Bilingualism: Language and Cognition 10 (3), 2007, 299–314 © 2007 Cambridge University Press doi:10.1017/S1366728907003082

Factors influencing L2 gender processing*

DENISA BORDAG
THOMAS PECHMANN
University of Leipzig

In four experiments we explored processes underlying L2 gender retrieval. We focused on L1 interference and on the influence of the L2 noun's termination. In Experiments 1 and 2 we tried to manipulate the intensity of L1 interference. We found that L2 speakers cannot eliminate or substantially reduce the interlingual interference neither when they know the response language long in advance in a situation in which code-switching is required (Experiment 1), nor when they are close to the monolingual mode (Experiment 2). Experiments 3 and 4 yielded evidence that gender typicality of the L2 noun's termination also exerts an influence on L2 processing, both in production and comprehension. L2 gender thus does not seem to be stored as a fixed feature as it is assumed for L1. Rather, our data support the assumption that it is computed anew each time when needed for processing. Further implications for modeling are discussed.

One of the topical questions concerning bilingual speech processing is the degree and scope of the interaction between the L1 and the L2 (or L3 etc.) language systems. Most models agree that L2 processing is not completely autonomous, but that it is affected by the earlier acquired L1 system. However, many questions still remain unanswered, e.g., which factors affect the degree of L2 activation relative to L1, at what levels the two systems interact, or how specific L2 features (e.g., grammatical gender) are represented and how this affects the possibility of L2 interference.

The interactive view is based on a generally accepted assumption that L2 speakers do not initially select the language-appropriate lexicon before making a lexical search, but rather that both lexicons are activated and searched simultaneously (De Bot, 1992; Green, 1993; Poulisse and Bongaerts, 1994; Hermans, Bongaerts, De Bot and Schreuder, 1998; Grosjean, 1998, but see Costa and Caramazza, 1999, for a different proposal). Consequently, it is important to find out how L2 speakers control their production, i.e., how it is achieved that only one language at a time is spoken while both are activated.

Most theories assume that the language systems can be at different levels of activation and that in order to speak one language rather than another, the activation of the target language must exceed that of the other language or languages (Paradis, 1981; Grosjean, 1988). Furthermore, it is assumed that the regulation of which language shall be spoken is achieved by the modification of the levels of activation in the language networks, rather than via a simple switch mechanism (Paradis, 1981; Grosjean, 1988; Grainger and Dijkstra, 1992; De Bot and Schreuder, 1993; Li, 1996).

The factors which influence the degree of activation of the language systems are numerous. Grosjean (1997, 1998) assumes that there are different language modes depending on how much the L1 system is activated relative to the L2 system. In the bilingual mode, both languages are active but the target language is more strongly activated. Speakers are in this mode when, for example, they speak with somebody with whom they can code-switch or mix languages. During such conversations, words or phrases from both languages are produced, though one language usually remains the base language. On the other hand, in the monolingual mode of Grosjean's continuum, one language, the base language, is active, whereas the other one is deactivated at least partially. As for L2 speakers, it is assumed that they cannot prevent their L1 from being at least partially activated and thus exerting an influence on their L2 production. The degree of their L1 activation may thus have an effect on the strength of the L1 transfer or interference.

In our research we addressed two questions concerning this topic: (1) Can L2 speakers effectively increase the L2 activation and suppress the L1 activation, if they know long enough in advance the language in which they will respond compared to a situation where the response language is indicated only very shortly in advance (Experiment 1)? (2) Can the L1 influence be reduced or even eliminated when L2 speakers are close to

* The research reported in this paper was supported by a grant from the Deutsche Forschungsgemeinschaft (German Research Council, DFG Pe/357-9). We thank Jana Hrachovcová, Nataša Ukropcová, Andreas Opitz and Michael Welt for their help with preparing and running the experiments as well as Dirk Janssen for his help with statistical analyses and Peter Ecke for his comments on an earlier draft of this manuscript.

Address for correspondence:
Denisa Bordag, University of Leipzig, Department of Linguistics, Beethovenstr. 15, D–04107 Leipzig, Germany
E-mail: *denisav@rz.uni-leipzig.de*

the monolingual end of Grosjean's continuum using only the L2 (Experiment 2)?[1]

Experimental evidence has been collected demonstrating that the L1 and L2 systems interact at the conceptual and phonological levels (De Groot, 1992; De Groot, Dannenburg and van Hell, 1994; Singleton, 1999; Dijkstra, Grainger and van Heuven, 1999; Costa, Caramazza and Sebastián-Gallés, 2000) and this issue is no longer controversial. The authors of these studies often assume parallel activation and interaction also at the level where grammatical features are stored, e.g., at the lemma level in Levelt's model (Levelt, 1989; Levelt, Roelofs and Meyer, 1999), but here the evidence is much less unequivocal. Basically, the only feature explored in this context thus far has been grammatical gender and the results are often contradictory.

Costa, Kovačić, Franck and Caramazza (2003) report several experiments in which they failed to obtain evidence for the interaction between L1 and L2 in grammatical gender processing. Subjects produced gender-marked noun phrases in their L2 which were either gender-congruent or gender-incongruent with their L1 translation equivalents. Naming latencies were statistically identical for both sets of items. In contrast, Bordag (2004, 2006) reports similarly designed experiments in which she observed the so-called gender interference effect from L1: subjects were slower when naming pictures with nouns and noun phrases when the L2 gender was incongruent with that of the corresponding L1 translation. When the same pictures were named by an L1 control group of subjects, no difference between the two sets was observed.

The question of whether the L1 gender affects the L2 gender processing is interesting for several reasons. First, as explained above, it can help to clarify whether the L1 and L2 systems interact at the level of grammatical encoding and whether this interaction can be modulated by, for example, subjects' position on the bilingual continuum. Second, it can help to specify how grammatical gender is represented and accessed. According to one class of models – as pointed out by Costa, Kovačić, Franck and Caramazza (2003) – the speed and efficiency with which the gender feature is retrieved depends to some extent on its activation level (e.g., Schriefers, 1993; Levelt, 2001; Vigliocco, Lauer, Damian and Levelt, 2002; Bordag, 2004). If a gender node is strongly activated and its competitors receive only very little activation, then it can be selected faster and more easily than when its activation is relatively low and/or there are other strongly activated competitors. In such a case, the competition for selection is more difficult and thus slower.

Alternatively, the automatic gender-access models (Caramazza et al., 2001; Schiller and Caramazza, 2002; Costa, Kovačić, Fedorenko and Caramazza, 2003) propose that the gender of a given noun becomes automatically available for further processing upon the selection of the noun's lexical node. In such models the gender access is a direct and automatic consequence of lexical selection and thus the activation levels of the target node and its competitors do not play any role.

According to the activation-dependent models (cf. Costa, Kovačić, Franck and Caramazza, 2003), L2 gender production should be faster and more accurate if the L1 and L2 translation equivalents have congruent genders than when their genders differ. When both the L1 and L2 nouns have the same gender, e.g., they are both feminine like *Kerze* (f, "candle") and *svíčka* (f, "candle") in German and Czech, respectively, the feminine gender node receives activation from two sources. On the other hand, if the L1 and L2 translations have different genders, e.g., *Burg* (f, "castle") and *hrad* (m, "castle"), then the feminine gender node receives activation from the L2 system and the masculine gender node from the L1 system. The target L2 gender thus has a highly activated competitor whose activation it must override in order to be selected. If it happens that the L1 masculine gender is more activated, then it is selected and the speaker produces a gender error.

