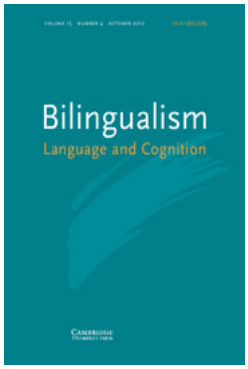

Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise

GEORGINA OLIVER, MARIANNE GULLBERG, FRAUKE HELLWIG, HOLGER MITTERER and PETER INDEFREY

Bilingualism: Language and Cognition / Volume 15 / Issue 04 / October 2012, pp 841 - 857
DOI: 10.1017/S1366728912000089, Published online:

Bilingualism: Language and Cognition 15 (4), 2012, 841–857 © Cambridge University Press 2012 doi:10.1017/S1366728912000089

Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise*

GEORGINA OLIVER
Max Planck Institute for Psycholinguistics, Nijmegen & University College London, Department for Speech, Hearing and Phonetic Sciences
MARIANNE GULLBERG
Max Planck Institute for Psycholinguistics, Nijmegen & Lund University, Centre for Languages and Literature
FRAUKE HELLWIG
Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour & Heinrich Heine University Düsseldorf, Institut für Sprache und Information
HOLGER MITTERER
Max Planck Institute for Psycholinguistics, Nijmegen
PETER INDEFREY
Max Planck Institute for Psycholinguistics, Nijmegen & Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour & Heinrich Heine University Düsseldorf, Institut für Sprache und Information

(Received: February 6, 2011; final revision received: February 2, 2012; accepted: February 14, 2012; first published online 27 March 2012)

This study investigated the development of second language online auditory processing with ab initio German learners of Dutch. We assessed the influence of different levels of background noise and different levels of semantic and syntactic target word predictability on word-monitoring latencies. There was evidence of syntactic, but not lexical-semantic, transfer from the L1 to the L2 from the onset of L2 learning. An initial stronger adverse effect of noise on syntactic compared to phonological processing disappeared after two weeks of learning Dutch suggesting a change towards more robust syntactic processing. At the same time the L2 learners started to exploit semantic constraints predicting upcoming target words. The use of semantic predictability remained less efficient compared to native speakers until the end of the observation period. The improvement and the persistent problems in semantic processing we found were independent of noise and rather seem to reflect the need for more context information to build up online semantic representations in L2 listening.

Keywords: sentence processing, second language, noise, word monitoring, longitudinal

Understanding spoken language is a complex matter. Most people learning a second language (L2) have experienced first-hand that listening comprehension in a second language is more difficult than in a first or native language (L1), especially in auditorily challenging circumstances such as on the telephone or with background noise as in a cafeteria. Listening to speech in noise is challenging even for native speakers, who have more difficulty identifying words the higher the background noise levels and the less predictable the word (e.g., Craig, 1988). However, a number of studies show that L2 speakers suffer a greater disadvantage than L1 speakers when listening to speech in noise even at advanced levels of proficiency although high L2 proficiency and early L2 acquisition seem to mitigate the effects (Florentine, 1985; Florentine, Buus, Scharf & Canevet, 1984; Garcia Lecumberri & Cooke, 2006; Kalikow, Stevens & Elliott, 1977; Mayo, Florentine & Buus, 1997; Nabelek & Donahue 1984; Roussohatzaki & Florentine, 1990; Shimizu, Makishima, Yoshida & Yamagishi, 2002; for exceptions, see Cutler, Weber, Smits & Cooper, 2004; Shi, 2010). Difficulties for L2 listeners have been found at the phonological level where the ability to identify phonemes is more adversely affected in L2 than in L1 listeners (Cutler, Cooke, Lecumberri & Pasveer 2007; Garcia Lecumberri & Cooke, 2006; Gong, 2006; Nabelek & Donahue, 1984), and at the semantic level where additional semantic context facilitates comprehension of single words in the sentence or the sentence as a whole in L1 but not in L2 listeners (Kilborn, 1992; Mayo et al., 1997; Shi, 2010). Similarly,

* This research was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). We would like to thank an anonymous reviewer and Jan Hulstijn for their helpful comments.

Address for correspondence:
Peter Indefrey, Heinrich Heine University Düsseldorf, Institut für Sprache und Information, Universitätsstr. 1, D-40225 Düsseldorf, Germany
Peter.Indefrey@phil.uni-duesseldorf.de

syntactic context improves comprehension for L1 listeners to a greater extent than for L2 listeners (Kilborn, 1992; Marslen-Wilson & Tyler, 1980). On the whole, then, non-native listeners' listening comprehension in noise seems to be less robust than native comprehension.

One way to account for the difficulties with L2 comprehension in noise is to relate them to the development of automatic processing in an L2. It is generally assumed that language processing in the beginning of L2 learning is highly controlled and effortful, and that during acquisition there is a development from controlled towards more automatic processing. The notion of automaticity is still not well defined in L2 research, but it is often discussed in terms of improved processing speed and accuracy, especially under stress, unintentional and effortless processing, and overall more efficient performance (e.g., DeKeyser, 2001; Segalowitz & Hulstijn, 2005; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz & Wood, 1998). It is sometimes argued that it involves ballistic processing, that is, processing that cannot be stopped once it has started (Segalowitz & Hulstijn, 2005). An assumption is that continuous exposure and use of the L2 leads to the restructuring of formerly controlled global processes into (partially) automatized sub-processes or modules which then do not tax working memory and operate in a ballistic manner (Favreau & Segalowitz, 1983). In this sense, automaticity is related to skill improvement via practice (see Newell & Rosenbloom, 1981) through a power law stating that plotting the logarithm of the time to perform a task against the logarithm of the trial number always yields a more or less straight line. According to DeKeyser (2001), processes once automatized do not further benefit from practice and additional working memory load does no longer interfere with these automatized processes. This more efficient functioning of processes leads to less effort in processing and thus increases the robustness of L2 comprehension when the L2 speaker is faced with stressors such as noise. Noise disturbs an auditory processing system which is not yet error-free and fail-safe in identifying L2 sounds and words and also adds to the processing load in the L2. The processing system is in a delicate balance and may just manage to function in the L2 when there are no stressors. Noise of any kind upsets this balance and could make functioning in an L2 harder or even impossible.

So far, we have discussed studies identifying difficulties in L2 listening comprehension. Another line of research investigated in how far the ability to build up comprehension in a second language is connected to properties of the languages involved (Odlin, 2003). L2 learners could draw on their specific L1 knowledge to help them understand the L2. That is, L2 listeners could be relying on cross-linguistic influence or positive transfer in processing the L2 (Sharwood-Smith & Kellerman, 1986; Odlin, 2003). In this sense, the more similar the L1 and the L2 are with respect to certain properties, the more robust the L2 processing of these properties or others that build on them may be even in noise. A large literature documents L1 transfer onto the L2 in offline tasks (see Odlin, 2003 for an overview) and in online lexical processing (Dijkstra, Grainger & van Heuven, 1999). By contrast, the precise role of transfer in online written L2 sentence processing is still under debate (e.g., Clahsen & Felser, 2006; Frenck-Mestre, 2005; Odlin, 2003; Roberts, Gullberg & Indefrey, 2008). Little is known about such transfer effects in online auditory sentence comprehension. In the case of learners learning a closely related L2, a facilitating influence of L1 syntax and semantics mitigating adverse noise effects could be expected but the extent and onset of such influences need to be further explored.

The aim of the present study was to complement the current knowledge on L2 comprehension in background noise with respect to several points. First, we wanted to obtain evidence on how L2 comprehension in background noise, and hence the robustness of L2 comprehension, might change over time when learning an L2 from scratch until a certain level of proficiency is reached. Secondly, we wanted to obtain further evidence on the differential impact of noise on phonological, syntactic, and semantic processing of auditory sentences and its possible interaction with increasing proficiency. To this end we chose a unique group of learners from a closely related L1 (German) to whom special intensive language courses are offered in which they achieve a level of proficiency enabling them to study at a Dutch university in a few weeks. The steep learning curve of German learners of Dutch is obviously due to the fact that German and Dutch share large parts of phonology, lexicon, and syntax so that German learners of Dutch can transfer many aspects of their L1 knowledge to the L2. However, we still lack knowledge on whether this transfer occurs during learning or during online processing and, more specifically, which aspects of L1 knowledge might be used during online auditory sentence processing. Our third research question, therefore, was whether there would be any evidence of the transfer of L1 knowledge at a very early stage of L2 learning when the participants would have had little opportunity to acquire L2 knowledge.

