



HAL
open science

Grammatical gender in translation

Denisa Bordag, Thomas Pechmann

► **To cite this version:**

Denisa Bordag, Thomas Pechmann. Grammatical gender in translation. *Second Language Research*, 2008, 24 (2), pp.139-166. 10.1177/0267658307086299 . hal-00570741

HAL Id: hal-00570741

<https://hal.science/hal-00570741>

Submitted on 1 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Grammatical gender in translation

Denisa Bordag and Thomas Pechmann

University of Leipzig

Received February 2007; revised July 2007; accepted July 2007

In three experiments native speakers of Czech translated bare nouns and gender-marked adjective + noun phrases into German, their second language (L2). In Experiments 1–3 we explored the so-called gender interference effect from first language (L1) as observed in previous picture naming studies (naming latencies were longer when the L1 noun and its L2 translation had different genders than when their genders were congruent). In Experiments 2 and 3 we investigated the influence of gender transparency in L2 (longer latencies when an L2 noun has a gender-atypical or gender-ambiguous termination than when its termination is gender-typical). Although both effects were observed in L2 picture naming, only the gender transparency effect could be demonstrated in L1 to L2 translation tasks. The resulting constraints on L2 gender processing during translation are discussed in the framework of bilingual speech production models.

Keywords: grammatical gender, translation, bilingualism, speech production, picture naming, Czech, German

I Introduction

Despite the initial controversies on this topic, most research evidence concerned with bilingual speech processing converges on the conclusion that both languages become activated in parallel and may compete for selection during speech production in L2 (De Bot, 1992; Green, 1993; Poulisse and Bongaerts, 1994; Grosjean, 1998; Hermans *et al.*, 1998).

Address for correspondence: Denisa Bordag, University of Leipzig, Department of Linguistics, Beethovenstraße 15, D-04107 Leipzig, Germany; email: denisav@rz.uni-leipzig.de

Many related questions, however, remain open. Although there is no reason to assume *a priori* that the interaction between first language (L1) and second language (L2) and their representations is the same for semantics, grammar, phonology and orthography, early bilingual research and models have not considered distinctions among levels of representation. Yet, as also pointed out by Kroll and Tokowicz (2005), the question about common vs. separate L1 and L2 systems may be answered differently for different aspects of language representation. Moreover, differences are to be expected for bilinguals whose particular languages constrain the possibilities of shared representations and whose language learning histories may be different and thus result in different arrangements and processing between their L1 and L2 systems. Particularly the age of acquisition (early vs. late bilinguals), proficiency (balanced vs. unbalanced bilinguals) and learning method (e.g. classroom translation method vs. immersion) may strongly affect the development of bilingual language representation and processing.

Another aspect that demonstrably affects the degree of L1 and L2 parallel activation and interaction is the type of task bilinguals perform. Previous research has shown that effects that were observed in picture naming cannot be replicated in translation, or only partially. As an example, Costa and Caramazza (1999) observed that semantic interference does not depend on the language of the distractor (i.e. a word that is presented together with the target picture and affects the naming latencies, although participants are instructed to ignore it). When Spanish–English bilinguals named pictures in Spanish (L1) with distractors in Spanish (L1) and in English (L2), a semantic interference effect – i.e. longer naming latencies when the target noun and the distractor word were semantically related – was detected in both cases. Contrary to the results in picture naming, Miller and Kroll (2002) observed semantic interference in translation only when the distractor appeared in the target language; for example, when the bilinguals translated from L1 into L2 and the distractor was in the L2, and not when the distractor was in the input language, i.e. in L1. Miller and Kroll conclude that cross-language competition may be reduced if a cue is available that allows production to proceed selectively. In translation, such a cue is present in the nature of the input (do not speak the input language, i.e. L1). According to Miller and Kroll, such language-specific information in the event that

initiates speech planning may prompt the production to proceed more selectively than in tasks like picture naming, where such a cue is absent.

In our study, we focus on the interaction of L1 and L2 systems at the level where grammatical information about L1 and L2 is stored. The theoretical starting point of this study is the serial, discrete model of Levelt (Levelt, 1989; Levelt *et al.*, 1999), the only production model that has been explicitly adopted for bilingual production (De Bot, 1992; 2003).

In this model (as well as in most other production models), one of the features that is processed at the grammatical level (also called the lemma level) is grammatical gender. It is represented independently of a word's phonological form and semantics. The independent representation of gender information is supported by a large number of experimental studies (for a review, see Schriefers and Jescheniak 1999), by studies investigating the tip-of-the-tongue phenomenon (Caramazza and Miozzo, 1997; Vigliocco *et al.*, 1997) as well as by the performance of aphasic patients (Badecker *et al.*, 1995; Avila *et al.*, 2001).

According to the Levelt model, each gender value (e.g. masculine, feminine and neuter in Czech or German) is represented as an abstract gender node. The selection of the node depends on the level of activation of the other gender nodes, its competitors. If there is a highly activated competitor, the selection is more difficult, and this is usually reflected in slower reaction times.

The grammatical level has long stood in the background of bilingual research, which is dominated by studies concentrating on parallel activation and interaction of the two systems at the semantic level (Weinreich, 1953; De Groot, 1993; Kroll, 1993; Francis, 1999) and at the phonological level (Roelofs, 2003; for exploration of the cognate facilitation effect, see, for example, Kroll *et al.*, 1998; Janssen, 1999; Costa *et al.*, 2000; Kroll *et al.*, 2000). Recently, however, several studies have been conducted that address bilingual processing at the grammatical level, especially the processing of grammatical gender. Although some authors (Costa *et al.*, 2003) have not found evidence of competition for selection between the L1 and L2 gender nodes of two translation equivalents in picture naming, other authors report systematic evidence for the so-called gender interference effect from L1, both in L2 picture naming (Bordag, 2004; Bordag and Pechmann, 2007) and in comprehension tasks (Lemhöfer *et al.*, 2003; Hagoort, 2003). The gender interference effect

manifests itself in longer naming latencies when L2 nouns (or NPs) that differ in gender from their L1 translation equivalents are produced, e.g. *grosses Haus* (neuter, German) vs. *velký dům* (masculine, Czech) [big house], compared to L2 nouns that have congruent gender with their L1 translation equivalent, e.g. *grosser Teller* (masculine, German) vs. *velký talíř* (masculine, Czech) [big plate].