In the framework of the automatic gender-access models, the retrieval of the gender feature does not depend on activation levels and therefore no systematic differences in naming latencies between words with a different and the same gender in L1 and L2 should be observed. More generally, the gender retrieval in such models is not affected by any possible influences which modify the activation levels. Such models would not only predict that there is no difference in gender production of gender-congruent and gender-incongruent nouns, but also that other factors, e.g., transparency/opacity of the L2 word's phonological form should not affect gender retrieval either. According to the activation-independent models, there would be no difference in the retrieval of e.g., the L2 word *Auge* (n, "eye") which has a termination atypical for its gender (*-e* is a typical German termination for feminine nouns) and the word *Kerze* (f, "candle") which has a gender termination typical for its gender. On the other hand, an activation-dependent model would predict faster reaction times for words with a typical termination than for words with an atypical termination

[1] Though there are studies supporting the account that the effects of the non-response language cannot be eliminated even in a completely monolingual setting, they have never focused on the level of grammatical encoding. As an example, Costa, Caramazza and Sebastián-Gallés (2000) showed that the L1 influence at the level of phonological encoding is not modified by subjects' position on the Grosjean's continuum. However, corresponding research concerning the level where grammatical features are stored, has not been reported so far. As there is no reason to assume a priori that the interaction between L1 and L2 semantics, grammar, phonology, and orthography is affected by the same factors and equally strongly (see also Kroll and Tokowicz, 2005), it is necessary to conduct similar research at all processing levels.

(under the assumption that there is feedback between the levels of phonological and grammatical encoding). If an atypically gender-marked neuter noun like *Auge* is to be produced, its phonological form, which has a gender marking typical for feminine nouns, will activate the feminine gender node. The neuter node thus has a strongly activated competitor whose activation level it must exceed in order to be selected. If the incorrect feminine node is more activated, then it is selected and the speaker incorrectly produces a feminine gender which is in accordance with the phonological form of the target word. If the noun termination is in accordance with its gender, i.e., gender typical as it is the case of the word *Kerze*, then the phonological form activates the feminine gender as well, thus allowing the feminine gender node to easily win the competition for selection. In our Experiments 3 and 4 we investigate the influence of phonological form on L2 gender processing.

The potential influence of phonological form (e.g., noun's termination) has also further important consequences for models of speech production. Most recent theories of speech production agree that lexicalization comprises two major subcomponents: the selection of grammatical features of a particular lexical item that adequately represents the intended concept and the item's phonological specification. However, the models differ both in how these two subcomponents interact (if at all), and in how the grammatical features are represented, whether as "a grammatical bundle" stored at the so-called lemma, as in the model of Levelt (1989), or independently from individual lexical items in a special syntactic network, as in the Independent Network Model of Caramazza (1997).

Serial, modular models of speech production, e.g., the Levelt model, propose that grammatical and phonological encoding proceed in two separate steps: first, one and only one lemma (with its grammatical features) is selected from multiple activated candidates, and second, this lemma alone is phonologically encoded. Phonological forms (e.g., word endings) have no influence on the selection of grammatical features (e.g., grammatical gender).

On the other hand, according to interactive models like the Interactive Activation Model of Dell (1986), phonological information can feed back to influence the processes on the higher level of grammatical encoding. In this case, the availability of a given word's phonological form can bias the probability that a particular grammatical feature (e.g., a masculine gender) will be selected instead of some alternative (e.g., a feminine or a neuter gender).

The interaction between the phonological and grammatical features is also inherent in the Independent Network Model of Caramazza. This model dispenses of the lemma notion and differentiates between three independent networks: the lexical-semantic, syntactic, and the network of phonological forms. During word production, first its lexical-semantic representation is selected. The activation then spreads to both the syntactic and the phonological network simultaneously. Caramazza differentiates between syntactic features which can be activated directly from the lexical-semantic network (e.g., word class or verb tense), and features like grammatical gender which cannot receive any activation from the semantic network. Under normal conditions, the activation coming from the lexical-semantic network is not sufficient to activate a syntactic feature to such a degree that it can be selected, so that additional activation from the network of phonological lexemes is necessary. The selection of grammatical features like gender cannot be achieved without the contribution of a word's phonological form. Thus, if we find evidence for the gender-transparency effect in our Experiments 3 and 4, it would speak not only in favor of activation-dependent processing, but also in favor of models postulating interaction between grammatical and phonological encoding.

The languages used in this study were Czech (L1) and German (L2). Both of these languages have the same number and type of genders: masculine, feminine, and neuter. Despite these similarities, the languages differ in how the three genders are assigned to their nouns. Approximately half of their translation equivalents have identical genders while the other half do not. Sometimes the gender can be derived from nouns' terminations but this is true only for transparently gender-marked nouns (see below).

Subjects in our experiments were late unbalanced bilinguals with intermediate and upper-intermediate knowledge of German. We decided to test this population rather than highly proficient bilinguals because previous research has shown that the magnitude of interlingual effects depends on subjects' proficiency. Ellis (2005) determined the actual size of cognate effects in several studies with an orthogonal design. Its magnitude varied between 15% and 19% when highly experienced bilinguals were the participants and was substantially larger when less-experienced bilinguals served as participants (25% in receptive tests and about 50% in productive tests). Moreover, Costa, Kovačić, Franck and Caramazza (2003) found no gender interference effect from L1 when testing highly proficient bilinguals in several languages and stipulated that it might be due to the very high proficiency of his subjects. Based on these findings, bilinguals of intermediate proficiency seemed to be a population in which the L1 gender interference is most likely to appear and who are able to deal with the given task (contrary to e.g., novice learners).

The first two picture-naming experiments of this study were designed to explore the factors affecting the L1 influence on L2 gender production. In Experiment 1, speakers of L1 Czech and L2 German randomly produced either bare nouns or gender-marked adjectives + nouns.

L1 fillers were also included to create optimal conditions for interlingual interference. Depending on the color of the background, which turned yellow or blue 1200 ms before pictureonset,[2] subjects had to respond either in their L1 or L2. We expected that reaction times should be significantly slower and the gender error rates significantly higher when subjects named pictures from the gender-incongruent condition, i.e. those whose names had different gender in L1 and L2, than when they named pictures from the gender-congruent condition.

In Experiment 2 we modified the position of subjects on the language continuum by excluding the L1 fillers and thus moving subjects closer to its L2 monolingual end. We assumed that this change could lead to a reduction or even elimination of the L1 interference. In both experiments we used the same materials, design and procedure as Bordag (2004) in order to make comparisons possible.

Experiment 1

In this experiment, subjects had to name pictures in their L2 German in two conditions: either with a bare noun (short condition) or with a size adjective ("*big*" or "*small*") and a noun (long condition). In both languages, adjectives have to agree in gender with their head nouns. Half of the nouns chosen for the experiment had a congruent gender in both languages, the other half did not. The aim was to investigate whether the production of nouns and noun phrases with a congruent gender in both languages would be still faster than the production of those with a different gender in German and Czech, when the sign determining the response language (color of the background) will be given long enough in advance. To create optimal conditions in which the L1 is most likely (and most strongly) activated, L1 fillers were included to keep subjects as close to the bilingual end of the bilingual continuum (Grosjean, 1997) as possible. Thus, one third (22) of the pictures was to be named in Czech and two thirds (44) in German.

Method

Participants

Eighteen subjects participated in this experiment. They were drawn from the same population as the subjects from the Bordag (2004) experiments (however, none of them participated in more than one experiment of this type). They were mostly exchange students studying for one or two semesters at the University of Leipzig with comparable German knowledge to that of the subjects in the former experiments (upper intermediate to advanced). They started to learn German after their 8th year (at age 13;2 on average) and had been learning German for 8.7 years on average. Many of them, however, experienced long periods when they were not using German at all. The proficiency of the subjects was judged on the basis of an informal interview in subjects' L2 at the beginning of the experiment, the information they provided in a questionnaire designed by Hermans et al. (1998), their rating of their familiarity with the experimental items and their results in the previous language studies. All participants had sufficient knowledge of German to study using this language at the University of Leipzig. Highly advanced subjects or subjects below the intermediate level were not tested or not included in the analysis. All subjects were paid for their participation.