We addressed these questions in a study on L2 Dutch sentence comprehension. German speakers learning Dutch as a second language in an intensive language course and a control group of native Dutch speakers were tested using a standard word-monitoring paradigm (Marslen-Wilson & Tyler, 1980) in which participants respond to a pre-specified target word in an auditorily presented sentence. This paradigm allows the assessment of sentence processing at different levels by embedding the target words in different types of (pseudo-)sentential contexts. Reaction times to target words in random word strings (random prose) reflect

bottom–up word recognition ability, because target words are unpredictable from the preceding context. Reaction times in syntactically correct but meaningless sentences (syntactic prose) relative to random prose reflect the ability to use syntactic structure as a predictive cue for upcoming target words. Reaction time in normal sentences (normal prose) relative to syntactic prose reflect the ability to use the semantic content of the preceding sentence context to predict upcoming target words. In addition to the manipulation of sentential contexts, we varied the position of the target words in the carrier sentences to obtain evidence on the amount of syntactic or semantic context information needed to facilitate target word recognition. Finally, we added varying levels of multi-speaker babble as background noise to test the robustness of L2 processing. Multi-speaker babble or "cafeteria noise" is a highly disruptive informational masker, which in contrast to energetic maskers such as pink or white noise is encountered in everyday situations.

We tested the native German speakers three times over an observation period of three to four months. Session 1 occurred within five days of starting to learn the L2 (Dutch), Session 2 occurred two weeks after Session 1, and Session 3 occurred three months after the end of the language course. Based on previous findings, we hypothesized that all speakers should be affected by noise, but L2 speakers should be more heavily affected than L1 speakers are. We also expected L2 learners to become less affected by noise over time with the development of processing automaticity. With respect to target word predictability we hypothesized that reaction times to embedded target words would become increasingly longer from normal prose over syntactic prose to random prose conditions for both L1 and L2 speakers. We also anticipated an emergent change in the effect of target word predictability over time for L2 speakers due to the learning of new words and syntactic structures in the L2 during the intensive language course. We had no specific hypothesis with respect to the effect of predictability in the first session. By that time the German participants had just started to learn the L2 Dutch. Under an assumption of no lexical or syntactic transfer between L1 and L2, they would have no knowledge of either syntactic structure or lexis and hence should be unable to use semantic or syntactic constraints for faster target word recognition. On the other hand, since the learners' native language was closely related to the target language, they might show semantic or syntactic context effects already in Session 1.

Methods

Participants

The participants were 12 native German speakers (10 female) with a mean age of 19 years (SD = 0.6) and a mean length of education of 13 years (SD = 0) recruited from two intensive Dutch language courses specifically tailored to German speakers, with a mandatory state Dutch language exam being taken at the end of the course. The intensive language courses consisted of language lessons for four-and-a-half or six weeks with six hours of teaching a day five days a week and additional homework every day. The participants were preparing to start a university education in the Netherlands for which a "pass" at the Dutch language state exam (Dutch "NT2-examen", version for higher education) is a prerequisite for a foreigner. The participants were recruited on the first day of the courses and took part in the first experimental session within the first five days. Except for a few days between arrival in the Netherlands and the beginning of the course, the participants reported no exposure to Dutch before the courses.

The participants completed a detailed language background questionnaire (Gullberg & Indefrey, 2003) on their language learning background, proficiency in foreign languages, valence toward Dutch and towards language learning in general (see Table 1).

Controls were 18 native Dutch speakers (15 female) with a mean age of 20.4 years (SD = 2.3) and a mean length of education of 13.3 years (SD = 1.6), who were students recruited from Radboud University Nijmegen, The Netherlands. All participants had normal or corrected-to-normal vision and normal hearing, and were paid for participation.

Materials

In an auditory word-monitoring task, participants listened for a visually pre-specified target word in stimulus sentences (mean length = 9.5 words, range = 7–12 words) and pressed a button as soon as they had recognized the target word. Target word positions in the sentence were distributed approximately equally across sentences (34% in the first third, 30% in the second third, and 36% in the last third of the sentence). Target words were predominantly concrete nouns of 14 semantic categories (animals, persons, body parts, buildings, parts of buildings, furniture, time expressions, kitchen utensils, tools, office utensils, sports, fruits and vegetables, clothing, and vehicles) chosen from the 1,000 most frequent Dutch nouns (CELEX; Baayen, Piepenbrock & van Rijn, 1998). The average length of the target words was 5.13 phonemes (SD = 2.24, range = 2–12). The number of syllables ranged from one to five (mean = 1.79, SD = 0.87).

For every target word, there were three sentences with different target predictability levels (normal prose, syntactic prose, random prose; see Table 2 for examples). Every sentence occurred with three different levels of background noise (no noise, some noise, loud noise).

Table 1. *Language learning background, Dutch proficiency, and valence towards Dutch of the German participants.*

Language learning background			
Languages spoken	Number of participants	Age of onset (years)	Duration of exposure (formal tuition)
English	all	11	8 years
French	9	15	4 years
Latin	1		6 years
Russian	1		1 year
Valence towards learning new languages, self-rating		Median (Range)	
How well do you like to learn new languages? (1 = dislike, 5 = like)		4.0 (3–5)	
How easy do you find learning new languages? (1 = difficult, 5 = easy)		3.0 (1–4)	
DUTCH			
Mean age of onset (SD)		19.00 (0.60) years	
Mean duration of exposure at Session 1 (SD)		1.50 (0.67) weeks	
Valence towards Dutch, self-rating (1 = disagree, 5 = agree)	Session 1 Median (Range)		Session 3 Median (Range)
I like to speak Dutch	4.0 (3–5)		5.0 (3–5)
I am confident using Dutch	3.0 (1–4)		3.0 (3–4)
I find it important to be good at Dutch	5.0 (4–5)		5.0 (4–5)
Dutch proficiency self-rating (1 = not good, 5 = very good)	Session 1 Median (Range)		Session 3 Median (Range)
Speaking	1.5 (1–3)		3.0 (1–4)
Listening	2.0 (1–3)		4.0 (3–4)
Writing	2.0 (1–2)		3.0 (2–4)
Reading	2.5 (2–4)		4.0 (2–5)
Grammar	1.0 (1–2)		3.0 (1–4)
Pronunciation	1.0 (1–3)		2.5 (2–4)
Dutch proficiency CITO cloze test	Session 1 Mean % (SD)	Session 2 Mean % (SD)	Session 3 Mean % (SD)
	41.2 (20.3)	61.5 (19.0)	79.4 (11.0)

Table 2. *Stimulus examples of the three predictability conditions.*

Predictability condition	Stimulus example
Normal prose (NP)	Mijn oma heeft in haar slaapkamer een **gordijn** met bloemetjesmotief. My grandma has in her bedroom a **curtain** with flowers on it.
Syntactic prose (SP)	In het **gordijn** slaapt men's ochtends vroeg al met ruzie. In the **curtain** you sleep already early in the morning with fighting.
Random prose (RP)	Aan even als **gordijn** kinderen wereld sluwe van de uit bij mooie. At just as **curtain** children world crafty from the out by beautiful.

We created syntactic and random prose sentences from normal prose sentences by first exchanging the target words randomly between sentences. In syntactic prose sentences, we then replaced all other content words pseudo-randomly with content words from other normal prose sentences, avoiding content words that were semantically related to the target word and adapting articles, number and adjectival endings to the newly inserted content words to yield syntactically correct sentences. In random prose sentences, all other words were pseudo-randomly replaced with words from other sentences.

