures. Learners who cannot profit from positive L1-L2 transfer have to resort to different types of processing – for instance, conscious rather than pre-conscious (cf. Gosselke Berthelsen et al., 2020) – but reach the same learning outcome.

A secondary finding in the behavioural data in the present study was an overall poor acquisition outcome for tone. Of all mismatch conditions, tone had the lowest accuracy and slowest response times. The inferior learning performance is likely based on two factors: a slight functional dissimilarity in the L1 and L2 tone in this study and a general difficulty in the acquisition of L2 tone. The tonal participants' L1 tone is associated with grammar but not itself grammatically meaningful, unlike the L2 tone in the present study. This might have negatively impacted on the TL1s' tone acquisition. Besides, L2 tone is generally difficult to acquire and even distinguish. To this effect, Yang and Chan (2010) found that even highly advanced learners of Chinese performed poorly in the perceptual discrimination of tone. For the least accurate tone contrast in their study, discrimination accuracy was below 30% in the advanced learners, compared to nearly 100% in the native speaker control group. This considerable discrimination difference is a clear indicator of the general perceptual and in turn acquisitional challenges associated with L2 tone. Coupled with the slight L1-L2 tone dissimilarity, this readily explains the lower acquisition outcome for L2 tone even in the tonal learners.

Interestingly, however, while there were no between-group differences for mismatch detection in the behavioural variables, we found a correlation between behavioural factors and the N400 such that response times were faster the larger the N400. This correlation centrally emerged both for vowel mismatch and in the tonal learners also for tone mismatch. The emergence of the correlation in the tonal learners (and a clear lack of correlation in the NTL1s) strengthens the suggested impact of the words' phonological properties, via pre-activation of linguistic content, in the processing and detection of errors in picture referents. Those learners, who react most strongly to picture mismatches neurophysiologically (visible in larger N400 effects), are also fastest at identifying a corresponding error behaviourally. Both outcomes are likely mediated through expectation-related processes.

Conclusions

In the current study, we investigated how behavioural error detection and neurophysiological error processing proceeded in novice learners. We found that mismatches elicited by pictures in word-picture pairs uniformly elicited an N400 response, regardless of whether the mismatch was lexicosemantic or morphosyntactic in nature. We believe this to be indicative of the increased difficulty of error processing in second language learners, making the study of grammatical ERP responses in violation paradigms a questionable measure of L2 proficiency and processing. The N400 emerged well within an hour of acquisition and was correlated with behavioural responses. Importantly, we only observed an N400 when the error was based on a familiar phonological property in the preceding word, highlighting the importance of cross-linguistic transfer. Besides the N400, we found a visual perception-related posterior N1 for the pictures which was delayed for learners who could, via phonological transfer, internalise all three sound-meaning relationships and thus automatically pre-activate three rather than only two of the picture's associated visual properties.

Acknowledgements. The authors gratefully acknowledge Lund University Humanities Lab. Special thanks to Frida Blomberg for help and support particularly during data acquisition, to Jonas Brännström for help with the setup of the auditory stimulus presentation, as well as to all participants for patiently and eagerly taking part in the experiments. This study was supported by Knut and Alice Wallenberg Foundation (Grant No. 2018.0454), Swedish Research Council (Grant No. 2018.00632), Craaford Foundation (Grant No. 2017.0006), Lundbeck Foundation (Grant No. R140-2013-12951; R164-2013-15801), Danish Council for Independent Research (Grant No. 6110-00486, project 23776), and Russian Foundation for Basic Research (Grant No. 19-313-51016).

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728921000134

References

Andersson A, Sayehli S and Gullberg M (2019) Language background affects online word order processing in a second language but not offline. *Bilingualism: Language and Cognition* **22**(4), 802–825. doi:10.1017/S1366728918000573

Barber H and Carreiras M (2005) Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience* **17**(1), 137–153. doi:10.1162/0898929052880101

Batterink L and Neville H (2013) Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience* **25**(6), 936–951. doi:10.1162/jocn_a_00354.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* (*Methodological*) **57**(1), 289–300. doi:10.2307/2346101

Boersma P (2001) Praat, a system for doing phonetics by computer. *Glot International* **5**(9/10), 341–345.