Materials

The target stimuli were 70 line-drawings from a database collected by the Max Plank Instituts in Nijmegen and in Leipzig. They have been regularly and successfully used in previous experiments. Pictures of 44 objects were chosen as experimental items, none of them were cognates. The gender of half of the nouns by which the objects were named was congruent with that of their translation equivalents (e.g., *strom* and *Baum*, both masculine), the gender of the other half differed from that of their translation equivalents (e.g., *obraz* (masculine) and *Bild* (neuter)). They were all easily-depictable concrete nouns with high frequency in both Czech and German. The majority of the experimental items were monomorphemic and inanimate. Each set consisted of 22 nouns: 8 masculine, 8 feminine, and 6 neuter. Their length ranged between one and three syllables and each set contained the same number of mono-, di- and trisyllabic masculine, feminine, and neuter nouns. The sets were also matched for the final phoneme to control for the possible influence of phonological form on gender selection. Pictures in each set were controlled for complexity and easiness of recognition (see Appendix A for the example pictures and Appendix B for the list of target stimuli).

Next to 44 experimental items, there were also 22 L1 filler items and 14 practice items. They were also very frequent in both languages. Some of the fillers were used as practice items as well.

Procedure

The experimental session had three parts and lasted approximately 40 minutes. Participants were tested individually. The stimuli appeared on a computer screen and participants' reaction times were recorded by a computer using the experimental software ERTS (Beringer, 1999).

The first part started with a short informal dialogue in the L2 to test whether the subject was a suitable candidate for the experiment. Subjects were then presented with instructions emphasizing both the speed and accuracy

[2] In Bordag (2004) the time lag between the language cue and the picture onset was only 300 ms. In Experiment 1 we wanted to test whether this variable affects the interference effect (see below).

of their responses and with a booklet containing all 70 pictures which would appear in the experiment later. Under each picture, its L1 name and its L2 translation equivalent were printed. Subjects were asked to rate on a seven-point scale (1) how well the picture depicted the L1 word, (2) how well it depicted the L2 word, and (3) how familiar they were with the L2 word. The ratings were done to ensure that participants really studied the names of the pictures and to obtain data for evaluations of the stimulus materials and of subjects' familiarity with the L2 words. They also allowed the experimenter to control for whether the words in the gender-congruent or gender-incongruent set were equally familiar and equally well depicted than the words in the other set. Later analyses showed this to be the case. (Total rating was 6.56 in the congruent and 6.59 in the incongruent group.)

In the second part of the experimental session the experiment proper was run. Before the experiment started, participants were presented with a sheet containing German and Czech examples of noun phrases which were to be used in the experiment's long condition. Attention was not explicitly drawn to grammatical gender.

The experimental session began with a block of 14 practice items. Each trial started with a visual fixation point (= ∗ =) presented on a grey background for 600 ms on the right half of the screen in the middle of an imaginary square in which the target picture was later displayed. Then, either a yellow or a blue background, indicating which language subjects were to use in their response, appeared for 1200 ms. In this respect the experiment differed from those reported in Bordag (2004, 2006). In her experiments the period was four times shorter (300 ms). This was a crucial difference because it gave subjects more time to raise the activation level of the target language – if possible. Half of the subjects in each experiment had to respond in German when a picture was displayed on a blue background and in Czech when the background was yellow. The other half of the subjects had the colors reversed. Following this pause, a target picture was presented in the imaginary square on the right side of the computer screen and marked with a black arrow pointing at it from the upper right part of the screen. In the long condition, two pictures of the same object appeared simultaneously on the screen: The target picture marked with an arrow in its usual position on the right part of the screen and either a larger or a smaller control picture of the same object on the left. When the left control picture was larger than the target picture, the subjects had to name the target picture with the adjective "small"; if the left control picture was smaller than the target picture, subjects had to name the target picture with the adjective "big". These adjectives are clearly gender-marked in nominative singular in German (e.g., *groß**e*** for feminine, *groß**er*** for masculine, *groß**es*** for neuter). There was, therefore, agreement between the head noun and the modifying adjective in the long condition. We included this condition, because previous research on gender production has revealed that some gender effects (e.g., the gender-congruency effect) emerge only when gender is required by the syntactic context, e.g., in a form of a gender-marked definite article (La Heij, Mak, Sander and Willeboordse, 1998). Naming latencies were measured from the onset of participant's utterance by a voice key.

The display color of the arrow was contingent on the participant's response. It turned white as soon as a vocal response was initiated. However, if no response was registered within 5000 ms (time out) the arrow turned red. When the participant was ready for the next trial, he or she pressed the space bar on the keyboard and after a pause of 300 ms the fixation point of the new trial was displayed.

Subjects were randomly assigned to one of two main randomization versions. Each item was presented in a long and in a short condition. The experiment had two parts. In each part subjects saw items in random order in the long and in the short condition. The items, which appeared in the long condition in the first part, appeared in the short condition in the second part and vice versa. The two main versions of the experiment differed with respect to which items were in the long condition first.

Individual randomizations within the two main versions were prepared for each subject with the following restrictions. No more than three trials of the same condition (long vs. short), the same size (big vs. small), and the same gender (masculine, feminine, neuter) were allowed to follow each other. The same language of response was allowed in maximally four successive trials.

In the third part of the experimental session, subjects filled out a questionnaire concerning their proficiency in the L2 and their language background. The same questionnaire as in Hermans et al. (1998) was used.

Results and discussion

A total of 389 (16.4%) responses were marked as incorrect. Observations were discarded due to voice-key errors, hesitations or stuttering, task errors (e.g., picture named in a wrong language), and errors specific for the long condition (gender errors, hesitations in the middle of the adjective or between the adjective and the noun and producing a wrong adjective, e.g., *gross* instead of *klein*). The cut-off was two standard deviations from a subject's mean response time in the long and short condition.

Analyses of variance (ANOVAs) were performed on the mean naming latencies per subject (F1) and per item (F2). The experiment had a 2×2 design, all factors were within-subjects. The factor 'length' was within-items, the factor 'congruency' between-items. Overall, naming latencies of pictures with names from the gender-incongruent set were slower than naming latencies of pictures with a congruent gender in both languages.

Table 1. *Results of Experiment 1 with reaction times (RT) in milliseconds, standard deviations (SD), and error rates with percentages counted for each experimental condition.*

	Congruent			Incongruent			
Condition	RT	SD	Errors	RT	SD	Errors	Interference effect
adj + noun	1490	555	87 (22%)	1622	578	123 (31%)	+132
noun	1163	333	34 (4.5%)	1265	339	18 (8.6%)	+102

Both the analysis with subjects as a random factor and with items as a random factor showed a significant effect of 'congruency' (congruent vs. incongruent): *F1* (1,17) = 16.37, $p < .01$ and *F2* (1,43) = 42.54, $p < .01$. The effect of 'length' (short vs. long) was significant as well (*F1* (1, 17) = 29.13, $p < .01$ and *F2* (1, 47) = 117.36, $p < .01$). There was no interaction between condition and length which means that as in the Bordag (2004, 2006) experiments, the gender-interference effect was obtained not only when pictures were named with adjectives, but also with bare noun naming (see Table 1 for a summary of the results). We postpone the discussion of this issue until after the presentation of Experiment 2.