Given the differences in picture naming and translation in, for example, the processing of semantic information, the immediate question arises as to whether similar differences could also be observed at the level of grammatical encoding. More particularly, do the gender representations of translation equivalents compete for selection only in L2 picture naming, or can we also find a similar cross-language effect in translation from L1 into L2? On the one hand, we can expect gender interference in translation, because if a gender marked NP (and maybe also a bare noun alone) has to be translated, e.g. from L1 into L2, both the L1 gender of the source language and the L2 gender of the target language must be accessed. Thus, an important pre-condition for competition for selection – i.e. activation of multiple gender nodes (if the L1 and L2 translations have different genders) – is satisfied. On the other hand, picture naming and translation differ in a number of important aspects that might affect the gender retrieval processes. As an example, both the L1 and L2 gender nodes might be activated, but at different time points, so that another important pre-condition for the competition, i.e. parallel activation of multiple gender nodes, would not be satisfied. The experiments reported in this study explore the interference factors affecting gender retrieval in translation.

The second research question addressed in our study concerns the processes of L2 gender selection in translation within the L2 system. Previous evidence from picture naming and gender monitoring suggests that L2 gender selection is influenced not only by interlingual interference (congruency vs. incongruency of the L1 gender), but also by intralingual interference from L2 phonological forms. This contrasts with the findings on adult L1 gender processing (but complies with the data on child L1 gender processing). In their studies, Bordag *et al.* (2006) and Bordag and Pechmann (2007) showed that the selection of grammatical gender at the level of grammatical encoding in L2 is affected by information from the lower level of phonological encoding, thus providing

support for cascaded processing in L2 (for converging evidence see, for example, Bates *et al.*, 1995; Andonova *et al.*, 2004): Participants were fastest when producing gender-marked noun phrases containing a noun with a gender-typical termination and slowest when the noun had a gender-atypical termination (see below for details). In this article we address the same topic, but in translation. We expect that the gender transparency effect (gender transparency or non-transparency of the L2 noun termination) should also be observed in translation but, again, task-specific procedures might render the influence of the word form ineffective. Experiments 2 and 3 in this study address this issue.

The expansion of research on the task of translation is important for several reasons. First, as already mentioned above, the evidence observed in one paradigm in L2 research often does not converge with evidence observed when participants perform a different task, suggesting that different processes, routes or levels of representation may be involved. Determining how processing is affected by the different tasks may help to clarify crucial issues concerning how bilinguals represent and process lexical items from the L2. Second, although very easy to perform (from a methodological point of view), the largely neglected task of translation has, to date, been only narrowly investigated in psycholinguistic research. In 1993, Snodgrass made the criticism that ‘relatively little information exists on performance in the translation task’ (Snodgrass, 1993: 85), and concludes that ‘the research on the bilingual lexicon using the technique of translation times needs more data rather than more theory’ (Snodgrass, 1993: 110). More than 10 years later, little has changed in this respect. New paradigms and research methods have been employed in L2 research (for review, see Kroll and De Groot, 2005), but only a small number of studies have focused on translation. Moreover, the topics addressed in this area of research are rather few. In most cases, the translation task is used to investigate asymmetries in bilingual processing (backward vs. forward translation) and to explore possible processing routes used by bilingual participants of different proficiencies by comparing the speed of their performance in word naming, picture naming and translation.

The empirical starting point of our research concerning the cross-language interference effect is a series of picture naming experiments with Czech–German unbalanced bilinguals reported by Bordag (2004; 2006)

and Bordag and Pechmann (2007). In these experiments, speakers of L2 Czech and German randomly produced either bare nouns or gender-marked adjectives + nouns under various conditions. The articulation latencies were significantly longer and the gender error rates significantly higher when participants named pictures from the gender-incongruent group – i.e. those whose names had different genders in L1 and L2 – than when they named pictures from the gender-congruent group. In the control experiments with native speakers who named the same pictures in their L1, no difference between the two sets of pictures was observed.

Bordag (2004) proposed that the gender interference effect from L1 is similar in nature to the gender-congruency effect obtained in gender experiments with picture–word distractors in native languages: If the gender of an L1 noun differs from that of its L2 translation equivalent, their lemmas compete for selection and, with them, their corresponding gender representations that they automatically activate. Selection of the L2 lemma is thus delayed compared to the condition where both translation equivalents have congruent genders just as in the picture–word distractor experiments in L1, in which the picture and the distractor gender nodes are assumed to compete for selection.

In the translation experiments described in this article we utilized the same materials as Bordag (2004) and Bordag and Pechmann (2007) to guarantee comparability. The participants' L1 was Czech and the L2 was German. Both of these languages have the same number and types 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 – e.g. *Teller* (masculine, German) vs. *talíř* (masculine, Czech) [plate] – while the other half does not, e.g. *Haus* (neuter, German) vs. *dům* (masculine, Czech) [house]. Sometimes the gender can be derived from a noun's termination, but this is only true for transparently gender-marked nouns (see below). The participants of the picture naming and translation experiments were drawn from the same population. They were late, unbalanced bilinguals who had learnt German mostly in a classroom setting, often using the grammar–translation method. At the time of testing, most of them were exchange students who came to study various subjects at the University of Leipzig for one or two semesters.

II Experiment 1

1 Method

a Participants: Eighteen native speakers of Czech (aged 21–32) participated in the experiment, most of them being exchange students at the University of Leipzig at the time of testing. All participants had been learning German for at least five years (10.2 years on average), but they often experienced long periods when they were either using German very rarely or not at all. They all started to learn German after the age of nine. They were drawn from the same population as participants of the picture naming experiments reported by Bordag (2004) and Bordag *et al.* (2006), and were paid for their participation. Their L2 proficiency ranged from upper-intermediate to advanced. This was determined on the basis of a questionnaire concerning participants' language background and L2 language history, their self-evaluation, evaluation by the experimenter (based on a conversation in L2) and participants' ratings of familiarity with the L2 vocabulary used in the study. For details see Appendix 1 and the section Procedure.

b Materials and design: The stimulus set (see Appendix 2) was identical to the one used in the picture naming experiments reported in Bordag (2004), and included 22 gender-congruent, 22 gender-incongruent and 22 filler nouns. The items from the two critical groups (congruent vs. incongruent) were controlled for length, frequency, familiarity and noun termination (most items had a gender-typical termination; the number of exceptions was kept constant across the two groups and the three genders). All nouns were concrete, highly imaginable, well depictable and well known in both languages. However, in contrast to the studies mentioned above, stimuli were not presented as pictures but as written words. Critical items were displayed in Czech, and had to be translated into German. The filler items were displayed in German, and had to be translated into Czech. The critical items are listed in Appendix 3.