All stimulus sentences were recorded from a female native Dutch speaker in a sound-proof booth directly onto Digital Audio Tape with a high-quality microphone (Sennheiser ME64). The sampling rate during digitization was 44,100 Hz. Stimuli for the training session (see below) were recorded from a male native Dutch speaker to prevent acoustic memorization of target words. All stimuli were read at about the same speed (average durations of normal/syntactic/random prose sentences = 2.92/3.12/2.86 s) and the speakers were instructed to read all sentences, including the random predictability stimuli, with as normal a sentence intonation as possible. They were not informed about the experimental design or the target words prior to recording. The resulting sound files were cut into single sentences and saved as single files.

To ensure equal durations of the target word tokens as spoken in different sentence types, we time-coded 260 sentence triplets corresponding to the three predictability conditions (780 sentences in total) for target word onset, uniqueness point, offset, and onset of the following word. The uniqueness point (the phoneme at which target words became different from all other Dutch words and hence unambiguously identifiable, as established by a native Dutch speaker using a standard Dutch dictionary) ranged from −6 (= the sixth phoneme before the end of the word) to 0 (= onset of following word, i.e. the word was not unique at the last phoneme) with a mean uniqueness point of −1.37 (SD = 1.79).

We then selected 117 target words with corresponding sentence triplets (351 sentence stimuli in total) such that the average duration from target word onset to the uniqueness point was approximately equal in the three predictability conditions. Word onset to uniqueness point duration was on average 331 ms in normal prose sentences, 327 ms in syntactic prose sentences, and 331 ms in random prose sentences. Twenty-seven additional stimulus sentences (9 × 3 predictability conditions) taken from the discarded sentences were used for warm-up trials during the main experiment.

The background noise consisted of multi-speaker babble. The method of producing three background noise conditions (no noise, some noise: SNR = 8dB, loud noise: SNR = 0dB, i.e. equally loud background noise and stimulus sentence) was adapted from Cutler et al. (2004). To create the background noise files, we recorded naturally spoken conversation data from three mixed-sex pairs of native Dutch speakers aged between 19 and 40 years. We recorded directly onto Digital Audio Tape using a Sennheiser ME64 microphone in a soundproof booth. The sampling rate during digitization was 32,000 Hz. Speakers spoke about informal topics (weekend plans, work, hobbies, moving house, etc.) for 10–20 minutes. Discarding the initial five minutes of each recording, from each speaker six 10s stretches were selected during which no background noises and no target words from the main experiment were present and he or she was speaking alone at a not too loud or soft tone. From these 60 files, eight randomly picked files were amplitude-equalized and added together to form one multi-speaker babble file. Each of the 117 sentence triplets was assigned a unique multi-speaker babble file. The stimulus sentences were centrally embedded in multi-speaker babble segments that were 500 ms longer at each end. Stimulus sentences and babble segments were combined at the signal-to-noise ratios 0 dB and 8 dB.

Procedures

Prior to taking part in the main experiment, the participants underwent an extensive test battery and a word training session.

Pre-tests

All participants were tested for normal hearing before the first session. German participants completed a detailed language background questionnaire before the first session and after the last session. Before all three sessions, the German participants completed a written cloze test of Dutch as a second language (Dutch Central Institute for Language Teaching, CITO). This type of test was chosen because cloze tests are efficient tools for a general assessment of language proficiency across a range of knowledge domains (productive and receptive vocabulary, grammar, semantics, pragmatics and orthography, see Hulstijn, 2010). Furthermore, we considered a gap-filling test to be suitable for repeated testing, because no correct response alternatives are provided and it is unlikely that memorizing any aspects of the test material would be beneficial on a re-test. Native speaker scores on this test were obtained from Dutch participants, who were matched to the German participants with respect to education and age. They did not participate in the main experiment.

Word training session

All participants completed a 10–30-minute self-paced training session on the target words used in the main experiment. The target words were presented in auditory and written form and illustrated with black and white

line-drawings. The participants were instructed to listen to and read every word at least once until they affirmed they were familiar with all the target words and understood their meaning. The Dutch control participants also completed the word training session and were instructed to listen to and read every word at least once.

Word-monitoring experiment

During the self-paced auditory word-monitoring task, participants monitored for a previously specified target word through high-quality Sennheiser headphones while seated in a soundproof booth in front of a computer screen with a two-button response box. Volume was adjusted to a comfortable level for each participant. The written target word was presented on a computer screen until the participants pressed the left-hand button on a two-button response box. Upon button-press, the auditory stimulus sentence started immediately and the participants pressed the right-hand button as soon as they had identified the target word. They were instructed not to wait with pressing the button until the target word was finished but to press the button as quickly as possible upon identification of the target word.

The German participants took part in the experiment three times: Session 1 was within five days of starting the intensive language course, Session 2 was two weeks after Session 1 and Session 3 was three months after the intensive language course had been completed. In addition to the language course, the German participants had had six weeks of tuition in Dutch at a Dutch university by Session 3. Dutch control participants took part in the experiment once.

In each session, participants were presented with a stimulus set of 351 stimulus sentences in nine blocks (three blocks of every predictability condition) of 39 sentences (13 per noise condition, resulting in a total of 117 per noise condition). The order of blocks was pseudo-randomized avoiding immediate repetition of predictability type. Each block started with a short warm-up period. Noise conditions varied randomly within each block. We created three stimulus sets with different noise levels for each sentence and different orders of sentences. The assignment of stimulus sets to sessions varied between participants. Across sessions, participants heard every sentence in the three different noise conditions.

Results

Language proficiency test

The mean score of German learners of Dutch on the Dutch language proficiency test was 41.2% correct (SD = 20.3) in Session 1, 61.5% correct (SD = 19.0) in Session 2, and 79.4% correct (SD = 11.0) in Session 3. Age and education matched native speakers scored on average 95.5% correct (SD = 3.5).

Word monitoring in noise

Button presses between 100 ms and 2,000 ms after target word onset were considered acceptable responses. Reaction times above or below three standard deviations from the average reaction time (RT) of a participant were considered as outliers and removed from further analysis. The remaining 91.6% of the trials of Dutch participants and 92.1% of the trials of German participants were entered into further analyses.

L1 and L2 speakers

Figure 1 shows the mean reaction times for Dutch participants and the first session of the German participants. Reaction times for Dutch and German participants were entered into a repeated measures analysis of variance (ANOVA, see Table 3) with the within-subjects factors Predictability (normal prose, syntactic prose, random prose) and Noise (no noise, some noise, loud noise), and the between-subjects factor Group (Dutch, German). The analysis showed significant main effects of Predictability and Noise. A significant main effect of group was only found in the item analysis. There were significant interactions between Predictability and Group and Predictability and Noise

Separate repeated measures ANOVAs for Dutch and German participants with the within-subjects factors Predictability and Noise showed significant main effects of Predictability and Noise for both groups. A significant Predictability × Noise interaction was found for both groups in the analysis of subject means but only for German participants in the analysis of item means. Bonferroni-corrected post-hoc comparisons showed for both groups significantly ($p < .000$) longer RTs for the random prose condition compared to the other prose conditions and significantly longer RTs for the loud noise condition compared to the other two noise conditions. RTs of the German but not the Dutch participants were significantly (subject means: $p < .041$, item means: $p < .001$) longer for the intermediate noise condition compared to the no noise condition.