Because we used the same design, procedure, and materials as Bordag (2004, 2006), we could compare our Experiment 1 with the previous experiments. We can conclude that the present results parallel those of Bordag (2004, 2006), despite the fact that we modulated the time gap between the presentation of the language cue and the picture onset. Contrary to our expectations, the earlier information about the response language did not facilitate picture naming: subjects were even slower than in the original Bordag (2004) experiments. Even more importantly, the gender-interference effect from L1 did not decrease under the new conditions, but was even larger, at least numerically. This suggests that knowing long in advance the response language does not enable subjects to rise its activation (or suppress the activation of their L1) effectively and thus to reduce the L1 interference at the level of grammatical encoding. The only deviation from the expected direction was in the number of gender errors. Whereas in the original Bordag (2004) experiment subjects produced significantly more gender errors in the incongruent condition, in this experiment the number of gender errors (47 in the gender-congruent and 65 in the gender-incongruent condition) failed to reach significance: $\chi^2 = 2.893$, $df = 1$, $p = .089$.

Experiment 2

In Experiment 1 as well as in the original Bordag (2004) experiments L1 fillers were included. This mixed-language design should increase the probability of detecting any influence of the L1 on the L2. In Experiment 2 we manipulated subjects' position at the bilingual continuum. The previous research on cognates has shown that even when close to the monolingual end, subjects cannot completely eliminate the cognate facilitation effect, which is assumed to originate at the level of phonological encoding. In Experiment 2 we wanted to test whether this holds also for grammatical effects like L1 gender-interference, or whether cross-language grammatical effects can be diminished if subjects are in the monolingual mode or very close to it. The crucial difference between this and the previous experiment thus was that in Experiment 2 all items had to be named in German. The experimental conditions were also monolingual: subjects were tested by a native German speaker, all instructions and conversations were also in German. Grosjean (1997) has shown that these factors affect subjects' position on the bilingual continuum.

Method

Participants

Again the same population of subjects was tested. None of them participated in any of the previous experiments on this topic. They started to learn German after their 9th year (at 12;4 on average) and had been learning German for 9;5 years on average. Their knowledge of German ranged from intermediate to advanced (but not highly proficient). They were all paid for their participation.

Materials

The same materials as in Experiment 1 were used. This time all items, including the fillers, were named in L2.

Procedure

The procedure of the experiment was similar to that of Experiment 1. The main difference was that this time all pictures had to be named in German. Thus there was no change of colors indicating the response language. Instead, the background stayed grey during the whole experiment. The target picture appeared 300 ms after the fixation point.

Results and discussion

A total of 274 (11.5%) responses were marked as incorrect. Observations were discarded mainly due to voice-key errors, gender errors, and task errors (picture named

Table 2. *Results of Experiment 2 with reaction times (RT) in milliseconds, standard deviations (SD), and error rates with percentages counted for each experimental condition.*

	Congruent			Incongruent			
Condition	RT	SD	Errors	RT	SD	Errors	Interference effect
adj + noun	1035	226	73 (18.4%)	1102	234	79 (20.0%)	+67
noun	865	183	15 (3.8%)	935	213	14 (3.5%)	+70

with a wrong adjective etc.). The cut-off was the same as in Experiment 1. Table 2 gives a summary of the results.

The experiment had the same design as Experiment 1. The ANOVAs again revealed a significant effect of the factor 'length': *F1* $(1,17) = 96.06$, $p < .01$ and *F2* $(1,47) = 103.69$, $p < .01$. Though numerically the gender-interference effect was smaller than in Experiment 1, the statistical analyses showed that it was highly significant both in F1 and F2: *F1* $(1,17) = 43.39$, $p < .01$ and *F2* $(1,47) = 9.14$, $p < .01$. The interaction was not significant: *F1* $(1,17) = 0.17$, $p = .69$ and *F2* $(1,47) = .18$, $p = .67$. The χ^2 test on the gender errors revealed that there was no significant difference between the 37 and 42 errors subjects made in the gender-congruent and gender-incongruent condition, respectively.

Thus, Experiment 2 provided evidence that the L1 gender also interferes with the L2 gender processing when subjects are in a monolingual mode (or at least much closer to it compared to the previous experiments). Obviously, subjects' position on the bilingual continuum is not a factor that can lead to the elimination of cross-language effects at the level of grammatical encoding.

The joint analyses of Experiment 1 and 2 ($2 \times 2 \times 2$ design with an additional between-subjects factor 'experiment') confirmed that the gender-interference effect was highly significant ($p < .001$). The factor 'length' was significant as expected, too. Crucially, the factor 'experiment' (Experiment 1 vs. 2) was also significant showing that subjects were significantly faster in Experiment 2 than in Experiment 1. The only significant interaction was the interaction between the factors 'experiment' and 'length': The difference between the long and short condition was larger in Experiment 1 than in Experiment 2. The interaction between 'experiment' and 'congruency' was only marginally significant in F2 and not significant in F1. This could be a slight hint that the interference effect was at least partially reduced compared to Experiment 1; the present data however do not allow for a strong conclusion in this respect.

In both Experiments 1 and 2 the gender-congruency effect was observed in both the short and the long condition. The consistent replication of this result with different subjects, modified designs, and different items (see also Experiment 3 in this study) makes it very unlikely that the effect in the short condition would have its origin in the particular material, subjects, or method used. Moreover, Bordag (2004) reports a control experiment with L1 German speakers, who named the same pictures as used in our Experiments 1 and 2 in their L1 and no differences between the two groups of pictures were observed. The results of the ratings also speak against this assumption.

The gender-congruency effect in the short condition contrasts with results of picture-word distractor experiments in L1s (Schriefers and Jescheniak, 1999). In these experiments, subjects name pictures in the presence of word distractors, which have either a congruent, or an incongruent gender with the name of the picture. One of the interpretations of this effect holds that subjects are slower in the gender-incongruent condition because the gender node of the picture competes with the gender node of the distractor and the resolution of this conflict manifests itself in longer naming latencies compared to the situation in which the picture and the distractor have congruent genders. This gender-congruency effect in L1s, which is in many aspects similar to the cross-language gender-congruency effect described in this study, was however observed only when gender was required by a noun's local syntactic environment (i.e. when overtly needed for production) and never when only bare nouns were named. Bordag (2004) offers several proposals for why the cross-language congruency effect appears in both short and long conditions. The most likely explanation seems to be the following.

In the monolingual picture-word distractor tasks, mental representations of the picture and the distractor have nothing in common: They merely appear together in the experimental conditions. On the other hand, translation equivalents are semantically closely related. They are assumed to share the same concepts or semantic features, at least those of concrete nouns (see the Revised Hierarchical Model by Kroll and Stewart, 1994; and the Distributed Feature Model by De Groot, 1992; and De Groot et al., 1994). One interpretation of the difference between the two effects thus holds that when there is a strong relationship between the target noun and its interfering counterpart, e.g., when they share the same concept, their grammatical features always compete for selection. When the lemmas do not share the same concept, the gender representations compete for selection only when a gender feature is necessary for production. Indirect evidence supporting this interpretation comes from a study by Schriefers and Teruel (2000). In their experiments with a picture-word distraction paradigm they also used semantically related distractors in both gender-congruent and gender-incongruent conditions and observed an interaction between semantic relatedness and gender congruency. The authors claim that "the competition between different gender nodes is strengthened by the semantic relatedness of target and distractor lemma" (Schriefers and Teruel, 2000, p. 1374). More research is, however, necessary to fully clarify this issue.