Similar to the picture naming experiments, the design also included the conditions short vs. long. In the short condition, a bare noun was presented, and participants had to produce its translation equivalent, e.g. *dům-Haus* (house). In the long condition, the noun was preceded by

either a small or large dot, which signalled that the noun had to be additionally modified by an adjective. If the dot was small, participants were instructed to use the adjective *klein* (small), e.g. · *dûm* – *kleines Haus* (small house), if the dot was large, the adjective *groß* (large), e.g. • *dûm* – *großes Haus* (large house) had to be used. Previous experiments investigating gender retrieval in both native and non-native languages have shown that overt gender realization (e.g. as a gender marked ending of an adjective or as a gender-marked article) might be a crucial factor for the manifestation of gender competition effects. Therefore both the long (overt gender realization on the adjective ending) and the short condition (only bare nouns) were included in the present experiments as well.

c Procedure: Each session started with an informal conversation with the participant in order to obtain a rough estimate of the participant's L2 proficiency, which helped to determine whether he or she was eligible for participation. Participants were tested individually, and the entire procedure lasted approximately 40 minutes. The experimental session consisted of three parts.

Before the experiment proper started, participants were presented with the instructions, which emphasized the importance of both speed and accuracy of their responses. Furthermore, participants received a booklet including all nouns used in the experiment, listed as German–Czech word pairs. Each word pair had two 7-point rating scales next to it and participants were asked to rate both the familiarity of each German word, and the adequacy of the Czech translation. The ratings were conducted, first, to ensure that all participants were using the same translation equivalents in the actual experiment, second, to collect data on the stimulus material, especially on the familiarity of the German words and, finally, to check that the words in the gender-congruent and gender-incongruent conditions were known equally well. Later analyses of the ratings revealed that the German words from the two conditions were equally familiar to all participants.

The presentation of the stimuli and the recording of the articulation latencies were controlled by a computer running ERTS (Experimental Run Time System; Beringer, 1999). Stimuli were presented on an NEC MultisyncXV 17-inch monitor with a screen resolution of 640 × 480 pixels. Articulation latencies were recorded with a Sony ECM-MS957

microphone that was connected to an ERTS voice-key and a DAT-recorder. The experimenter, seated behind the participant, listened to his/her responses via headphones and checked their accuracy. The experiment started with a 14-item training block, which was identical for all participants. The items from this block did not appear among experimental items, but were included in the booklet. The training block was run twice, and participants received feedback from the experimenter on the accuracy of their responses (e.g. a wrong translation or adjective, too slow response), yet not on potential gender errors (i.e. a wrong adjective ending). The experimental items were presented in two blocks separated by a short break. During the critical trials participants received no feedback.

Participants initiated each trial by pressing the space bar. After 300 ms of blank screen, a white fixation point (= * =) was displayed on black background in the centre of the screen for 1000 ms. After a break (a blank screen for 300 ms), the target stimulus was presented. In the short condition the noun was displayed alone, and in the long condition the noun was preceded by a small or a large dot. The onset of articulation triggered the voice-key, and the response latency was digitally recorded. The colour of the displayed word (initially white) changed depending on the participant's response. If the voice-key recorded a response, the word turned blue; if no response was recorded within 3000 ms after the stimulus onset, the word turned red signalling time-out.

Participants were randomly assigned to 2 randomized versions of the experiment. Each item was presented in both the long and the short condition. Items in the long and short conditions followed each other in a pseudo-random order. Half of the participants saw half of the items first in the long condition, and the other half of the participants saw these items in the short condition first. The two versions of the experiment thus differed in which items were presented in the long condition first.

Each participant was presented with a unique randomization of one of the two versions of the experiment. Randomizations were created by applying the following two restrictions:

- no direct succession of more than three trials of the same condition (short vs. long or congruent vs. incongruent), the same adjective (*klein* vs. *groß*), or the same gender (masculine, feminine, neuter);
- no direct succession of more than four trials in the same language.

The order of stimulus presentation was also balanced with respect to the number of congruent vs. incongruent items appearing in a language switch/no-switch condition and a length (short vs. long) switch/no-switch condition.

In the last part of the experimental session, participants answered a questionnaire concerning their knowledge of German, and their general language background.

2 Results and discussion

The data were screened for errors. Observations for the response time (RT) analysis were discarded mainly due to voice-key errors, gender errors and task errors (a phrase translated with a wrong adjective, produced in a wrong language, etc.). The cut-off was two standard deviations from a participant's mean response time in the long and in the short condition. Altogether, 16.6% of observations were excluded from the analyses.

Analyses of variance (ANOVAs) were performed on the mean naming latencies per participant (F_1) and per item (F_2) with the factors 'length' (2 levels, within participants, within items) and 'gender congruency' (2 levels, within participants, between items). Overall, naming latencies in the long condition were longer than in the short condition: $F_1(1, 17) = 58.85, p < .001$ and $F_2(1, 21) = 221.542, p < .001$.

More importantly, however, we did not observe any gender interference effect in the analyses of reaction times: Nouns and NPs from the gender-incongruent set were translated as fast as nouns and NPs with a congruent gender in both languages ($F_s < 1$).

Table 1 Results of Experiment 1 with reaction times in milliseconds and error rates with percentages counted for each experimental condition

Condition	Congruent			Incongruent			Interference effect
	RT	SD	Errors	RT	SD	Errors	
Adjective + noun	1250	335	107 (27%)	1246	374	112 (28%)	-4
Noun	959	254	28 (7%)	968	245	52 (13%)	+9

Concerning the gender errors alone (in the long condition), participants made more gender errors in the gender-incongruent (47 errors) than gender-congruent condition (25 errors): chi squared = 6.722, $df = 1$, $p = .010$.

These results contrast sharply with those of Bordag (2004; 2006) and Bordag and Pechmann (2007), who report a gender interference effect in both the short and long condition in picture naming experiments using the same materials. Before discussing this difference, we report the results of Experiment 2.

III Experiment 2

Having observed no significant difference between the two critical conditions (congruent vs. incongruent) in Experiment 1, we decided to replicate the task with another set of materials, which also yielded a significant gender interference effect in picture naming (Experiment 3 in Bordag and Pechmann, 2007). However, this set of items was selected to investigate not only the influence of the L1 gender, but also the influence of gender transparency of the L2 noun termination. In the picture naming experiment reported by Bordag and Pechmann, participants had the least difficulty with items with a gender-typical termination and most difficulties with items with a gender-atypical termination. Gender transparency was carefully controlled in the materials used in our Experiment 1 reported above (nouns from the gender-congruent and gender-incongruent conditions were matched for their termination), but that experiment was not designed to explore this issue (e.g. nouns with a gender-atypical termination were, apart from a few exceptions, almost missing). In the materials used in Experiment 2, the factor of gender transparency was systematically manipulated.