In sum, native Dutch participants recognized the target words on average 19 ms faster than German L2 participants, but this difference was only significant by items not by subjects. In both groups, RTs were longer in the random prose condition and at the loudest noise level compared to the other levels of these factors. Overall, noise affected both groups equally (no significant interaction of Noise and Group), but unlike native Dutch participants German participants recognized target words more slowly at an intermediate noise level compared to no noise. In both groups, an interaction of Noise

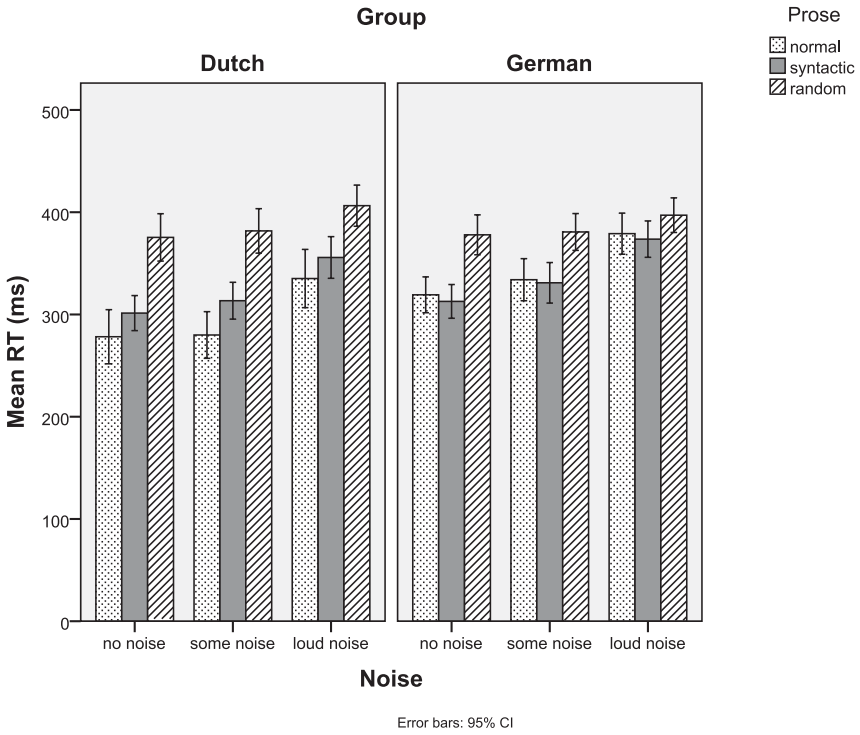

Figure 1. Mean target word recognition latencies for Dutch native speakers and the first session of German learners of Dutch. Bars show push-button response latencies after target word onset for the three predictability conditions and the three noise levels collapsed over all target word positions.

and Predictability reflected a stronger adverse effect of noise on target recognition in normal and syntactic prose compared to random prose. This effect was stronger in the German group and did not reach significance in the item analysis of the Dutch group.

Finally, the two groups were differentially affected by the prose conditions. Considering that semantic and syntactic constraints build up over the sentence, we investigated this interaction between Predictability and Group more closely by conducting further ANOVAs in which the position of the target word in the sentence was taken into account (see Table 3). Note that across participants a particular target word appeared in all combinations of the factors Group, Noise, and Predictability but not in all sentence positions to limit the number of item repetitions to three per session. Therefore, no ANOVAs on item means were conducted with the factor Sentence Position.

An overall repeated measures ANOVA with the within-subjects factors Sentence Position (first, second and third), Noise, and Predictability and the between-subjects factor Group again showed main effects of Noise and Predictability as well as an interaction of the two factors but no other significant interaction involving the factor Noise. There was no main effect of Group but an interaction of Predictability and Group. The factor Sentence Position had a main effect and interacted with Predictability and Group. In addition, there was a Sentence Position × Predictability × Group interaction. As can be seen in Figure 2, target words were recognized faster towards the end of the sentence. From the second third of the sentence onwards, Dutch listeners recognized target words in normal prose sentences faster than target words in syntactic prose and faster than German listeners. Separate ANOVAs with the within-subjects factor Predictability and the between-subjects factor Group for the three Sentence Positions confirmed that there was no main effect of Group or Predictability × Group interaction in the first third of the sentence, no main effect of Group but a significant Predictability × Group interaction in the second third of the sentence, and both a main effect of Group and a significant Predictability × Group interaction in the last third of the sentence. Separate analyses for the two groups with Bonferroni-corrected pairwise post-hoc comparisons showed that from the second third of the sentence onwards reaction times to targets in normal prose sentences were significantly faster than reaction times to targets in syntactic prose sentences for Dutch listeners (second third: $p = .006$; third: $p = .000$) but not for German listeners (second third: $p = 1$; third third: $p > .2$).

Figure 3 shows reaction times for both groups in the last part of the sentence. Whereas reaction times in syntactic prose sentences were similar in the two groups, Dutch participants showed clear evidence of exploiting the additional semantic information in normal

Table 3. *Summary of ANOVA results on word monitoring in noise showing main effects and interactions of the factors Noise (no noise, some noise, loud noise), Predictability (normal prose, syntactic prose, random prose), Sentence Position (target words appeared in the first, second, or last third of the sentence), and Group (Dutch, German). Greenhouse–Geisser corrections for violations of the sphericity assumption were applied where appropriate.*

Repeated measures ANOVAs on subject and item means with factors Noise, Predictability, and Group						
Overall	F_1	*df*	Partial η^2	*F2*	*df*	Partial η^2
Predictability	57.74***	2, 56	.67	31.54***	1.88, 201.38	.23
Noise	133.25***	2, 56	.83	125.44***	2, 214	.54
Group	*ns*	2, 56	.16	28.61***	1, 107	.21
Predictability × Group	5.48**	2, 56	.25	15.03***	2, 214	.12
Predictability × Noise	9.29***			5.20***	4, 428	.05
Other interactions	*ns*			*ns*		
Dutch group						
Predictability	39.49***	2, 34	.70	46.36***	1.90, 219.95	.29
Noise	87.87***	2, 34	.84	64.32***	1.90, 219.84	.36
Predictability × Noise	4.75**	4, 68	.22	*ns*		
German group						
Predictability	49.87***	2, 22	.82	15.78***	2, 214	.13
Noise	53.65***	2, 22	.83	70.81***	2, 214	.40
Predictability × Noise	4.81*	2.35, 25.85	.30	3.57**	4, 428	.30

Repeated measures ANOVA on subject means with factors Sentence Position, Noise, Predictability, and Group			
	F_1	*df*	Partial η^2
Sentence Position	91.22***	1.42, 39.68	.77
Noise	121.24***	2, 56	.81
Predictability	61.01**	2, 56	.69
Group	*ns*		
Sentence Position × Predictability	25.94***	3.09, 86.38	.48
Sentence Position × Group	8.00**	2, 56	.22
Noise × Predictability	8.68***	4, 112	.24
Predictability × Group	7.50**	2, 56	.21
Sentence Position × Predictability × Group	5.53***	4, 112	.17
Other interactions	*ns*		
Last third of sentence			
Repeated measures ANOVAs on subject means with factors Noise and Predictability			
Dutch group			
Noise	76.35***	2, 34	.82
Predictability	79.36***	2, 34	.82
Noise × Predictability	*ns*		
German group			
Noise	63.31***	2, 22	.85
Predictability	35.83***	2, 22	.77
Noise × Predictability	3.43*	4, 44	.24

ns = not significant, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

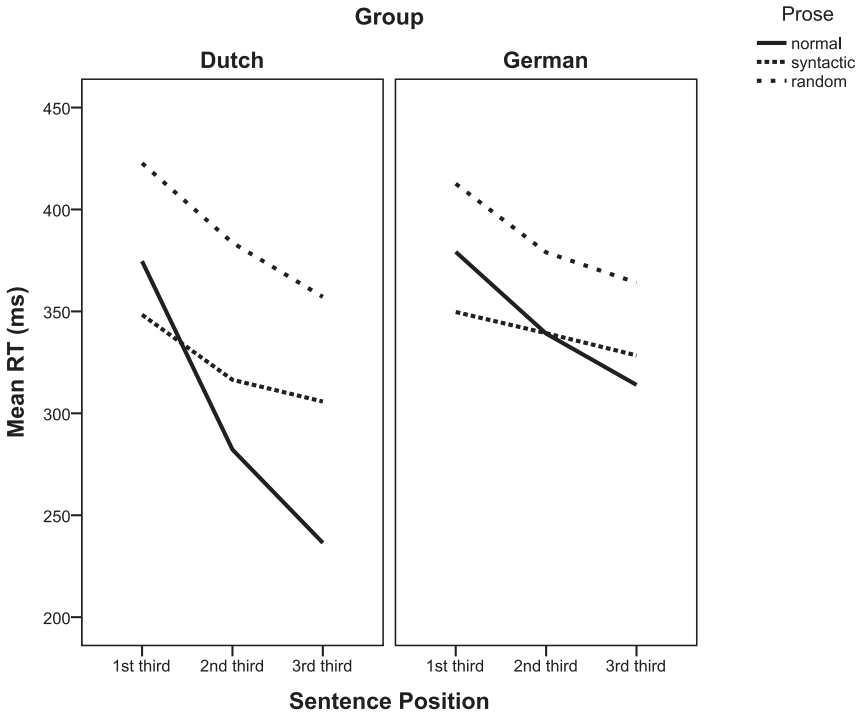

Figure 2. Mean recognition latencies for Dutch native speakers and German learners of Dutch in the three predictability conditions for target words in different sentence positions collapsed over noise levels.