Experiment 3

In this experiment we wanted to investigate both the influence of the L1 grammatical gender and of the phonological form of the L2 word on the L2 production. Though the gender cues in German are not always reliable, the following regularities can be observed with monomorphemic nouns (cf. Köpcke and Zubin, 1983; Mills, 1986): Feminine nouns end typically with a schwa *-e* (e.g., *die Kerze* "the candle"), masculine and neuter nouns with a consonant (e.g., *der Baum* "the tree" or *das Buch* "the book"). A small number of monomorphemic feminine nouns can also end with a consonant (e.g., *die Burg* "the castle") and vice versa, a small number of masculine and neuter nouns end with an *-e* (e.g., *der Käse* "the cheese" or *das Ende* "the end"). From this point of view, there is one typical gender termination, i.e., *-e* for feminine nouns, one ambiguous gender termination, i.e., a consonant for masculine and neuter nouns, and two atypical terminations, i.e., a consonant for feminine nouns and an *-e* for masculine and neuter nouns.

In similarly designed experiments Bordag, Opitz and Pechmann (2006) observed that English learners of German were slower and made more errors when they produced gender-marked nominal phrases containing a noun with a gender-atypical or ambiguous termination than when they produced similar phrases containing nouns with a termination typical for their gender. As there is no grammatical gender in English, it was assumed that the influence of L1 gender played no significant role in those experiments. On the other hand, in the present experiment with Czech learners of German we take both factors, gender congruency and gender transparency, into account.

Method

Participants

Eighteen subjects took part in the experiment. All of them were Czech native speakers from the same population as the subjects in the previous experiments. However, they were more fluent than the subjects in Experiments 1 and 2, because they were tested at the end of their exchange program, i.e. after at least 3 months of intensive exposure to German. They started to learn German after their 6th year (at 10;7 on average) and had been learning German for 13 years on average, but they often experienced long periods when they were either using German very rarely or not at all. Their knowledge of German ranged from upper intermediate to advanced. They were all paid for their participation. None of them participated in the previous experiments.

The same participants took also part in Experiment 4, reported below.

Materials

The set of experimental items consisted of three groups: Group A contained 16 nouns with a gender-typical termination, Group B contained 16 nouns with a gender-ambiguous termination, and Group C contained 16 nouns with a gender-atypical termination.[3] The gender-typical Group A consisted of German feminine nouns with a typically feminine termination *-e*, the gender-ambiguous Group B consisted of masculine and neuter nouns ending with a consonant, and the atypical Group C contained feminine nouns ending with a consonant and masculine and neuter nouns ending with an *-e*. Half of the nouns in each gender-typical, gender-ambiguous, and gender-atypical group had the same gender as their translation equivalents (gender-congruent condition), the other half had a different gender than their L1 translation (gender-incongruent condition). The words in the three groups were carefully selected on the basis of a rating study and were matched for frequency, length, familiarity, and degree of formal similarity between L1 and L2.[4]

Procedure

Participants were tested individually. At the beginning of each session they reviewed a booklet containing all 58 pictures that were to appear in the experiment later. Each picture was labeled with the German noun it depicted. Subjects were asked to rate on a seven-point scale how well each picture depicted the given noun and how familiar they were with each German noun. The aim of the ratings was

[3] The set of materials used in this experiment was very similar to the set of items used by Bordag, Opitz and Pechmann (2006). Only a few items had to be changed to balance the compared sets of items with regard to a new condition, namely the gender congruency.

[4] In the L1 version of the experiment (German native speakers naming the pictures in German), which could be taken as a control experiment, no difference between the naming latencies of the three groups was observed (see also Bordag et al., 2006).

Table 3. *Results of Experiment 3 regarding 'gender congruency' with reaction times (RT) in milliseconds, standard deviations (SD), and error rates with percentages counted for each experimental condition.*

	Congruent			Incongruent			
Condition	RT	SD	Errors	RT	SD	Errors	Interference effect
adj + noun	909	224	63 (14.6%)	944	217	51 (11.8%)	+ 35
noun	719	165	19 (4.4%)	749	153	29 (6.7%)	+ 30

to draw subjects' attention to the pictures and to check whether the items in the three Groups A, B, C and in the gender-congruent and gender-incongruent group were equally well depicted and well known. The subsequent analyses showed that this was the case.

Next, participants read the instructions for the subsequent part of the experiment. Their attention was especially drawn to the importance of both speed and accuracy.

There were again two naming conditions in the experiment. In the short condition, pictures were named with bare nouns (e.g., *Haus*), in the long condition with a complex noun phrase in the form of adjective + noun (e.g., *großes Haus*).

Two versions of the experiment were prepared. In the first version, all nouns had to be named first in the short condition (bare noun) and after that in the long condition (adjective and noun). In the second version the two conditions were presented in a reverse order. Nine subjects were tested in the first version, nine were tested in the second version.

The adjectives used in this experiment were *groß* "big" and *klein* "small". Half of the items had to be named with *groß*, the other half with *klein*. This equal distribution was also maintained within the three groups of critical items (typical, ambiguous, atypical).

The order of the items in each condition was pseudo-randomized for each subject with the following restrictions: no more than two items from the same group and no more than two items of the same gender were allowed to follow each other.

The experiment started with a block of 10 practice items. Subjects were corrected, if their response was incorrect, however gender errors were not corrected. The practice block was followed by a short break after which the first part of the experiment with critical items started, followed by another break. There was again a block with practice items (now in the other condition) followed by a break before the second part of the experiment started. Each trial began with a 600 ms fixation sign (=∗=) at the center of the right half of the screen, where the target picture was about to appear. A pause of 300 ms followed after the fixation sign had disappeared.

In the short condition (bare noun naming), a picture then appeared at the right half of the screen marked by a black arrow. The subject named the single picture with a bare noun. In the long condition, two versions of the same picture appeared. The size of the picture at the right half of the screen was identical with that of the picture in the short condition and it was again marked with an arrow. The second picture at the left part of the screen was bigger or smaller compared with the picture on the right. The subjects had to name the right (arrow-marked) picture with an adjective *groß-* "big" or *klein-* "small" depending on the size of the marked target picture compared with the control picture on the left.

The onset of the utterance was recorded by a voice-key. The black arrow turned white when the response time was measured and the subject had to request the next trial by pressing space bar. If there was no signal recorded within a 2000 ms period after the appearance of the stimuli (no utterance of the subject), the arrow turned red. This was a "time out" notification for the participant and experimenter.

The experimenter coded whether the subject's response was correct and recorded the complete form of an incorrect utterance. One experimental session took about 30 minutes.

Results and discussion

The data were screened for errors and outliers. Trials in which the voice key malfunctioned, the subject hesitated, stuttered or gave incorrect responses and those trials in which naming latencies were more than three standard deviations away from the mean of the participant and condition (short vs. long) were removed from the data. Naming latencies were submitted to two separate ANOVAs, one with the factor gender 'congruency' and one with the factor gender 'transparency'. Table 3 gives a summary of the results.

In the first ANOVA with two factors 'congruency' (two levels, between-items) and 'length' (two levels, within-items) , both the factor 'length' and 'congruency' were significant: 'length' $F1\ (1,17) = 40.54$, $p < .01$ and $F2\ (1,47) = 198.78, p < .01$; 'congruency' $F1\ (1,17) = 29.44$,

Table 4. *Results of Experiment 3 regarding 'gender transparency' with reaction times (RT) in milliseconds, standard deviations (SD), and error rates with percentages counted for each experimental condition.*

	A (typical)			B (ambiguous)			C (atypical)		
Condition	RT	SD	Errors	RT	SD	Errors	RT	SD	Errors
adj + noun	914	211	15 (5.2%)	919	224	59 (20.5%)	948	230	40 (13.9%)
noun	722	146	19 (6.6%)	726	161	10 (3.5%)	753	270	19 (6.6%)

$p < .01$ and *F2* (1,47) = 7.21, $p < .05$. The interaction between these two factors was not significant: *F1* (1,17) = 0.39, $p = 0.54$ and *F2* (1,47) = .01, $p = .96$. This result shows that we again observed a significant gender-congruency effect in both the short and the long condition. Though significant, the effect was numerically smaller than in the previous experiments (about 30 ms). Together with the fact that the overall naming latencies in this experiment were much shorter than in the previous experiments, we attribute this difference to the higher proficiency level of the subjects in this experiment. As noted above, contrary to the subjects in the previous experiments they were tested towards the end of their one semester stay at the University of Leipzig, i.e. after a period of intensive exposure to German. Moreover, a closer look at our data on these subjects revealed that before their participation in this experiment they all had participated in at least one picture naming experiment (but with different materials and on a different topic). Thus, they were more proficient not only as far as their German knowledge was concerned, but also in their experience with psycholinguistic experiments.