Bowden HW, Steinhauer K, Sanz C and Ullman MT (2013) Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia* 2492–2511. doi:10.1016/j.neuropsychologia.2013.09.004

Braun B and Johnson EK (2011) Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics* **39**, 585–594. doi:10.1016/j.wocn.2011.06.002

Callaway E and Halliday R (1982) The effect of attentional effort on visual evoked potential N1 latency. *Psychiatry Research* **7**(3), 299–308. doi:10.1016/0165-1781(82)90066-X

Cunillera T, Càmara E, Toro JM, Marco-Pallares J, Sebasián-Galles N, Ortiz H, . . . Rodríguez-Fornells A (2009). Time course and functional neuroanatomy of speech segmentation in adults. 48(3), 541–553. doi:/10.1016/j.neuroimage.2009.06.069

Davidson DJ and Indefrey P (2009) Plasticity of grammatical recursion in German learners of Dutch. *Language and Cognitive Processes* **24**(9), 1335–1369. doi:10.1080/01690960902981883

de Diego-Balaguer R, Rodríguez-Fornells A and Bacoud-Lévi A-C (2015) Prosodic cues enhance rule learning by changing speech segmentation mechanisms. *Frontiers in Psychology* **6**(1478). doi:10.3389/fpsyg.2015.01478

de Diego-Balaguer R, Toro JM, Rodríguez-Fornells A and Bachoud-Lévi A-C (2007) Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLos ONE* **2**, 11. e1175. doi:10.1371/journal.pone.0001175

Dittinger E, Scherer J, Jäncke L, Besson M and Elmer S (2019) Testing the influence of musical expertise on novel word learning across the lifespan using a cross-sectional approach in children, younger adults and older adults. *Brain and Language* **198**(104678). doi:10.1016/j.bandl.2019.104678

Federmeier KD and Kutas M (2001) Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**(1), 202–224. doi:10.103I77////0022778-7393.27.1.202

Fort A, Besle J, Giard M-H and Pernier J (2005) Task-dependent activation latency in human visual extrastriate cortex. *Neuroscience Letters* **379**, 144-–48. doi:10.1016/j.neulet.2004.12.076

Foucart A and Frenck-Mestre C (2011) Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition* **14**(3), 379-399. doi:10.1017/S136672891000012X

Foucart A and Frenck-Mestre C (2012) Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language* **66**, 226–248. doi:10.1016/j.jml.2011.07.007

François C, Cunillera T, Garcia E, Laine M and Rodríguez-Fornells A (2017) Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia* 56–67. doi:10.1016/j.neuropsychologia.2016.10.006

Friederici AD (2002) Towards a neural basis of auditory sentence processing. *TRENDS in Cognitive Science* **6**(2), 78–84. doi:10.1016/S1364-6613(00)01839-8

Friederici AD, Steinhauer K and Pfeifer E (2002) Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *PNAS* **99**(1). doi:10.1073/pnas.012611199

Ganis G, Kutas M and Sereno MI (1996) The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience* **8**(2), 89–106. doi: 10.1162/jocn.1996.8.2.89

Gillon Dowens M, Guo T, Guo J, Barber H and Carreiras M (2011) Gender and number processing in Chinese learners of Spanish - Evidence from Event Related Potentials. *Neuropsychologia* **49**, 1651–1659. doi:10.1016/j.neuropsychologia.2011.02.034

Gillon Dowens M, Vergara M, Barber HA and Carreiras M (2010) Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience* **22**(8), 1870–1887. doi:10.1162/jocn.2009.21304