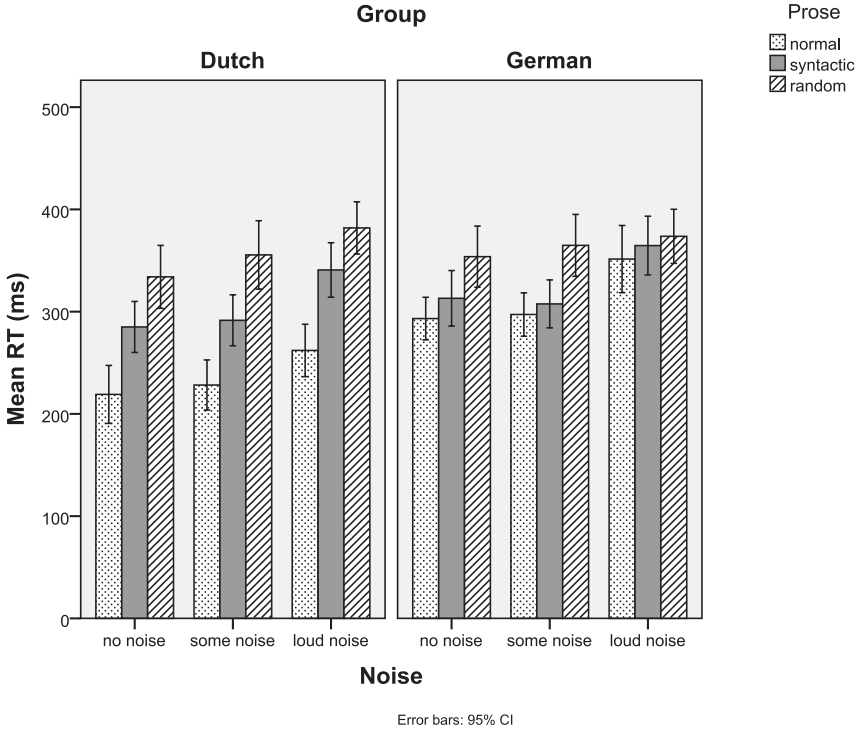

Figure 3. Mean recognition latencies for Dutch native speakers and German learners of Dutch for target words in the last third of the sentence.

Table 4. *Summary of ANOVA results on word monitoring in noise (German particants) showing main effects and interactions of the factors Session (1st, 2nd, 3rd), Noise (no noise, some noise, loud noise), and Predictability (normal prose, syntactic prose, random prose). Greenhouse–Geisser corrections for violations of the sphericity assumption were applied where appropriate.*

Repeated measures ANOVAs on subject means with factors Session, Noise, and Predictability			
Overall	F_1	*df*	Partial η^2
Session	9.71**	2, 22	.47
Noise	164.75***	2, 22	.94
Predictability	159.01***	2, 22	.94
Session × Noise	3.83**	2.26, 24.84	.26
Session × Predictability	3.18*	4, 44	.22
Noise × Predictability	4.62**	2.48, 27.27	.30
Session × Noise × Predictability	*ns*		
Session 1: see Table 3, German group			
Session 2			
Noise	48.46***	1.28, 14.06	.82
Predictability	81.01***	2, 22	.88
Noise × Predictability	*ns*		
Session 3			
Noise	25.14***	2, 22	.70
Predictability	104.31***	2, 22	.91
Noise × Predictability	*ns*		

ns = not significant, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

prose sentences for faster target recognition. German participants were less successful than Dutch participants in using syntactic information under conditions of loud noise. These observations were confirmed by repeated measures ANOVAs with the factors Noise and Predictability for the two groups (Table 3). Both groups showed main effects of Noise and Predictability but only the German participants a significant interaction of Noise and Predictability. Post-hoc Bonferroni-corrected ($p < .05$) paired samples *t*-tests showed significant differences between all three prose conditions under conditions of loud noise for Dutch participants (normal – syntactic: $t(17) = 7.56$; normal – random: $t(17) = 12.71$; syntactic – random: $t(17) = 4.33$), but no differences for German participants (all *p*s > .5).

In sum, both groups showed faster response times for sentences with syntactic structure compared to random prose already in the first third of sentence. Native Dutch but not German listeners exploited the additional semantic constraints provided by normal prose sentences for faster target word recognition from the second third of sentences onwards. Although noise did not affect the two groups differentially overall, loud noise hindered German listeners from using syntactic information.

L2 speakers over time

To identify effects of the acquisition of Dutch, the German participants' target word recognition latencies from the three experimental sessions (at the beginning, during, and three months after the language course) were entered into an overall repeated measures ANOVA with the within-subjects factors Session (first, second, third), Noise, and Predictability (see Table 4). In addition to main effects of Noise, and Predictability and an interaction of Noise × Predictability that were already observed in the analysis of the first session of the German group, this analysis showed a main effect of Session, reflecting a decrease of reaction times over sessions (see Figure 4). There were also interactions between Session and Noise, mainly reflecting a steeper decrease of reaction times between Sessions 2 and 3 in the loud noise compared to the other two noise conditions, and between Session and Predictability.

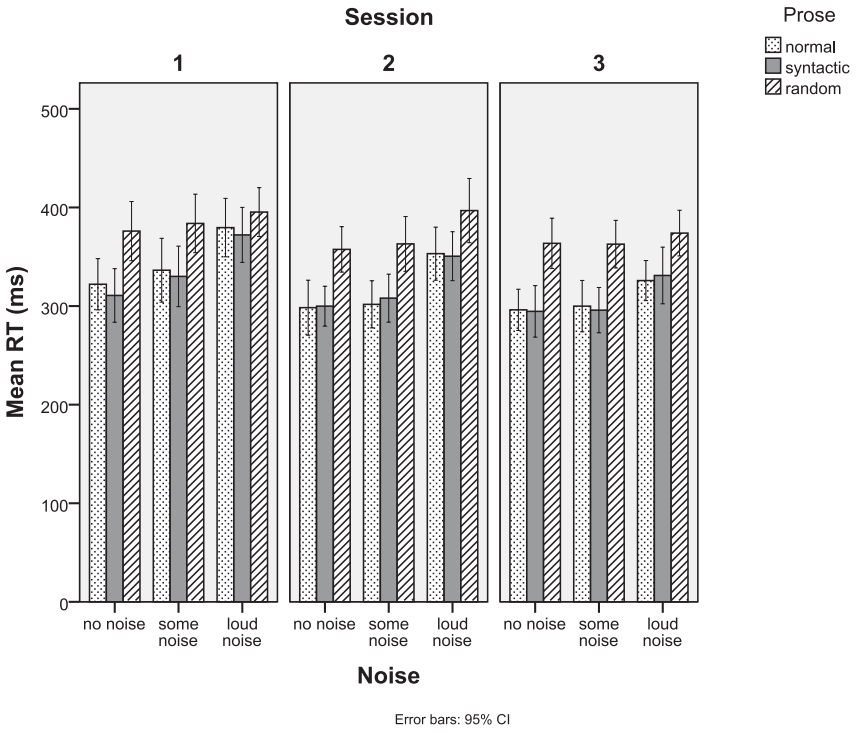

Figure 4. Mean target word recognition latencies in the three experimental sessions of German learners of Dutch collapsed over all target word positions.

To investigate the nature of the observed interaction between Session and Predictability more closely, we conducted separate ANOVAs for Sessions 2 and 3 with the within-subjects factors Predictability and Noise. These ANOVAs showed main effects of Predictability and Noise in both sessions. Bonferroni-corrected pairwise post-hoc comparisons showed significantly longer RTs for random prose sentences and longer RTs for the loud noise condition compared to the other levels of both factors in both sessions. Unlike the first session, RTs were not significantly longer in the intermediate noise compared to the no noise conditions. An interaction of Noise and Predictability reflecting a stronger adverse effect of loud noise in normal and syntactic compared to random prose sentences that was observed in the first session (see above) was no longer found in Sessions 2 and 3.