The analyses of errors showed no significant differences between the number of errors in the gender-congruent and gender-incongruent conditions: 'congruency' *F1* (1,17) = .06, $p = 0.82$ and *F2* (1,46) = .02, $p = .89$. The factor 'length' was again highly significant: *F1* (1,17) = 21.21, $p < .01$; *F2* (1, 46) = 15.21, $p < .01$. The interaction between these two factors failed to reach significance: *F1* (1,17) = 3.65, $p = .07$; *F2* (1,46) = 2.29, $p = .14$. Table 4 gives a summary of the results.

In the second ANOVA with the factors 'gender transparency' (three levels, between-items) and 'length' (two levels, within-items) , the factor 'length' was again significant in both F1 and F2: *F1*(1,17) = 40.98, $p < .01$; *F2* (1,30) = 311.52, $p < .01$.Crucially, the factor 'gender transparency' was significant in F1: *F1* (2,35) = 10.45, $p < .01$ and though it failed to reach significance in *F2* (2,61) = 2.24, $p = .12$, the post-hoc Scheffé test ($p < .05$) was significant in both F1 and F2: Group C differed significantly from both Groups A and B (which were statistically identical). The interaction between 'group' and 'length' was not significant: *F1* (2,35) = .14, $p = .87$; *F2* (2,61) = 0.01, $p = .99$, showing that the gender-transparency effect emerged both in the long and in the short condition.

The analyses of error rates revealed also a significant effect on factor 'gender transparency' (*F1* (2,34) = 9.32, $p < .01$ and *F2* (2,60) = 4.38, $p < .05$). The Scheffé test ($p < .05$) for both F1 and F2 showed that Group A differed significantly from both Groups B and C which were statistically identical. This corresponds also to the number of gender errors alone (1 gender error in Group A, 38 in Group B, and 21 in Group C). The factor 'length' was significant, too: *F1* (1,17) = 21.21, $p < .01$ and *F2* (1,30) = 34.10, $p < .01$ showing that subjects made more errors in the long than in the short condition. The increase of errors in the long condition was carried by the gender errors in the ambiguous and atypical group (indeed subjects could not make any gender errors in bare noun naming). The interaction between the factors 'gender transparency' and 'length' was also significant (*F1* (2,34) = 14.664, $p < .01$ and *F2* (2,60) = 9.48, $p < .01$), an expectable result given that the increase of errors concerned only groups B and C in the long condition. Compared to the analyses of reaction times, where only the atypical group C differed from the two other groups, the analyses of errors suggest processing difficulties also in gender production of NPs with a noun with an ambiguous termination.

The results of this experiment yielded at least two important findings. First, the gender-interference effect from L1 was replicated with different materials and under different conditions. Similarly to Experiment 2, subjects were tested in monolingual settings using their L2 only. In contrast to both Experiments 1 and 2, subjects named the pictures in the long and short conditions in separate blocks: half of the subjects named all items with a bare noun first and then with the adjective + noun phrase, the other half of the subjects named the pictures in the reversed order. Thus it could not be argued that the gender-interference effect in the short condition is only a transfer effect from the long condition. Obviously, the L1 grammatical gender also interferes with the L2 production when bare nouns are produced (for a detailed discussion on this topic see Bordag, 2004).

The second important finding is that gender-transparency affects L2 production. Subjects were slower when they named pictures with an atypical termination (Group C) and made more errors when the nouns had an atypical or ambiguous termination. In this respect the results replicate those of Bordag et al. (2006) with English native speakers: the subjects in these experiments had the most difficulty with the production of phrases containing nouns with an atypical gender termination, less difficulty when the nouns had an ambiguous termination and were fastest and made the least number of errors when the nouns had a typical termination. Crucially, however, this is the first experiment in which both the gender-interference effect and the gender-transparency effect were obtained at the same time.

Experiment 4

This experiment directly followed Experiment 3. The same subjects were tested. In this experiment we examined whether the processing of gender-atypical or gender-ambiguous nouns leads to slower responses and/or higher error rates in a grammatical judgment task. A similar experiment was reported by Bordag et al. (2006) with English learners of German. They obtained evidence that the noun's termination affects subjects' grammaticality judgment.

Under the assumption that the phonological form affects gender processing in grammaticality judgment tasks, we would expect that a noun phrase with a gender agreement error should be more easily judged as incorrect if the noun has a typical termination than when its termination is ambiguous or atypical (but in accordance with the adjectival gender marking). For instance in the phrase **dieses Blume* "this(n) flower(f)" the feminine noun *Blume* "flower" has a phonological form that is typical for the feminine gender (disyllabic + termination *-e*), but is (incorrectly) marked with a masculine demonstrative pronoun. According to our assumption, subjects should be quick to detect the gender mismatch. On the other hand, a phrase containing a noun with a gender-atypical ending like *Affe* "monkey", masculine , also incorrectly gender-marked (e.g., **diese Affe* "this(f) monkey(m)"), should lead to longer decision latencies and more errors: The phonological form of the noun (misleadingly) indicates the feminine gender and the pronoun agrees with this (misleadingly indicated) feminine gender.

As far as the difference between the gender-congruent and gender-incongruent nouns is concerned, it was impossible to investigate the effect in this experiment. It was not possible for the forms of the demonstrative pronoun *dieser, dieses, diese* to be congruent with the gender that corresponded to the noun's termination in the ambiguous and atypical group and, at the same time, congruent with the gender of the L1 translation equivalent in the gender-incongruent group. In the group of nouns which were both gender-typical and gender-congruent (half of the feminine nouns ending with *-e* from Group A) this was virtually impossible because they would have to be combined with a masculine or neuter form of the determiner (to achieve a no-response in the grammaticality judgment), yet they were all feminine in Czech. We therefore had to refrain from exploring both effects in one experiment and focused only on the gender-transparency effect.

Method

Materials

Noun phrases consisting of a German demonstrative pronoun *dieser* (m), *diese* (f), *dieses* (n) + a noun were elicited. Critical items were the same 48 nouns as in Experiment 3. All critical items were combined with pronoun forms that mismatched their gender, but corresponded (where possible) to the gender, which could be expected according to their termination (e.g **diese*(f) *Käse* (m) "this cheese"; "cheese" is a masculine noun in German, but has a typical feminine termination and is combined here with a feminine pronoun; or **dieses*(n) *Burg*(f) "this castle"; "castle" is a feminine noun in German with a typical phonological form of a masculine or neuter noun and is combined here with a neuter noun).

In Group A (typical, feminine), eight nouns were combined with the masculine form of the pronoun and eight with the neuter form of pronoun (e.g., **dieses Kiste* "this(n) box(f)"; **dieser Straße* "this(m) street(f)"). In the ambiguous Group B, masculine items were combined with the neuter form of the pronoun and vice versa (e.g **dieses Mond* "this(n) moon(m)"; **dieser Herz* "this(m) heart(n)"). In the atypical Group C, all items were combined with pronouns with a gender, which would be expected according to their termination (e.g., **diese Auge* "this(f) eye(n)"; where the phonological form of *Auge* wrongly indicates feminine gender and the pronoun is in agreement with this wrongly expected gender).