1 Method

a Participants: Eighteen native speakers of Czech (aged 23–35) participated in the experiment. They were drawn from the same population as participants of Experiment 1. They all started to learn German after the age of nine and had been using the language for 11.5 years on average. For details, see Appendix 1.

b Materials and design: We used the same set of materials as Bordag and Pechmann (2007)¹ in their Experiment 3 (picture naming). Sixteen items had a gender-typical termination (Group A), 16 a gender-ambiguous termination (Group B) and 16 a gender-atypical termination (Group C). The gender-typical group consisted of German feminine nouns with a typically feminine termination -e, the gender-ambiguous group consisted of masculine and neuter nouns ending with a consonant, and the atypical group contained feminine nouns ending with a consonant and masculine and neuter nouns ending with an -e (compare Köpcke and Zubin, 1983; Mills, 1986). To balance the experiment with regard to the number of different genders, masculine and neuter fillers were included as well. Half of the nouns in each gender-typical, ambiguous and atypical group had the same gender as their translation equivalents (gender-congruent condition), and 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 L2.² Nouns in the typical vs. ambiguous vs. atypical as well as in the gender-congruent vs. gender-incongruent group in this experiment were matched for frequency and length in L1 (Czech) and for formal similarity between L1 (Czech) and L2 (German). All nouns were presented in Czech and had to be translated into German; there were no fillers to be translated in the opposite direction. In all other respects, the design was identical to that of Experiment 1. The items are listed in Appendix 3.

¹ Four items were exchanged compared to Experiment 3 in Bordag and Pechmann (2007) to achieve better matching of length of the nouns in the critical groups. (The length of L1 nouns was not controlled in the picture naming experiment of Bordag and Pechmann, 2007.)

² Average ratings of the experimental items used in the Bordag and Pechmann (2006) study by the German (L1) and English (L2) natives (the scale was from 0 to 6; 6 was the best mark).

German is	Gender transparency		L1–L2 similarity		Familiarity	
	L2	L1	L2	L1	L2	L1
Group A	5.43	4.23	3.20	2.68	4.90	–
Group B	3.50	3.49	2.74	2.56	5.07	–
Group C	2.05	3.09	3.20	2.46	4.99	–

c Procedure: The procedure was the same as in Experiment 1. Additionally, the items were pseudo-randomized in such a way that no more than 3 items from the gender typical, ambiguous or atypical termination followed each other in successive trials.

2 Results and discussion

The data were screened for errors and outliers applying the same criteria as in the previous experiment (18.2%). Naming latencies were submitted to two separate ANOVAs, one with the factor ‘gender congruency’ and one with the factor ‘gender transparency’.³

In the first ANOVA with the factors ‘length’ (2 levels, within participants, within items) and ‘gender congruency’ (2 levels, within participants, between items), the factor ‘length’ was significant: $F_1(1, 17) = 31.792$, $p < .001$; $F_2(1, 23) = 137.790$, $p < .001$. The factor ‘gender congruency’ was, however, insignificant and the interaction between these two factors was not significant either (all analyses $F < 1$).

The analyses of errors showed no significant differences between the number of errors in the gender-congruent and gender-incongruent conditions either. The factor ‘length’ was again highly significant: $F_1(1, 17) = 46.095$, $p < .001$, $F_2(1, 23) = 28.802$, $p < .001$. The interaction between these two factors was not significant ($F < 1$).

Table 2 Results of Experiment 2 regarding ‘gender congruency’ with reaction times in milliseconds and error rates with percentages counted for each experimental condition

Condition	Congruent			Incongruent			Interference effect
	RT	SD	Errors	RT	SD	Errors	
Adjective + noun	1127	425	116 (27%)	1129	406	120 (28%)	+2
Noun	851	253	55 (13%)	878	271	46 (11%)	+27

³ We decided to report the two separate ANOVAs, and not one ANOVA, with the factors ‘length’, ‘gender congruency’ and ‘gender transparency’, because of the inherent problem with F2, which such an analysis would imply: The factors ‘gender congruency’ (2 levels) and ‘gender transparency’ (3 levels) are between-item factors. To perform an F2 where both these factors would be included, six (2×3) items would always have to be merged into one ‘superitem’, so that the F2 would be run only over 8 merged items (total of 48 items divided by 6). However, we ran the F1 with all three factors (‘length’, ‘gender congruency’ and ‘gender transparency’), and the interaction ‘gender congruency’ vs. ‘gender transparency’ was not significant. This was true of both Experiment 2 and Experiment 3. Consequently, no important information is lost when reporting two separate ANOVAs, and the analyses are more precise.

In the second ANOVA with the factors ‘length’ (2 levels, within participants, within items) and ‘gender transparency’ (3 levels, within participants, between items), the factor ‘length’ was again significant in both F1 and F2: $F1(1, 17) = 31.545, p < .001$; $F2(1, 15) = 318.518, p < .001$.

Crucially, the factor ‘gender transparency’ was significant both in F1 ($F1(2, 34) = 6.631, p < .01$) and in F2 ($F2(2, 30) = 6.071, p < .01$). According to a *post hoc* Scheffé test ($p < .05$), Group A differed significantly from both Groups B and C (which were statistically identical) in F1, but not in F2. The interaction between ‘group’ and ‘length’ was marginally significant in F1 ($F1(2, 34) = 3.056, p = .060$) and significant in F2 ($F2(2, 30) = 4.379, p < .05$), suggesting that the interference effect appeared only in one of the two (long vs. short) conditions. The separate analyses for these two conditions revealed that the effect was significant only in the long condition ($F1(2, 34) = 5.766, p < .01$; $F2(2, 30) = 6.399, p < .01$), but not in the short condition ($F < 1$). The Scheffé test confirmed that in the long condition, Group A was faster than Groups B and C, which were statistically identical.

The analyses of error rates mirrored the analyses of reaction times. The factor ‘gender transparency’ was significant in both F1 ($F1(2, 34) = 10.154, p < .001$) and in F2 ($F2(2, 30) = 9.324, p < .001$). 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 pattern is also reflected in the number of gender errors alone (4 gender errors in Group A, 62 in Group B and 49 in Group C). The factor ‘length’ was significant, too: $F1(1, 17) = 46.095, p < .001$; $F2(1, 15) = 35.209, p < .001$. The interaction between these two factors was also significant ($F1(2, 34) = 16.681, p < .001$; $F2(2, 30) = 13.800, p < .001$), confirming what is obvious from the numerical differences themselves, namely that the effect appeared only in the long condition.