We also conducted separate analyses for the three sentence positions to identify effects of semantic constraints as they develop over the course of the sentence. From the second session onwards German listeners were able to exploit semantic constraints. They showed significantly faster (Bonferroni-corrected paired samples *t*-tests, $p < .01$) target word recognition in the last third of normal prose compared to syntactic prose sentences (see Figure 5). RTs between normal and syntactic prose conditions in the second third of the sentence in Sessions 1–3 and in the last third of the sentence in the first session did not differ significantly (all *p*s $> .4$).

As can be seen in Figure 6, loud noise did not compromise the newly acquired ability to exploit semantic constraints for faster word recognition in the last third of the sentence in Sessions 2 and 3. A repeated measures ANOVA of RTs to target words in the last third of the sentence under loud noise with the within-subjects factors Session (2, 3) and Predictability showed a main effect of Predictability ($F_1(2,22) = 40.12$, $p = .000$, partial $\eta^2 = .79$). There was no main effect of Session ($F_1 = 2.15$, $p > .1$) and Session did not interact with Predictability ($F_1 < 1$). Bonferroni-corrected post-hoc pairwise comparisons showed significant differences between normal compared to syntactic prose sentences ($p = .013$) and syntactic compared to random prose sentences ($p = .000$).

Assuming that the ability to use the semantic content of normal prose sentences for faster target word recognition was related to the acquisition of Dutch over the observation period, faster RTs in the last third of normal compared to prose sentences should be correlated with the L2 proficiency level that was reached. This was indeed the case. Scores on the CITO cloze test of Dutch L2 proficiency in the third session were significantly correlated ($r = .71$, $p = .009$) with the RT difference between target words in the last third of syntactic and normal prose sentences (see Figure 7). A correlation was already present in the second session but failed to reach significance ($r = .50$, $p = .102$).

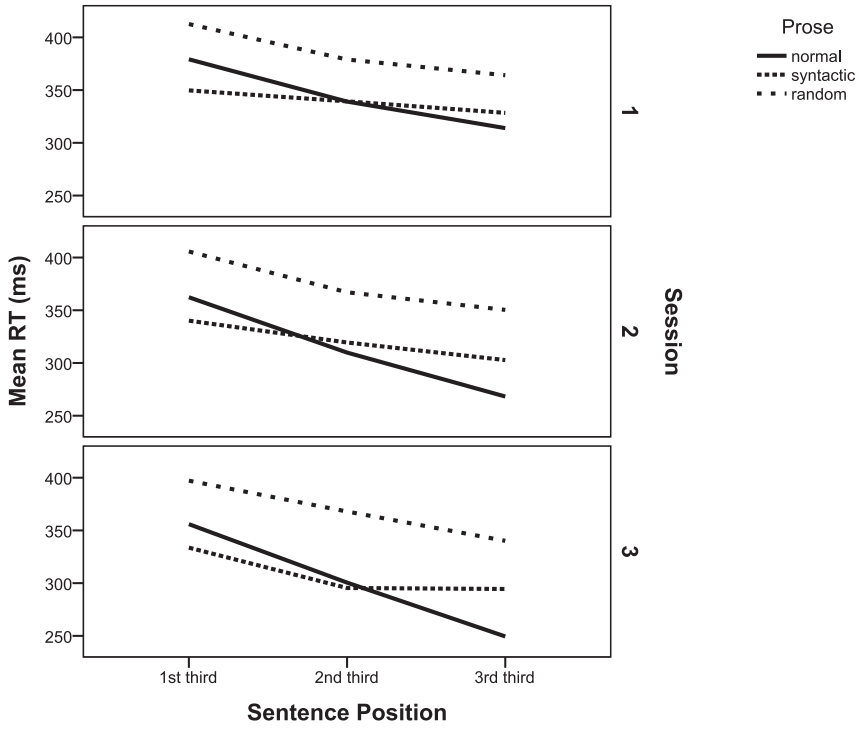

Figure 5. Mean recognition latencies in the three predictability conditions for target words in different sentence positions in the three experimental sessions of German learners of Dutch (collapsed over noise levels).

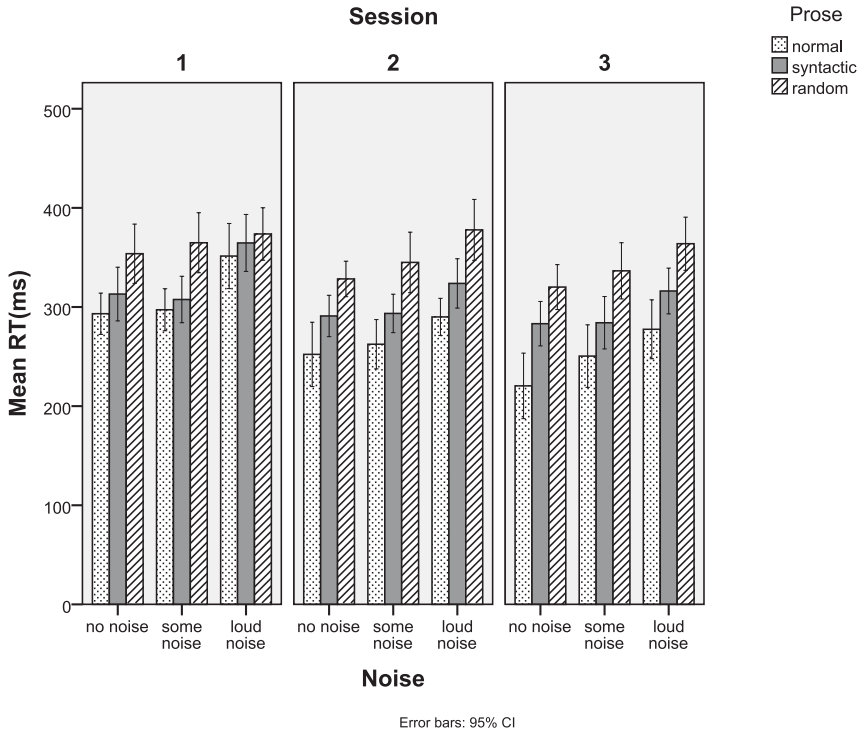

Figure 6. Mean recognition latencies for target words in the last third of the sentence in the three experimental sessions of German learners of Dutch.

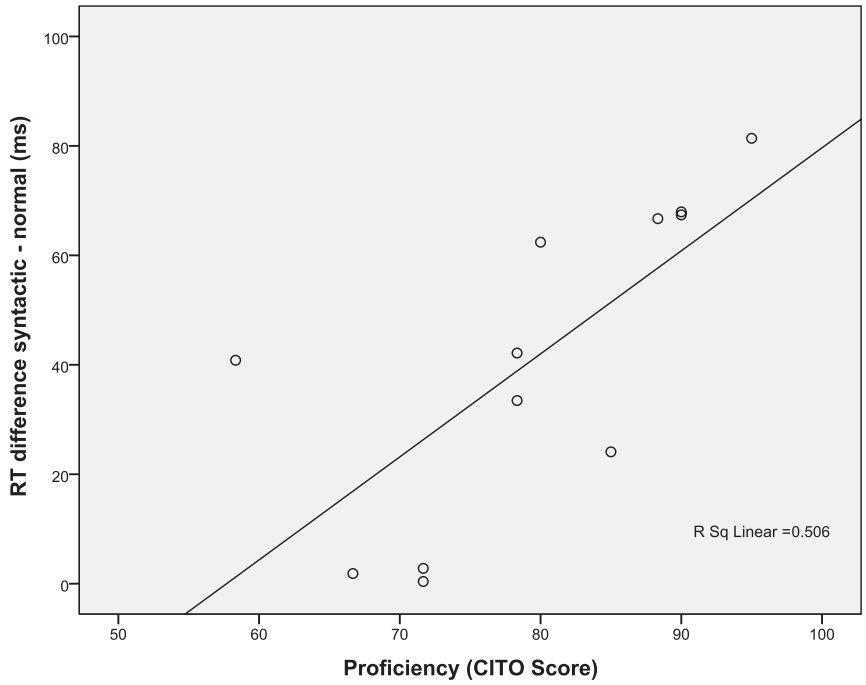

Figure 7. Reaction time advantage for target words in the last third of meaningful compared to syntactic prose sentences (RT difference syntactic – normal) plotted against Dutch proficiency (CITO) scores of the twelve German participants in the third experimental session.