Thus, the set consisted of 48 grammatically incorrect critical items (three groups, each containing 16 items). In order to balance the number of nouns of a particular gender, 36 grammatically incorrect filler phrases were added. Consequently, there were noun phrases with 28 feminine, 28 masculine and 28 neuter nouns, to which subjects had to respond with "no". To balance the experiment with respect to the number of yes- and no-responses, additional 84 grammatically correct noun phrases were added as fillers (e.g., *dieser Baum* "this(m)

Table 5. *Results of Experiment 4 with reaction times (RT) in millisecond, standard deviations (SD), and error rates with percentages counted for each experimental condition.*

	A (typical)			B (ambiguous)			C (atypical)		
Condition	RT	SD	Errors	RT	SD	Errors	RT	SD	Errors
	1184	272	35 (12.2%)	1221	307	98 (34%)	1248	302	81 (28.1%)

tree(m)"). Altogether there were 168 items. Forty-eight of them were critical items and 120 fillers. Ten noun phrases (three f, three n, four m) were used as practice items.

Procedure

The participants were tested individually. They were presented written instructions. The stimuli appeared on a computer screen and participants' reaction times and accuracy were recorded by a computer using the experimental software ERTS.

The experiment started with a practice block of 10 items which was identical for all subjects. The main part of the experiment started after a short break. In each trial, after a fixation sign (=*=) of 500 ms, a noun phrase in the form of a demonstrative pronoun + a noun was presented in the center of the monitor. Participants had to decide whether it was grammatically correct by pressing a YES- or NO-button. After pressing one of these buttons, the item disappeared from the screen and after a pause of 500 ms the next trial started. If the subject did not respond within a period of 2500 ms after the presentation of the stimulus, there was a "time out" notification on the screen and after a pause of 500 ms the next trial started. The items were equally distributed in four blocks, each starting with a break, which could be terminated by the subject. Subjects were not given feedback on the correctness of their judgements. The order of the items was pseudo-randomized for each subject with the following restrictions: no more than three identical decisions and no more than two items of the same group or the same gender were allowed to follow each other. One experimental session took about 15 minutes.

Results and discussion

The data were again screened for errors and outliers. Trials in which the subject gave an incorrect response and those trials in which decision latencies were more than three standard deviations away from the mean of the participant were removed from the data. Table 5 gives a summary of the results.

The numerical differences go in line with our expectation that subjects should be fastest with Group A, slower with Group B, and slowest with Group C. This is reflected in the analysis of reaction times, which was significant in F1 and marginally significant in F2, despite the high percentages of errors and thus limited statistical power: $F1$ $(2,39) = 4.71$, $p < .05$ and $F2$ $(2,57) = 2.58$, $p = .09$. The Scheffé test ($p < .05$) showed that Groups A and C differ from each other, but not from Group B.

The analysis of error rates was also significant: $F1$ $(2,34) = .21$, $p < .01$ and $F2$ $(2,45) = 7.13$, $p < .01$. The Scheffé test revealed that Groups B and C differed significantly from Group A, but not from each other. Taken together, the results of this experiment and of Experiment 3 clearly show that subjects have the most difficulty with processing atypically gender-marked nouns, more difficulty with ambiguously marked nouns and the least difficulty with nouns with a typical termination.

General discussion

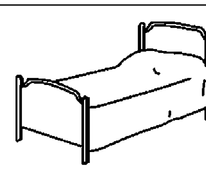
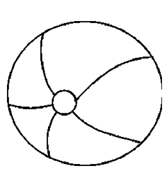
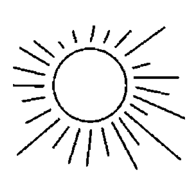
The four experiments reported in this study yielded several important findings concerning L2 gender processing. In the first three experiments we showed that the gender-interference effect from L1 is a robust effect which can be replicated under various conditions. Experiment 2 moreover revealed that unbalanced bilinguals cannot completely deactivate the L1 system or block its influence, even when they are close to the L2 monolingual end of Grosjean's continuum. These results speak in favor of a model, which assumes interaction between the L1 and L2 systems not only at the conceptual and phonological level, but also at the level of grammatical encoding. Moreover, contrary to Costa et al.'s (2003) proposal that the gender systems of L1 and L2 are language specific, our results support a representation, where the three gender nodes are shared by both L1 and L2, at least when the two gender systems have the same number and type of gender nodes as it is the case in Czech and German. If the two languages had distinct gender nodes, whether or not a given word and its translation equivalent had congruent or incongruent genders, it would not have any representational or functional implications. Our data, however, show that the factor congruent vs. incongruent gender affects gender processing suggesting that the two languages share their gender nodes.

In Experiments 3 and 4 we demonstrated that the L2 gender production is affected not only by the L1 grammatical gender, but also by the phonological form of the L2 noun, especially by its termination: Subjects had the least difficulty with nouns with a gender-typical termination and the most difficulty with nouns with a gender-atypical termination. These results speak in favor of a model, which allows interaction between the levels of phonological encoding and grammatical encoding, e.g., the Interactive Activation model of Dell (1986) or, in this respect, the Independent Network model of Caramazza (1997). In contrast, they pose a problem for serial, modular models of speech production like the Levelt (1989) model, which assume that first the selection processes at the grammatical level are completed, and only then the phonological encoding starts, making it thus impossible for phonological features to affect the selection processes (e.g., gender selection) at the higher grammatical level. Evidence in support of an interactive model of L2 processing is reported also by Lemhöfer, Spalek and Schriefers (submitted), who observed a larger cross-language gender effect for cognates than for non-cognates in both a production and a comprehension experiment.

Both the gender-interference effect and the gender-transparency effect support the notion of gender selection based on activation levels. Since the L2 gender selection is affected by both the L1 gender and the L2 phonological form, it seems very unlikely that the gender access could be a direct and automatic consequence of lexical selection, where the activation levels of the target node and its competitors do not play any role. Rather, the influence of the L1 gender and of the phonological form affects the activation levels of the target gender node and its competitors. This renders the competition for selection easy, if all activation collides just on one node, and more difficult, if there are more strongly activated competitors. Our data thus speak for an activation-dependent model, which assumes that gender information is not stored as a firm feature at each lemma, but is computed each time it is needed anew on the basis of all available information (including stored information based on input and previous competition results).

It is, however, important to mention that different gender-retrieval mechanisms may be at work in the L1 and L2 production. Bordag et al. (2006) report that they found the gender-transparency effect in L2, but not in L1 German. They hypothesize that the differences between the L1 and L2 data may reflect differences between the acquisitional stages of the two groups of subjects: A phonological form (or an L1 gender) may play a role in gender retrieval in initial or even low advanced stages of acquisition and may become irrelevant later. This could explain seemingly contradictory data from L1 and L2 subjects, as well as from intermediate and low advanced L2 speakers in this study and highly-proficient bilingual speakers in the studies like Costa, Kovačić, Franck and Caramazza (2003). As proposed also by Costa et al. (2003), more research is obviously needed to address the impact of degree of proficiency (or age of acquisition etc.) on the role of L1 or of L2 phonological factors in speech processing.