Gosselke Berthelsen S, Horne M, Brännström KJ, Shtyrov Y and Roll M (2018) Neural processing of morphosyntactic tonal cues in second-language learners. *Journal of Neurolinguistics* **45**, 60–78. doi:10.1016/j.jneuroling.2017.09.001

Gosselke Berthelsen S, Horne M, Shtyrov Y and Roll M (2020) Different neural mechanisms for rapid acquisition of words with grammatical tone in learners from tonal and non-tonal backgrounds: ERP evidence. *Brain Research* **1729**(146614). doi:10.1016/j.brainres.2019.146614

Grey S, Sanz C, Morgan-Short K and Ullman MT (2018) Bilingual and monolingual adults learning an additional language: ERPs reveal differences in syntactic processing. *Bilingualism: Language and Cognition* **21**(5), 970–994. doi:10.1017/S1366728917000426

Guarjardo LF and Wicha NY (2014) Morphosyntax can modulate the N400 component: Event related poentials to gender-marked post-nominal adjectives. *NeuroImage* **91**, 262–272. doi:10.1016/j.neuroimage.2013.09.077

Gullberg M, Roberts L, Dimroth C, Veroude K and Indefrey P (2010) Adult language learning after minimal exposure to an unknown natural language. *Language Learning* **60**(S2), 5–24. doi:10.1111/j.1467–9922.2010.00598.x

Hallé PA, Chang Y-C and Best CT (2004) Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* **32**, 395–42. doi:10.1016/S0095-4470(03)00016-0

Hanna J, Shtyrov Y, Williams J and Pulvermüller F (2016) Early neurophysiological indices of second language learning. *Neuropsychologia* **82**, 18–30. doi:10.1016/j.neuropsychologia.2016.01.001

Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers* **27**, 46–51. doi:10.3758/BF03203619

Havas V, Laine M and Rodríguez-Fornells A (2017) Brain signatures of early lexical and morphological learning of a new language. *Neuropsychologia* **101**, 47–56. doi:10.1016/j.neuropsychologia.2017.04.005

Havas V, Taylor J, Vaquero L, de Diego-Balaguer R, Rodríguez-Fornells A and Davis MH (2018) Semantic and phonological schema influence spoken word learning and overnight consolidation. *Quarterly Journal of Experimental Psychology* **71**(6), 1469–1481. doi:10.1080/17470218.2017.1329325

Hed A, Schremm A, Horne M and Roll M (2019) Neural correlates of second language acquisition of tone-grammar associations. *The Mental Lexicon* **14**(1), 98–123. doi:10.1075/ml.17018.hed

Hollingshead AB (1975). *Four factor index of social status*. Unpublished manuscript, Yale University, New Haven, CT.

ISO. (2004). ISO 389-8. Acoustics: Reference zero for the calibration of audiometric equipment. Part 8: Reference equivalent threshold sound pressure levels for pure tone and circumaural earphones. *International Organization for Standardization 389-8.*

Jung T-P, Makeig S, Humphries C, Lee T-W, McKeown MJ, Iragui V and Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**, 136–178. doi:10.1111/1469-8986.3720163

Kim A and Osterhout L (2005) The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language* **52**(2), 205–225. doi:10.1016/j.jml.2004.10.002

Kimppa L, Kujala T, Leminen A, Vainio M and Shtyrov Y (2015) Rapid and automatic speech-specific learning mechanism in human neocortex. *NeuroImage* **118**, 282–291. doi:10.1016/j.neuroimage.2015.05.098

Krott A and Lebib R (2013) Electrophysiological evidence for a neural substrate of morphological rule application in correct wordforms. *Brain Research* **1496**, 70–83. doi:10.1016/j.brainres.2012.12.012

Kutas M and Federmeier KD (2014) Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology* **62**, 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas M and Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* **207**(4427), 203–205. doi:10.1126/science.7350657