In sum, over the span of the three sessions target word recognition in German learners of Dutch became more similar to native Dutch speakers in two respects. First, subtle adverse effects of noise such as longer reaction times at an intermediate noise level compared to no noise and a reduced ability to use syntactic information for faster target word recognition under conditions of loud noise disappeared by Session 2. Second, the German learners of Dutch began using the semantic information in normal prose sentences for faster target word recognition from the second session onwards. However, their use of semantic information remained limited to the last third of the sentence, whereas Dutch listeners exploited semantic constraints already earlier on in the sentence.

Discussion

In this study, we investigated the auditory sentence processing ability of German learners of Dutch from the first days of an intensive language course until three months after the completion of the course. We used a standard word-monitoring task (Marslen-Wilson & Tyler, 1980) that allows the assessment of sentence processing at different levels by embedding the target words in different types of contexts. In addition, we embedded the sentences in increasing noise levels to assess the stress resistance of sentence comprehension as an indicator of the degree of automatic processing at different processing levels.

Given that noise affected the sentence processing levels in the two listener groups differentially, its effects are best discussed against the background of noise-independent differences between L1 and L2 listeners. We will therefore first discuss L1 and L2 listeners' performance with respect to word recognition, syntactic, and semantic comprehension, then changes in the L2 listeners' performance over time, and finally the differential impact of noise for the two listener groups and changes of that impact over time for the L2 listeners.

Similarities and differences between L1 and L2 listeners

Our study first of all showed that native German speakers who had just begun to learn Dutch performed just as well as native Dutch speakers on recognizing target words occurring at unpredictable positions in random prose context. Considering previous studies showing that the typological distance between native and target language affects monitoring performance at a phonemic level (Cutler et al., 2004; Cutler et al., 2007) it is likely that the closeness of German and Dutch played a role in our finding and that it cannot necessarily be generalized to more distant language pairs.

Furthermore, our results showed that both the German learners and the native Dutch speakers used syntactic cues for target word recognition. This ability required the

identification of syntactic cues in the continuous auditory input and the knowledge that these cues predicted an upcoming target noun. That the German participants were able to exploit syntactic predictability already in the first days of learning Dutch suggests a cross-linguistic influence or transfer in the area of syntax. It is not clear which syntactic cues the L2 speakers exploited, but possibilities include articles and possessive pronouns, which are similar in both languages concerned. Knowledge of their L1, for example, could have alerted them to the fact that a noun would be following shortly after an article. Recall, however, that due to practical constraints we could only test our participants after a few days of learning Dutch. Syntactic cues such as articles, signaling noun phrase structures are highly frequent in the course materials and could have been learned (with or without transfer) in the few days before our first experimental session. Insofar as this was the case, online transfer of L1 knowledge during the experiment would not have been necessary to exploit syntactic context. Although we cannot exclude this possibility in the present experiment, other data from the same group of learners (Davidson & Indefrey, 2009) show transfer of a syntactic structure that they could not have learned, because it is different between Dutch and German. In Davidson and Indefrey (2009) the German learners initially preferred the German word order of sentence-final verb clusters in Dutch sentences. It seems, therefore, at least plausible that L1 syntactic knowledge was also used in the present experiment.

Finally, the native Dutch speakers showed a classic pattern of a semantic predictability effect evolving over the course of sentence. Compared to syntactic prose, reaction times to target words in normal prose sentences became faster in the middle third of the sentence and the difference between the two conditions reached a 70 ms difference in the last third of the sentence. The German learners of Dutch, by contrast, showed no evidence for using semantic context information in their first experimental session. This was probably due to the fact that they had little lexical knowledge of the L2 at this time, and most L2 words were essentially meaningless to them. Even though the proportion of cognate words between Dutch and German is high (see below), the German learners could not know which Dutch words corresponded to similar sounding German words before having encountered the Dutch words plus their meaning at least once. Therefore, they could not use the semantic context to predict when a target word would occur in a stimulus sentence.

Changes in L2 listeners over time

In the German learners, a semantic predictability effect emerged two weeks after the first session and reached a 45 ms difference between semantic and syntactic prose in the last session. The size of this effect in Session 3 was correlated with scores in a Dutch proficiency test. The ability to build an online semantic representation of the context preceding the target words, therefore, seems to rely on the acquisition of Dutch rather than simply develop with the passage of time or experience with the task.

Interestingly, even in the third session, the semantic predictability effect for the learners was limited to the last part of the sentence suggesting that the German L2 speakers of Dutch, who by that time already studied at a Dutch university, still needed more semantic context than native Dutch speakers to successfully predict upcoming target words. Given the relatively short observation period, we cannot say whether this remaining difference would disappear after longer exposure to Dutch. Results by Mayo et al. (1997) and Shi (2010) suggest that even L2 speakers who have spoken the L2 since infancy may not reach L1 speakers' level of performance on word recognition in meaningful auditory sentences, especially under adverse listening conditions. Our results on semantic predictability differ from those obtained by Kilborn (1992), whose native German speakers of L2 English did not exploit the semantic information contained in a normal prose condition even though they were highly proficient and had lived in the United States for years. L2 vocabulary size, therefore, does not seem to be the only necessary component for successful online semantic processing. Considering again the typological closeness of German and Dutch one may assume that at least two further aspects help German L2 speakers of Dutch compared to English to build a semantic sentence representation before the target word comes up. One is the option to use syntactic L1 structures already discussed above which might free resources for semantic integration. A second aspect concerns differences in the proportions of cognate words and false friends for German and Dutch compared to German and English. Based on an analysis of a professional translation database, Schepens, Dijkstra and Grootjen (2012) report a much higher proportion of (near) cognates between Dutch and German (42%) than between English and German (15%). Due to the use of high frequency words, the proportion of (near) cognates among both target and context words was even higher (63%) in our materials. Cognate words are recognized (Caramazza & Brones, 1997; de Groot & Nas, 1991) and translated (de Groot, 1992) more quickly than non-cognates in isolation. A facilitatory cognate effect has also been shown for L2 sentence reading (Odlin, 1989). In addition, our L2 learners may have taken advantage of cognate vocabulary in Dutch and German in order to learn the lexis of the L2 more quickly and easily (Odlin, 2003) once they recognized the phonological correspondences between German and Dutch and were able to segment Dutch words from a continuous sound stream. Finally, not only are there less phonologically similar words

with a similar meaning between German and English, but German L2 speakers of English are also more strongly discouraged to exploit such correspondences. The proportion of words that are phonologically similar to a German word but have a different meaning ("false friends") is about twice as high in English as in Dutch (Schepens et al., 2012).

In sum, the results from this study show that syntax may be transferred before semantics in the auditory processing of sentences of a closely related L2. In the present case, syntactic transfer may depend in part on the cognate status of function words such as articles. Their limited number and high frequency may enable language learners to identify and use them for mapping L1 onto L2 syntax already after a few days of L2 contact. However, this finding does not necessarily mean that syntactic representations are built up as efficiently as in L1 speakers. This question was investigated by manipulating background noise. We will now turn to the effects of this manipulation and their interaction with syntactic and semantic predictability.

Effects of noise

Overall, the native German speakers were only marginally slower than native Dutch listeners and were not significantly more affected by noise. In particular, German listeners detected target words when they were not predictable from syntactic or semantic context under all noise conditions as fast as native speakers. Both groups showed the largest increase in reaction times between the intermediate and the loud noise level. There were two subtle learner-specific noise effects reflected in interactions with predictability conditions and sentence position. First, German learners, like Dutch controls, were able to use the syntactic context to predict upcoming target words but their ability suffered more from loud noise than that of the native speakers. Second, German learners showed longer reaction times for semantic and syntactic prose sentences at an intermediate noise level compared to no noise whereas native speakers showed no such increase. Both effects were no longer detectable from the second experimental session onwards.