Table 3 Results of Experiment 2 regarding ‘gender transparency’ with reaction times in milliseconds and error rates with percentages counted for each experimental condition

Condition	A (typical)			B (ambiguous)			C (atypical)		
	RT	SD	Errors	RT	SD	Errors	RT	SD	Errors
Adjective + noun	1059	391	38 (13%)	1180	436	110 (38%)	1149	417	88 (31%)
Noun	846	271	39 (14%)	870	278	32 (11%)	878	238	30 (10%)

Neither in Experiment 1 nor in Experiment 2 did we observe the gender interference effect that was reported in picture naming. However, in Experiment 2, we were able to replicate the intralingual gender transparency effect. These results suggest that during the task participants had to complete, the L1 gender did not interfere with the L2 gender in the incongruent condition. Before concluding that different processes are involved during gender retrieval in picture naming and translation, it is necessary to explore whether the lack of the gender interference effect is not caused by the specific design of our experiments. In Experiments 1 and 2, participants did not have to perform pure translation of noun phrases, but the size adjective that they had to produce in the long condition (which is the carrier of the gender marking) was determined by the size of a dot that appeared in front of the word. One might assume that under these circumstances the production system avoids the selection of the L1 gender, rendering no competition for selection between the L1 and L2 gender node necessary. In Experiment 3, we decided to change the task into a pure translation task, in which participants translated complex noun phrases from their L1 into their L2 in the long condition.

IV Experiment 3

Experiment 3 is a replication of Experiment 2, except that participants translated either bare nouns or complex NPs consisting of a gender-marked adjective and a noun into their L2.

1 Method

a Participants: Fifteen native speakers of Czech (aged 23–29) participated in the experiment. They were drawn from the same population as participants of Experiments 1 and 2. They all started to learn German after the age of ten and had been using the language for 9.8 years on average. None of them participated in Experiment 2. For details, see Appendix 1.

b Procedure: The same as in Experiment 2.

c Materials and design: The same set of materials was used as in Experiment 2. The design was also the same as in Experiment 2 with

the exception that instead of a dot an appropriate size adjective in the correct gender form appeared in front of the noun in the long condition. Thus, participants had to translate NPs like *velký strom* ('big(M) tree') or *malá ruka* ('small(F) hand') into their L2: *großer Baum* or *kleine Hand*, respectively. The L1 gender of the noun was thus explicitly marked on the ending of the adjective, which should increase the probability that it is activated and selected in L1.

2 Results and discussion

After screening the data for errors (14.6%), naming latencies were submitted to two separate ANOVAs, one with the main factor 'gender congruency' and one with the main factor 'gender transparency'. Despite the difference in the presentation of the adjective prompt, which could have increased the probability of L1 gender retrieval, no gender interference effect from L1 was observed in this experiment either. In the first ANOVA with 'congruency' as a main factor, only the factor 'length' was significant: $F1(1, 14) = 22.531, p < .001$; $F2(1, 23) = 126.280, p < .001$. The factor 'congruency' was not significant either in F1 or in F2, nor was the interaction between these two factors ($F_s < 1$).

The analyses of errors did not show significant differences between the number of errors in the gender-congruent and gender-incongruent conditions either. The factor 'length' was again highly significant: $F1(1, 14) = 19.443, p < .001$, $F2(1, 23) = 25.703, p < .001$. The interaction between these two factors was not significant ($F < 1$). In the ANOVA with the main factor 'gender transparency', the factor 'length' was again highly significant: $F1(1, 14) = 21.723, p < .001$; $F2(1, 15) = 199.439, p < .001$.

Table 4 Results of Experiment 3 regarding 'gender congruency' with reaction times in milliseconds and error rates with percentages counted for each experimental condition

Condition	Congruent			Incongruent			Interference effect
	RT	SD	Errors	RT	SD	Errors	
Adjective + noun	1078	376	74 (21%)	1069	305	75 (21%)	-9
Noun	867	200	30 (8%)	852	192	31 (9%)	-15

The factor ‘gender transparency’ was also significant both in F1 ($F(2, 28) = 13.820, p < .001$) and in F2 ($F(2, 30) = 5.066, p < .01$). According to a *post hoc* Scheffé test ($p < .05$), Group A differed significantly from both Groups B and C (which were statistically identical) in F1. In F2, Groups A and B differed from each other, but Group C lay between them, differing neither from A nor from B. The interaction between ‘group’ and ‘length’ was significant in F1 ($F(2, 28) = 5.163, p < .05$), and marginally in F2 ($F(2, 30) = 2.137, p = .08$). The separate analyses for the long and short conditions revealed that the effect was significant in both F1 and F2 in the long condition ($F(2, 28) = 10.571, p < .001$; F2: $F(2, 30) = 4.672, p < .05$), but not in the short condition (Scheffé for the long condition: Group A differed significantly from Groups B and C, which did not differ from each other).

The numerical differences between the error rates in the six cells mirror the results of the analyses of reaction times, although statistically the differences were not always significant. The factor ‘gender transparency’ was significant neither in F1 ($F(2, 28) = 1.576, p = .225$) nor in F2 ($F(2, 30) = 1.721, p = .128$), but the Scheffé test revealed that in both analyses Groups B and C differed marginally from Group A ($p = 0.09$). This pattern is also reflected in the number of gender errors alone (2 gender errors in Group A, 33 in Group B and 37 in Group C). The factor ‘length’ was significant, too: $F(1, 14) = 19.443, p < .001$; $F(1, 15) = 24.230, p < .001$. The interaction between these two factors was also significant in F1 ($F(2, 28) = 5.529, p < .01$), but failed to reach significance in F2 ($F(2, 30) = 2.162, p = .124$). Scheffé tests run separately on the long and short conditions yielded significant differences ($p < 0.05$) between Group A and Groups B and C (which were statistically identical) only in the long condition, but not in the short condition. Summarizing, despite the difference in design, the results of Experiment 3 mirror those of the previous

Table 5 Results of Experiment 3 regarding ‘gender transparency’ with reaction times in milliseconds and error rates with percentages counted for each experimental condition

Condition	A (typical)			B (ambiguous)			C (atypical)		
	RT	SD	Errors	RT	SD	Errors	RT	SD	Errors
Adjective + noun	1017	301	34 (14%)	1107	350	56 (23%)	1101	371	59 (25%)
Noun	848	191	21 (9%)	872	211	22 (9%)	857	184	18 (8%)

experiments: We observed no gender interference effect from L1, but again found evidence for the intralingual gender transparency effect.