Kutas M and van Petten C (1990) Electrophysiological perspectives on comprehending written language. *New Trends and Advanced Techniques in Clinical Neurophysiology* **41**, 155–167. doi:10.1016/b978-0-444-81352-7.50020-0

Lehmann D and Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology* **48**(6), 609–621. doi:10.1016/0013-4694(80)90419-8

Leminen A, Kimppa L, Leminen MM, Lehtonen M, Mäkelä JP and Shtyrov Y (2016) Acquisition and consolidation of novel morphology in human neocortex: A neuromagnetic study. *Cortex* **83**, 1–16. doi:10.1016/j.cortex.2016.06.020

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press

Lozano-Argüelles C, Sagarra N and Casillas JV (2019) Slowly but surely: Interpreting facilitates L2 morphological anticipation based on suprasegmental and segmental information. *Bilingualism: Language and Cognition* 1–11. doi:10.1017/S1366728919000634

Malins JG, Desroches AS, Robertson EK, Newman RL, Archibald LM and Joanisse MF (2013) ERPs reveal the temporal dynamics of auditory word recognition in specific language impairment. *Developmental Cognitive Neuroscience* **5**, 134–148. doi:10.1016/j.dcn.2013.02.005

Maris E and Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177–190. doi:10.1016/j.jneumeth.2007.03.024

McKean C, Letts C and Howard D (2013) Functional reorganization in the developing lexicon: separable and changing influences of lexical and phonological variables on children's fast-mapping. *Journal of Child Language* **40**, 307-335. doi:10.1017/S0305000911000444

McLaughlin J, Osterhout L and Kim A (2004) Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience* **7**(7), 703–704. doi:10.1038/nn1264

McLaughlin J, Tanner D, Pitkänen I, Frenck-Mestre C, Inoue K, Valentine G and Osterhout L (2010) Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning* **60**(Suppl. 2), 123–150. doi:10.1111/j.1467-9922.2010.00604.x

Molinaro N, Barber HA, Caffarara S and Carreiras M (2014) On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, 1–4. doi:10.1016/j.cortex.2014.06.009

Morgan-Short K, Steinhauer K, Sanz C and Ullman MT (2012) Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience* **24** (4), 993–947. doi:10.1162/jocn_a_00119

Mueller JL, Hahne A, Fujii Y and Friederici AD (2005) Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience* **17**(8), 1229–1244. doi:10.1162/0898929055002463

Mueller JL, Hirotani M and Friederici AD (2007) ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners. *BMC Neuroscience* **8**(18). doi:10.1162/0898929055002436

Nash AJ and Williams CS (1982) Effects of preparatory set and task demands on auditory event-related potentials. *Biological Psychology* **15**, 15–31. doi:10.1016/0301-0511(82)90028-X

Neville H, Nicol JL, Barss A, Forster KI and Garret MF (1991) Syntactically based sentence processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience* **3**(2), 151–165. doi:10.1162/jocn.1991.3.2.151

Nigam A, Hoffman JE and Simons RF (1992) N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience* **4**(1), 15–22. doi:10.1162/jocn.1992.4.1.15

Ojima S, Nakata H and Kakigi R (2005) An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience* **17**(8), 1212–1228. doi:10.1162/0898929055002436

Osterhout L and Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* **31**, 785–806. doi:10.1016/0749-596X(92)90039-Z

Osterhout L, Poliakov A, Inoue K, McLaughlin J, Valentine G, Pitkanen I, Frenck-Mestre C and Hirschensohn J (2008) Second-language learning and changes in the brain. *Journal of Neurolinguistics* **21**(6), 509-521. doi:10.1016/j.jneuroling.2008.01.001

Partanen E, Leminen A, de Paoli S, Bundgaard A, Skjold Kingo O, Krøjgaard P and Shtyrov Y (2017) Flexible, rapid and automatic neocortical word form acquisition mechanism in children as revealed by neuromagnetic brain response dynamics. *NeuroImage* **155**, 450–459. doi:10.1016/j.neuroimage.2017.03.066