Appendix A: Examples of pictures used in the experiments

picture	congruent	picture	incongruent
	*Ohr* (n) *ucho* (n) “ear”		*Bett* (n) *postel* (f) “bed”
	*Ball* (m) *míč* (m) “ball”		*Fuss* (m) *noha* (f) “foot”
	*Kerze* (f) *svíčka* (f) “candle”		*Sonne* (f) *slunce* (n) “sun”

Appendix B: The list of experimental items used in Experiments 1 and 2

Gender-congruent condition						Gender-incongruent condition					
German (L2)		Czech (L1)		*English*		German (L2)		Czech (L1)		*English*	
Ball	m	míč	m	*ball*		Fuss	m	noha	f	*foot*	
Hund	m	pes	m	*dog*		Stuhl	m	židle	f	*chair*	
Zug	m	vlak	m	*train*		Kopf	m	hlava	f	*head*	
Baum	m	strom	m	*tree*		Wein	m	víno	n	*wine*	
Brief	m	dopis	m	*letter*		Fisch	m	ryba	f	*fish*	
Zahn	m	zub	m	*tooth*		Schrank	m	skříň	f	*wardrobe*	
Teller	m	talíř	m	*plate*		Sessel	m	křeslo	n	*arm-chair*	
Teppich	m	koberec	m	*carpet*		Spiegel	m	zrcadlo	n	*mirror*	
Maus	f	myš	f	*mouse*		Burg	f	hrad	m	*castle*	
Hand	f	ruka	f	*hand*		Stadt	f	město	n	*town*	
Kerze	f	svíčka	f	*candle*		Nase	f	nos	m	*nose*	
Karte	f	mapa	f	*map*		Sonne	f	slunce	n	*sun*	
Dusche	f	sprcha	f	*shower*		Brücke	f	most	m	*bridge*	
Katze	f	kočka	f	*cat*		Kirche	f	kostel	m	*church*	
Tafel	f	tabule	f	*blackboard*		Butter	f	máslo	n	*butter*	
Krone	f	koruna	f	*crown*		Tomate	f	rajče	n	*tomato*	
Herz	n	srdce	n	*heart*		Schiff	n	lod'	f	*ship*	
Feld	n	pole	n	*field*		Bett	n	postel	f	*bed*	
Ei	n	vejce	n	*egg*		Pferd	n	kůň	m	*horse*	
Ohr	n	ucho	n	*ear*		Dorf	n	vesnice	f	*village*	
Fenster	n	okno	n	*window*		Eis	n	zmrzlina	f	*ice-cream*	
Meer	n	moře	n	*sea*		Geschenk	n	dárek	m	*present*	

Appendix C: The list of critical items used in Experiments 3 and 4

Group A (Gender Typical Items)

German	Gender	Czech	Gender (in-)congruecy	English
Brücke	f	most	gi	bridge
Wolke	f	mrak	gi	cloud
Glocke	f	zvon	gi	bell
Blume	f	květina	gc	flower
Kerze	f	svíčka	gc	candle
Zunge	f	jazyk	gi	tongue
Schlange	f	had	gi	snake
Kirche	f	kostel	gi	church
Sonne	f	slunce	gi	sun
Birne	f	hruška	gc	pear
Mütze	f	čepice	gc	cap
Fliege	f	moucha	gc	fly
Straße	f	silnice	gc	street
Katze	f	kočka	gc	cat
Kirsche	f	třešeň	gc	cherry
Nase	f	nos	gi	nose

Group B (Gender Ambiguous Items)

German	Gender	Czech	Gender (in-)congruecy	English
Messer	n	nůž	gi	knife
Dach	n	střecha	gi	roof
Blatt	n	list	gi	leaf
Glas	n	sklenice	gi	glass
Hemd	n	košile	gi	shirt
Fenster	n	okno	gc	window
Herz	n	srdce	gc	heart
Buch	n	kniha	gi	book
Kopf	m	hlava	gi	head
Mond	m	měsíc	gc	moon
Zahn	m	zub	gc	tooth
Topf	m	hrnec	gc	pot
Schlüssel	m	klíč	gc	key
Stein	m	kámen	gc	stone
Stern	m	hvězda	gi	star
Brief	m	dopis	gc	letter

Group C (Gender Atypical Items)

German	Gender	Czech	Gender (in-)congruecy	English
Stadt	f	město	gi	town
Burg	f	hrad	gi	castle
Zwiebel	f	cibule	gc	onion
Gabel	f	vidlička	gc	fork
Nuss	f	ořech	gi	nut
Mauer	f	zed'	gc	wall
Wurst	f	buřt	gi	salami
Ampel	f	semafor	gi	traffic lights
Insel	f	ostrov	gi	island
Butter	f	máslo	gi	butter
Maus	f	myš	gc	mouse
Hand	f	ruka	gc	hand
Käse	m	sýr	gc	cheese
Hase	m	zajíc	gc	rabbit
Affe	m	opice	gi	monkey
Auge	n	oko	gc	eye

References

Beringer, J. (1999). *Experimental run time system*. Frankfurt: BeriSoft Cooperation.

Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. In S. Foster-Cohen (ed.), *EUROSLA Yearbook 4 (2004)*, pp. 203–230. Amsterdam: John Benjamins.

Bordag, D. (2006). *Psycholinguistische Aspekte der Interferenzerscheinungen in der Flexionsmorphologie des Tschechischen als Fremdsprache*. Hildesheim: G. Olms Verlag.

Bordag, D., Opitz, A. & Pechmann, T. (2006). Gender processing in L1 and L2: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1090–1101.

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

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. & Alario, F. X. (2001). A crosslinguistic investigation of determiner production. In E. Dupoux (ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*, pp. 209–226. Cambridge, MA: MIT Press.

Costa, A. & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2, 231–244.

Costa, A., Caramazza, A. & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283–1296.

Costa, A., Kovačić, D., Fedorenko, E. & Caramazza, A. (2003). The gender congruency effect and the selection of freestanding and bound morphemes: Evidence from Croatian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1270–1282.

Costa, A., Kovačić, D., Franck, J. & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, 6(3), 181–200.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13, 1–24.

De Bot, K. & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 191–214. Amsterdam: John Benjamins.

De Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representation. In R. Frost & L. Katz (eds.), *Orthography, phonology, morphology, and meaning*, pp. 389–412. Amsterdam: Elsevier.

De Groot, A. M. B., Dannenburg, L. & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 33(5), 600–629.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dijkstra, A. F. J. [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(4), 496–518.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305–352.

Grainger, J. & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (ed.), *Cognitive processing in bilinguals*, pp. 207–220. Oxford: North-Holland.

Green, D. W. (1993). Towards a model of L2 comprehension and production. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 249–278. Amsterdam: John Benjamins.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, 3, 233–274.

Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In A. de Groot & J. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 225–254. Mahwah, NJ: Erlbaum.

Grosjean, F. (1998). Transfer and language mode. *Bilingualism: Language and Cognition*, 1(3), 175–176.

Hermans, D., Bongaerts, T., De Bot, K. & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1(3), 213–229.

Köpcke, K.-M. & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik*, 11, 166–182.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Kroll, J. F. & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 531–553. New York: Oxford University Press.

La Heij, W., Mak, P., Sander, J. & Willeboordse, E. (1998). The gender congruency effect in picture-word tasks. *Psychological Research*, 61, 209–219.

Lemhöfer, K., Spalek, K. & Schriefers, H. (submitted). Cross-language effects of grammatical gender in bilingual word recognition and production.

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

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98(23), 13464–13471.

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

Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language*, 35, 757–774.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German*. Berlin: Springer Verlag.

Paradis, M. (1981). Neurolinguistic organization of a bilingualism. *LACUS Forum*, 7, 486–494.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15(1), 36–57.

Schiller, N. O. & Caramazza, A. (2002). The selection of grammatical features in word production: The case of plural nouns in German. *Brain and Language*, 81, 342–357.

Schriefers, H. (1993). Syntactic processes in the production of noun phases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 841–850.

Schriefers, H. J. & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production. *Journal of Psycholinguistic Research*, 28(6), 575–600.

Schriefers, H. J. & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology*, 26(6), 1368–1376.

Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.

Vigliocco, G., Lauer, M., Damian, M. F. & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 46–58.

Received February 24, 2006
Revision received June 16, 2006
Accepted August 13, 2006