V General discussion

In three translation experiments, we observed no cross-language gender interference effect, but two of them yielded a gender transparency effect. The L2 gender transparency effect observed in Experiments 2 and 3 is consistent with the results of picture naming experiments reported by Bordag *et al.* (2006). The influence of this factor on gender selection supports the hypothesis that L2 grammatical gender is not stored as a fixed feature, as is assumed for adult L1 gender representation (but see different evidence concerning L1 acquisition: Mills, 1986; Brooks *et al.*, 1993; Taraban and Kempe, 1999; Kempe *et al.*, 2003), but rather computed each time anew, when needed, on the basis of various available pieces of information, for example the phonological form of the word. On the assumption that L2 gender would be stored as a fixed feature, no differences between gender retrieval of gender-typical and atypical/ambiguous nouns would be expected: They would all be stored in a uniform fashion. The fact that the retrieval of nouns with a gender transparent termination is faster (i.e. easier) speaks for a computation of the gender value in which gender transparency plays a role.

The influence of phonological form on gender selection also speaks in favour of models postulating an interaction between the levels of phonological and grammatical encoding (e.g. Dell, 1986). If gender transparency/opacity affects the selection of the appropriate gender node at the grammatical level, then at least two conditions must be satisfied. First, there must be a temporal overlap between grammatical and phonological encoding. If the processes of grammatical feature selection were completed before the beginning of phonological encoding – as defined by the serial, modular model of Levelt – no influence of L2 noun termination on gender selection would be possible. Second, there must be a flow of information from the lower level of phonological encoding to the higher level of grammatical encoding for the same reasons. Overall, the data from both picture naming and translation are in accordance with cascaded models of speech processing (for a detailed argumentation and review of converging evidence, see Bordag *et al.*, 2006).

Considered independently, the absence of the gender interference effect could be interpreted in favour of a model in which the L1 and L2 gender representations do not compete for selection, either because the L1 and L2 systems each have a functionally autonomous gender representation (gender features are not shared across languages), or because gender selection is not a competitive, but rather an automatic process (for a detailed overview of gender access accounts, see Costa *et al.*, 2003).

However, in light of the results of picture naming experiments in which the same items were used and in which the gender interference effect was obtained (Bordag *et al.*, 2006; Bordag and Pechmann, 2007), it seems more likely that the reason for the absence of the interference effect is not due to a representation that excludes the possibility of competition between the L1 and L2 gender nodes during L2 processing, but rather that it is the difference between the tasks (picture naming vs. translation) that is responsible for the seemingly contradictory results.

As mentioned in the introduction, different effects have been observed for picture naming and translation in previous research as well. The usual explanation of these differences is that the system employs a different route for each of the two tasks. Most authors agree that picture naming is conceptually mediated, i.e. in order to name a picture it is necessary to first access that picture's meaning before accessing its name. In a way, naming pictures can be viewed as a type of translation task in which pictorial stimuli are 'translated' into verbal output (Snodgrass, 1993: 87). For translation from L1 into L2, two possible routes have been proposed. The so-called concept mediation proceeds in similar stages to picture naming: L1 word recognition – concept retrieval – L2 word retrieval. On the alternative route, word association proceeds in the following stages: L1 word recognition – L2 word retrieval. The word association route assumes L2 word retrieval via direct links between the L1 words and their L2 translation equivalents. The two routes were first proposed by Potter *et al.* (1984) and correspond to Weinreich's (1953) distinction between compound and subordinative bilingualism, respectively. The later Revised Hierarchical Model (Kroll and Stewart, 1994) attempts to combine these routes into a single developmental model in which the bilingual lexicon develops from a representational structure best captured by word association at low levels of proficiency to a structure best captured by concept mediation at higher proficiency levels. The intuitive

argument for word association (Snodgrass, 1993) is that many language learners (including the participants in our study) are taught vocabulary words in L2 by direct word-to-word translation from L1, which could lead to the development of strong direct word-to-word connections and result in translating via the word association route.

The studies conducted so far do not provide unequivocal evidence concerning the two processing routes relevant for translation. Evidence has been collected in support of both interpretations (for an overview, see Snodgrass, 1993; Dijkstra, 2005), and various factors have been identified that might have an effect on which route is used. The most important factor seems to be L2 proficiency. Kroll and Curley (1988) found evidence that novice bilinguals use the word association link, whereas more proficient bilinguals showed more evidence of the use of the concept mediation route.

As discussed in the introduction, one of the most serious problems of early and recent bilingual models is that they fail to distinguish among levels of representation. This also holds for the concept mediation vs. word association distinction. What exactly is meant by 'word'? Does it comprise both the grammatical and the word form information (fusion of lemmas and phonological forms in Levelt's model), or it is restricted only to the word form information as suggested, for example, by Dijkstra (2005), who interprets word association as follows: 'the Russian word form *kniga* would directly activate the English word form *book* either via a direct orthographic connection (between scripts) or through a phonological connection' (Dijkstra, 2005: 189).

At first sight, the latter interpretation might seem very appealing in our case: Only the L1 word form is activated and from it the activation spreads directly to the word form of its L2 equivalent. During this process, L1 grammatical features like gender are not activated at all and thus do not compete for selection with their L2 counterparts.

However, a closer look at the translation of syntactic units – like complex NPs or even just nouns in their non-canonical naming form – reveals that such a mechanism is hardly viable. Let us assume that an L1 plural form (*dogs*) is to be translated into L2 (*Hunde*). In order to retrieve the appropriate plural form, the plural gender value (represented at the grammatical level) must be transferred as well, so that the plural ending *-s* can be encoded at the phonological level. In our Experiment 3,

participants translated noun phrases with adjectives inflected for gender (and actually also for number and case). How could the correctly inflected L2 word forms be retrieved (irrespective of whether they are stored as whole units, or, more likely, are constructed online) if their syntactic specifications were not activated first? Moreover, the task the participants performed in Experiments 1 and 2 surely involved concept activation of at least the adjective, which was not presented linguistically, but iconically. The inflection of the adjective could not be encoded at all if the grammatical properties of the L2 noun that controls the adjective had not been specified first. Some of these values (e.g. number) are transferred from the L1 noun, others (e.g. gender) are language specific, but their values at the grammatical level must be specified so that the appropriate forms of the L2 nouns and their agreeing elements can be encoded at the form level. That means that the bundle of grammatical properties of the L2 target noun is a *sine qua non* condition for the selection of the target (inflected) word forms. Translation processes of inflected forms thus cannot be reduced to the involvement of the phonological level only.