Pelzl E, Lau EF, Guo T and DeKeyser R (2019) Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition* **41**, 59–86. doi:10.1017/S0272263117000444

Pratarelli ME (1994) Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition* **24**, 137–157. doi:10.1006/brcg.1994.1008

Roll M, Gosselke S, Lindgren M and Horne M (2013) Time-driven effects on processing grammatical agreement. *Frontiers in Psychology* **4**(1004), 1–8. doi:10.3389/fpsyg.2013.01004

Roll M, Horne M and Lindgren M (2007) Object shift and event-related brain potentials. *Journal of Neurolinguistics* **20**(6), 462–481. doi:10.1016/j.jneuroling.2007.04.001

Roll M, Söderström P, Mannfolk P, Shtyrov Y, Johansson M, van Westen D and Horne M (2015) Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain & Language* **150**, 14–21. doi:10.1016/j.bandl.2015.07.009

Rossi S, Gugler MF, Friederici AD and Hahne A (2006) The impact of proficiency on syntactic second-langauge processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience* **18**(12), 2030–2048. doi:10.1162/jocn.2006.18.12.2030

Sagarra N and Casillas JV (2018) Suprasegmental information cues morphological anticipation during L1/L2 lexical access. *Journal of Second Language Studies* **1**(1), 31–59. doi:10.1075/jsls.17026.sag

Schremm A, Novén M, Horne M and Roll M (2019) Brain responses to morphologically complex verbs: An electrophysiological study of Swedish regular and irregular past tense forms. *Journal of Neurolinguistics* **51**, 76–83. doi:10.1016/j.jneuroling.2019.01.006

Shen G and Froud K (2019) Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: an ERP study. *Bilingualism: Language and Cognition* **22**(2), 253-265. doi:10.1017/S136672891800038X

Silva S, Vigário M, Leone Fernandez B, Jerónimo R, Alter K and Frota S (2019) The sense of sounds: Brain responses to phonotactic frequency, phonological grammar and lexical meaning. *frontiers in Psychology* **10** (681). doi:10.3389/fpsyg.2019.00681

So CK and Best CT (2010) Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* **53**(273), 273–293. doi:10.1177/0023830909357156

Steinhauer K (2014) Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics* **35**(4), 393–417. doi:10.1093/applin/amu028

Tanner D, McLaughlin J, Herschensohn J and Osterhout L (2012) Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition* **16**, 367–382. doi:10.1017/S1366728912000302

Tokowicz N and MacWhinney B (2005) Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition* **27**, 173–204. doi:10.1017/S0272263105050102

Unsworth N, Heitz RP, Schrock JC and Engle RW (2005) An automated version of the operation span task. *Behavior Research Methods* **37** (3), 498–505. doi:10.3758/BF03192720

Wayland RP and Li B (2008) Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics* **36**, 250–267. doi:10.1016/j.wocn.2007.06.004

Wicha NY, Moreno EM and Kutas M (2003) Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending line drawings within a written sentence in Spanish. *Cortex* **39**, 483–508. doi:10.1016/S0010-9452(08)70260-0

Williams SM (2013). *A major revision of the Edinburgh Handedness Inventory.* doi:10.13140/2.1.5180.9284

Yang C and Chan MK (2010) The perception of Mandarin Chinese tones and intonation by American learners. *Journal of the Chinese Language Teachers Association* **45**(1), 7–36.

Yang J and Li P (2019) Mechanisms for auditory perception: A neurocognitive study of second language learning of Mandarin Chinese. *Brain Sciences* **9** (139). doi:10.3390/brainsci9060139

Yu K, Li L, Chen Y, Zhou Y, Wang R, Zhang Y and Li P (2019) Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology* **56**:e13448. doi:10.1111/psyp.13448

https://doi.org/10.1017/S1366728921000134 Published online by Cambridge University Press