We used the same type of multi-speaker babble background noise and the same noise levels as Cutler et al. (2004). Our results on the detection of visually pre-specified unpredictable target words are compatible with their finding that listeners whose L1 is closely related to the L2 target language are no more affected by noise than native listeners in a phoneme detection task. In the random prose condition, a phonemic representation of the target word had to be recognized in an incoming language stream without predictive cues and thus in principle this task was of a similar nature as, albeit more complex than, a phoneme detection task.

We cannot exclude that with more extreme noise levels or a different type of background noise a difference between Dutch and German listeners might have been found. For example, Hoen et al. (2007) found that multi-speaker babble based on four speakers had a worse effect upon L2 auditory comprehension than multi-speaker babble based on eight speakers. The stronger detrimental effect could be due to a heightened degree of informational masking with fewer speakers. Our procedure composing multi-speaker babble from six speakers thus may not have been maximally interfering with comprehension. Note, however, that Cutler et al. (2007) found differences in phoneme detection between native English speakers and Spanish L2 speakers of English with the same type and level of background noise we used here and hence it is in principle suitable for detecting comprehension difficulties in L2 speakers.

Previous studies investigating L2 speakers' perception of meaningful sentence stimuli in noise have mostly reported a stronger impact of noise on performance compared to native speakers (Florentine et al., 1984; Florentine, 1985; Kilborn, 1992; Mayo et al., 1997; Roussohatzaki & Florentine, 1990; but see Bradlow & Bent, 2002). In our study with L2 learners from a closely related language there was no overall stronger effect of noise on the German compared to the Dutch listeners (see also Shi, 2010, for a similar finding in late onset Russian and Spanish learners of English). On the other hand, the relatively stronger effect of noise on sentences with syntactic structure we observed in the first session supports at least one mechanism that has been suggested to cause difficulties for L2 learners. If noise is seen as a stressor taxing processes that are less robustly represented (or possibly less automatized, DeKeyser, 2001; Segalowitz & Hulstijn, 2005), then the ability of our German learners to use syntactic constraints to predict upcoming target words must draw on linguistic knowledge that is less robustly represented than in native listeners. This, in turn, means that the transfer of L1 syntactic knowledge we assume to underlie the observed ability to exploit syntactic predictability is not the same as simply using L1 knowledge representations for L2 processing. For example, even if the predictive value of words providing syntactic cues could be derived from L1 knowledge, the recognition of such words may still have been less efficient. It is also conceivable that, although L1 syntactic knowledge was retrieved for building up L2 syntactic structures during the parsing of L2 sentences, the process of building up L2 syntactic structures may nonetheless be less efficient.

In our L2 learners, this situation changed already two weeks later, when noise no longer affected disproportionately the use of syntactic constraints for faster target word recognition. Note that this change cannot simply be due to noise being less of a stressor in

the second session because the participants might have become used to it. The huge effect size of the factor Noise did not change between Session 1 and 2, and the effect of noise on the recognition of syntactically unpredictable target words in random prose (RT difference between no noise and loud noise) even increased from 19 ms in Session 1 to 31 ms in Session 2 (see Figure 4) indicating that the amount of processing resources needed for bottom–up target word recognition was certainly not reduced. The disappearance of a syntax-specific noise effect, therefore, seems to reflect a true qualitative change towards more robust syntactic processing.

Summary and conclusions

This study combines several issues that have so far been investigated separately in different types of L1 and L2 speakers. From first exposure onwards, we followed the development of L2 auditory processing in a relatively homogenous group of L2 learners who achieved a relatively high level of proficiency during the time course of the study. At the phonemic level, the use of natural connected speech stimuli added an extra dimension compared to single phoneme, syllable or single word stimuli by including segmentation and co-articulation effects. We varied the predictability of target words in syntactic and semantic contexts and investigated the influence of different levels of background noise in both L1 and L2 speakers.

The development of auditory processing skills in learners of Dutch from a closely related first language (German) resembled a fast motion picture of L2 acquisition. Their performance suggests phonological and syntactic transfer from the beginning, the development of noise-resistant syntactic processing, and the ability to build online semantic sentence representations within two weeks. In contrast to noise resistance of syntactic processing, however, online semantic processing did not reach a native-like level with higher L2 proficiency. The improvement and the persistent problems in semantic processing we found were independent of noise and rather seem to reflect the need for more context information to build up online semantic representations in L2 listening.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1998). The CELEX Dutch database. Release N32. http://www.mpi.nl/world/celex (accessed October 10, 2004).

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112 (1),* 272–284.

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society, 13,* 212–214.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10 (12),* 564–570.

Craig, C. H. (1988). Effect of three conditions of predictability on word-recognition performance. *Journal of Speech and Hearing Research, 31,* 588–592.

Cutler, A., Cooke, M., Lecumberri, M. L., & Pasveer, D. (2007). L2 consonant identification in noise: Cross-language comparisons. Presented at the Interspeech '07, Antwerp.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America, 116 (6),* 3668–3678.

Davidson, D. J., & Indefrey, P. (2009). Electrophysiological responses to crossed versus nested structures in German learners of Dutch. *Language and Cognitive Processes, 24,* 1335–1369.

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1001–1018.

De Groot, A. M. B., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30,* 90–123.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (ed.), *Cognition and second language instruction*, pp. 125–151. New York: Cambridge University Press.

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496–518.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first-and second-language reading of fluent bilinguals. *Memory and Cognition, 11 (6),* 565–574.

Florentine, M. (1985). Speech perception in noise by fluent, non-native listeners. *Journal of the Acoustical Society of America, 77,* s106.

Florentine, M., Buus, S., Scharf, B., & Canevet, G. (1984). Speech reception thresholds in noise for native and non-native listeners. *Journal of the Acoustical Society of America, 75,* s84.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research, 21,* 175–198.

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America, 119 (4),* 2445–2454.

Gong, J. (2006). Comparing non-native and native speech perception in noise. Masters thesis, University of Sheffield.

Gullberg, M., & Indefrey, P. (2003). Language background questionnaire. Nijmegen: Max Planck Institute for Psycholinguistics. http://www.mpi.nl/research/researchprojects/the-dynamics-of-multilingual-processing (accessed January 30, 2005).

Hoen, M., Meunier, F., Grataloup, C.-L., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., & Collet, L. (2007). Phonetic

and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication, 49,* 905–916.

Hulstijn, J. H. (2010). Measuring second language proficiency. In E. Blom & S. Unsworth (eds.), *Experimental methods in language acquisition research (EMLAR)*, pp. 185–199. Amsterdam: John Benjamins.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61 (5),* 1337–1351.

Kilborn, K. (1992). On-line integration of grammatical information in a second language. In R. Harris (ed.), *Cognitive processing in bilinguals*, pp. 337–350. Amsterdam: Elsevier.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

Mayo, L., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40,* 686–693.

Nabelek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *Journal of the Acoustical Society of America, 75,* 632–634.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (ed.), *Cognitive skills and their acquisition*, pp. 1–55. Hillsdale, NJ: Lawrence Erlbaum.

Odlin, T. (1989). *Language transfer*. New York: Cambridge University Press.

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*, pp. 436–486. Malden, MA: Blackwell.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). On-line pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30 (3),* 333–357.

Roussohatzaki, M., & Florentine, M. (1990). Perception of American-English in noise by Greek listeners. *Journal of the Acoustical Society of America, 87,* s72.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition, 15 (1),* 157–166.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 371–388. Oxford: Oxford University Press.

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14,* 369–385.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics, 19,* 53–67.

Sharwood-Smith, M., & Kellerman, E. (1986). Crosslinguistic influence in second language acquisition: An introduction. In E. Kellerman & M. Sharwood Smith (eds.), *Crosslinguistic influence in second language acquisition*, pp. 1–9. New York: Pergamon Press.

Shi, L.-F. (2010). Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech, Language, and Hearing Research, 53,* 821–835.

Shimizu, T., Makishima, K., Yoshida, M., & Yamagishi, H. (2002). Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx, 29,* 121–125.