These considerations bring us to the conclusion that at least in situations where grammatical processing is involved, retrieval of at least some L1 grammatical features is obligatory in translation, so that they can be encoded in L2 (e.g. number), and that the two languages involved in translation must interact at a higher level than the level of word forms, because the L1 syntactic specifications are necessary for appropriate L2 inflected forms to be encoded. The highest candidate level for cross-language interaction would thus be either the level of grammatical encoding (with semantically and grammatically specified lemmas) or the conceptual level, as suggested by the concept mediation hypothesis (this does not exclude interaction at the lower levels as well, i.e. the cognate effect).

If we cannot explain the absence of the gender interference effect in translation with the version of the word association model in which L1 lemmas do not get activated at all (Dijkstra, 2005), a different explanation must be pursued. A very plausible account is captured by a hypothesis concerning the time course of the activation of the gender features in L1 and L2 in translation vs. picture naming.

In picture naming, where the gender interference effect was observed, it is interpreted as follows: The activation spreads from the concept

common to both L1 and L2 lemmas in parallel to the level of grammatical encoding. Consequently, the L1 and L2 gender nodes are activated at the same time (or just one gender node, if the nouns have congruent genders) and can compete for selection.

In forward translation, first the L1 word forms and their lemmas are activated. Only then does the activation spread to the lemma of the L2 translation equivalent (if the lemma level is the level of cross-language interaction). The L2 gender node thus becomes activated only a short time after the L1 gender node. If the level of cross-language interaction in translation is the conceptual level, then the units are activated in the following order: L1 word form – L1 lemma – concept – L2 lemma – L2 word form. In this case, the L2 gender node is activated with even more delay after the L1 gender node than in the previous case. Crucially, in both scenarios described above, the L2 gender node is activated only after the L1 gender node, which may render any competition for selection between these two nodes ineffective.

Obviously, the explanation of the absence of the L1 gender interference effect in translation as compared to picture naming can only be tentative at the moment and further research of not only gender retrieval, but also of processes during translation in general are necessary to clarify this issue. The hypothesis of Miller and Kroll (2002) that a language-specific cue (e.g. the spelling of the word to be translated) may cause the production to proceed more selectively could also offer an alternative explanation for the absence of the gender interference effect in translation. The results concerning the gender transparency effect are, on the other hand, clear-cut and converge with the data from L2 picture naming studies. They supply further evidence for interaction between the word form level and the level of grammatical encoding and challenge distinct serial models of speech production.

VI References

- Andonova, E., D'Amico, S., Devescovi, A. and Bates, E. 2004: Gender and lexical access in Bulgarian. *Perception and Psychophysics* 66, 496–507.
- Ávila, C., Lambon-Ralph, M., Parcet, M.A., Geffner, D. and González-Darder, J.M. 2001: Implicit word cues facilitate impaired naming performance: Evidence from a case of anomia. *Brain and Language* 79, 185–200.

- Badecker, W., Miozzo, M. and Zanuttini, R.** 1995: The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57, 193–216.
- Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S. and Hernandez, A.** 1995: Gender and lexical access in Italian. *Perception and Psychophysics* 57, 847–62.
- 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 Foster-Cohen, S., editor, *EUROSLA Yearbook 4*. Amsterdam: Benjamins, 203–30.
- 2006: Psycholinguistische Aspekte der Interferenzerscheinungen in der Flexionsmorphologie des Tschechischen als Fremdsprache [Psycholinguistic aspects of the interference phenomena in the inflectional morphology of Czech as an L2]. Hildesheim: G. Olms Verlag.
- Bordag, D. and Pechmann, T.** 2007: Factors influencing L2 gender processing. *Bilingualism: Language and Cognition* 10, 299–314.
- Bordag, D., Opitz, A. and Pechmann, T.** 2006: Gender processing in L1 and L2: the role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32, 1090–1101.
- Brooks, P.-J., Braine, M.-D., Catalano, L., Broda, R.-E. and Sudhalter, V.** 1993: Acquisition of gender-like noun subclasses in an artificial language: the contribution of phonological markers to learning. *Journal of Memory and Language* 32, 76–95.
- Caramazza, A. and Miozzo, M.** 1997: The relation between syntactic and phonological knowledge in lexical access: evidence from the tip-of-the-tongue phenomenon. *Cognition* 64, 309–43.
- Costa, A. and 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, 231–44.
- Costa, A., Caramazza, A. and Sebastian-Galles, N.** 2000: The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26, 1283–96.
- Costa, A., Kovacic, D., Franck, J. and Caramazza, A.** 2003: On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition* 6, 181–200.
- De Bot, K.** 1992: A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics* 13, 1–24.
- 2003: Bilingual speech: from concepts to articulation. In Henrici, G., Königs, F.G. and Zöfgen, E., editors, *Mündliche Produktion in der Fremdsprache [Speech production in L2]* 32. Tübingen: Gunter Narr Verlag, 92–103.

- De Groot, A.M.B.** 1993: Word-type effects in bilingual processing tasks: support for a mixed-representational system. In Schreuder, R. and Weltens, B., editors, *The bilingual lexicon*. Amsterdam: John Benjamins, 27–51.
- Dell, G.S.** 1986: A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283–321.
- Dijkstra, A.F.J.** 2005: Bilingual visual word recognition and lexical access. In Kroll, J.F. and De Groot, A.M.B., editors, *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press, 178–201.
- Francis, W.S.** 1999: Cognitive integration of language and memory in bilinguals: semantic representations. *Psychological Bulletin* 125, 193–222.
- Green, D.W.** 1993: Towards a model of L2 comprehension and production. In Schreuder, R. and Weltens, B., editors, *The bilingual lexicon*. Amsterdam: John Benjamins, 249–77.
- Grosjean, F.** 1998: Transfer and language mode. *Bilingualism: Language and Cognition* 1, 175–76.
- Hagoort, P.** 2003: Mysteries of the bilingual mind. Paper presented at the Language Learning Round Table: The Cognitive Neuroscience of Second Language Acquisition, Edinburgh.
- Hermans, D., Bongaerts, T., De Bot, K. and Schreuder, R.** 1998: Producing words in a foreign language: can speakers prevent interference from their first language? *Bilingualism: Language and Cognition* 1, 213–29.
- Janssen, N.** 1999: Bilingual word production: the time course of lexical activation in a mixed language context. Unpublished Master's thesis, University of Nijmegen, The Netherlands.
- Kempe, V., Brooks, P., Mironova, N. and Fedorova, O.** 2003: Diminutivisation supports gender acquisition in Russian children. *Journal of Child Language* 30, 1–15.
- Köpcke, K.M. and Zubin, D.** 1983: Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache [Cognitive organization of gender assignment of one-syllable nouns in German]. *Zeitschrift für Germanistische Linguistik* 11, 166–82.
- Kroll, J.F.** 1993: Accessing conceptual representations for words in a second language. In Schreuder, R. and Weltens, B., editors, *The bilingual lexicon*. Amsterdam: John Benjamins, 53–81.
- Kroll, J.F. and Curley, J.** 1988: Lexical memory in novice bilinguals: the role of concepts in retrieving second language words. In Gruneberg, M.M., Morris, P.E. and Sykes, R.N., editors, *Practical aspects of memory: current research and issues, volume 2*. London: Wiley, 389–95.
- Kroll, J.F. and De Groot, A.M.B.** 2005: *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press.
- Kroll, J.F. and Stewart, E.** 1994: Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33, 149–74.

- Kroll, J.F. and Tokowicz, N.** 2005: Models of bilingual representation and processing: looking back and to the future. In Kroll, J.F. and De Groot, A.M.B., editors, *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press, 531–53.
- Kroll, J.F., Dijkstra, A.F.J., Janssen, N. and Schriefers, H.J.** 2000: Selecting the language in which to speak: experiments on lexical access in speech production. Paper presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans, LA.
- Kroll, J.F., Michael, E. and Sankaranarayanan, A.** 1998: A model of bilingual representation and its implications for second language acquisition. In Healy, A.F. and Bourne, L.E., editors, *Foreign language learning: psycholinguistic experiments on training and retention*. Mahwah, NJ: Lawrence Erlbaum, 365–95.
- Lemhöfer, K., Schriefers, H.J. and Spalek, K.** 2003: Cross-language effects of grammatical gender on visual word recognition in bilinguals. Poster presented at the 13th Conference of the European Society for Cognitive Psychology, Granada.
- Levelt, W.J.M.** 1989: *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M., Roelofs, A.P.A. and Meyer, A.S.** 1999: A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–75.
- Miller, N.A. and Kroll, J.F.** 2002: Stroop effects in bilingual translation. *Memory and Cognition* 30, 614–28.
- Mills, A.** 1986: *The acquisition of gender: a study of English and German*. Berlin: Springer.
- Potter, M.C., So, K.-F., von Eckardt, B. and Feldman, L.** 1984: Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior* 23, 23–38.
- Poulishse, N. and Bongaerts, T.** 1994: First language use in second language production. *Applied Linguistics* 15, 36–57.
- Roelofs, A.P.A.** 2003: Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and Cognitive Processes* 18, 175–204.
- Schriefers, H. and Jescheniak, J.D.** 1999: Representation and processing of grammatical gender in language production: a review. *Journal of Psycholinguistic Research* 28, 575–600.
- Snodgrass, J.G.** 1993: Translating vs. picture naming: similarities and differences. In Schreuder, R. and Weltens, B., editors, *The bilingual lexicon*. Amsterdam: John Benjamins, 83–114.
- Taraban, R. and Kempe, V.** 1999: Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics* 20, 119–48.
- Vigliocco G, Antonioni, T. and Garrett, M.F.** 1997: Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8, 314–7.
- Weinreich, U.** 1953: *Language in contact*. The Hague: Mouton.

Appendix 1 Information about participants

L2	Experiment 1		Experiment 2		Experiment 3	
	Mean	SD	Mean	SD	Mean	SD
Age	24.6	3.1	23.2	2.8	25.0	3.1
Length of acquisition (in years)	10.2	3.1	11.5	2.8	9.8	2.8
Onset of acquisition	12.4	2.3	11.8	2.1	12.7	2.0
Self evaluation (1–5, 1 best)	1.8	0.7	1.6	0.6	1.8	0.8
Evaluation of the experimenter (1–5, 1 best)	1.5	0.7	1.3	0.7	1.5	0.6
Familiarity with items (7 high – 1 low)	6.7	0.2	6.7	0.1	6.8	0.2
Picture–word correspondence (7 high – 1 low)	6.8	0.2	6.7	0.2	6.8	0.2
Number of errors in the experiment	16.6 (min 10, max 24)	4.4	18.7 (min 12, max 25)	4.3	14 (min 7, max 20)	4.3

Appendix 2 List of experimental items used in Experiment 1

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	

Note: m (masculine), f (feminine), n (neuter) after each item indicates its grammatical gender.

Appendix 3 List of experimental items used in Experiments 2 and 3**Table 6** Group A (Gender typical items)

German	Gender*	Czech	Gender (in)congruency	English
Brücke	f	most	incongruent	bridge
Wolke	f	mrak	incongruent	cloud
Glocke	f	zvon	incongruent	bell
Blume	f	kvetina	congruent	flower
Kerze	f	svíčka	congruent	candle
Woche	f	týden	incongruent	week
Schlange	f	had	incongruent	snake
Kirche	f	kostel	incongruent	church
Sonne	f	slunce	incongruent	sun
Waffe	f	zbraň	congruent	weapon
Mütze	f	čepice	congruent	cap
Fliege	f	moucha	congruent	fly
Straße	f	silnice	congruent	street
Katze	f	kočka	congruent	cat
Schule	f	škola	congruent	school
Nase	f	nos	incongruent	nose

Note: *Indicates the gender of the German item.

Table 7 Group B (Gender ambiguous items)

German	Gender*	Czech	Gender (in)congruency	English
Messer	n	nůž	incongruent	knife
Dach	n	střecha	incongruent	roof
Blatt	n	list	incongruent	leaf
Glas	n	sklenice	incongruent	glass
Hemd	n	košile	incongruent	shirt
Fenster	n	okno	congruent	window
Herz	n	srdce	congruent	heart
Buch	n	kniha	incongruent	book
Kopf	m	hlava	incongruent	head
Schritt	m	krok	congruent	step
Zahn	m	zub	congruent	tooth
Topf	m	hrnec	congruent	pot
Schlüssel	m	klíč	congruent	key
Stein	m	kámen	congruent	stone
Stern	m	hvězda	incongruent	star
Brief	m	dopis	congruent	letter

Note: * Indicates the gender of the German item.

Table 8 Group C (Gender atypical items)

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

Note: *Indicates the gender of the German item.