



**HAL**  
open science

## Transfer effects in learning a second language grammatical gender system

Laura Sabourin, Laurie A. Stowe, Ger J. de Haan

► **To cite this version:**

Laura Sabourin, Laurie A. Stowe, Ger J. de Haan. Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 2006, 22 (1), pp.1-29. 10.1191/0267658306sr259oa . hal-00572092

**HAL Id: hal-00572092**

**<https://hal.science/hal-00572092>**

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.

# Transfer effects in learning a second language grammatical gender system

**Laura Sabourin, Laurie A. Stowe and Ger J. de Haan**  
*University of Groningen*

Received November 2003; revised June 2004; accepted September 2004

In this article second language (L2) knowledge of Dutch grammatical gender is investigated. Adult speakers of German, English and a Romance language (French, Italian or Spanish) were investigated to explore the role of transfer in learning the Dutch grammatical gender system. In the first language (L1) systems, German is the most similar to Dutch coming from a historically similar system. The Romance languages have grammatical gender; however, the system is not congruent to the Dutch system. English does not have grammatical gender (although semantic gender is marked in the pronoun system). Experiment 1, a simple gender assignment task, showed that all L2 participants tested could assign the correct gender to Dutch nouns (all L2 groups performing on average above 80%), although having gender in the L1 did correlate with higher accuracy, particularly when the gender systems were very similar. Effects of noun familiarity and a default gender strategy were found for all participants. In Experiment 2 agreement between the noun and the relative pronoun was investigated. In this task a distinct performance hierarchy was found with the German group performing the best (though significantly worse than native speakers), the Romance group performing well above chance (though not as well as the German group), and the English group performing at chance. These results show that L2 acquisition of grammatical gender is affected more by the morphological similarity of gender marking in the L1 and L2 than by the presence of abstract syntactic gender features in the L1.

## **I Introduction**

Adult second language (L2) acquisition is different from first language (L1) acquisition in many ways. One obvious difference is in ultimate attainment between L1 and L2 speakers. It is rare that adult L2 learners – no matter how much exposure they have had to the L2 – acquire perfect native-like competence in the L2. An L2 speaker may behave like a native speaker in some ways, yet, in other ways, never reach the same proficiency as natives. Depending on the language being learned, some L2 constructions are learnable while others seem not to be. This apparently depends partly on the structure of the L1 since it seems that transfer from the L1 can help in L2 acquisition in some cases. An example of an L2 phenomenon that is particularly difficult to learn is grammatical gender (also called noun class), which is a lexical property of nouns. The adult L2 acquisition of grammatical gender and the influence of transfer from L1 on this process is the focus of the research presented here.

There are three logical possibilities for the degree of transfer in L2 learning: no transfer, partial transfer or full transfer. In the case of no transfer there should be no effects of the L1 on the L2. Partial transfer refers to the idea that in at least the initial state of learning the L2, some L1 properties are carried over into the L2 grammar. Full transfer is said to occur when properties of the L1 determine the entire L2 grammar, at least initially. For both the full and partial transfer views, the role of transfer may be different at different stages of acquisition. The no transfer position predicts that no differences will occur between L2 learners from different L1s. As it is generally accepted that there are at least some effects of L1 on the learning of the L2 grammar (White, 1985; Vainikka and Young-Scholten, 1996; Hawkins and Chan, 1997), this position of no transfer is not supported and will not be considered further. Both the partial transfer and full transfer positions predict that differences will be found between L2 learners from different L1s at some stage of acquisition. Full transfer claims that all aspects from the L1 are taken over into the L2 grammar, at least in the initial state. Partial transfer represents the case where only part of the L1 grammar is used in the L2 grammar. In the literature, proponents of partial transfer are not in agreement over what part of the L1 is transferred and what is not transferred.

Besides the degree of language transfer there are also two types of transfer that have been proposed. One type is the transfer of surface features from one language to another (surface transfer). This might include such things as the transfer of surface word order between languages or the transfer of morphologically similar gender marking. German, having a gender system morphologically similar to the Dutch system, may qualify for this type of transfer. The other type is the transfer of more abstract features of language (also known as deep transfer). This could involve the transfer of abstract syntactic categories that exist in both languages, but which do not have similar morphological exponents, e.g. the transfer by Romance speakers of their gender category to the learning of the Dutch gender system.

There has been a lot of research on both the comprehension and production of L2 grammatical gender (Andersen, 1984; Rogers, 1984; 1987; Finneman, 1992; Shelton, 1996; Myles, 1995; Hawkins, 1998; Dewaele and Véronique, 2000; Franceschina, 2001; 2002; White *et al.*, 2004). These studies have shown that L2 gender errors are frequent, that overgeneralization to one form occurs, that accuracy depends on the actual amount of use of the L2 and not on the amount of classroom exposure and that gender agreement seems to be more difficult when the agreeing element is structurally more distant from the noun it has to agree with.

Relatively less research has, however, been done on L1 effects on gender acquisition. Specifically relevant to the current study is the research by Franceschina (2001; 2002) and White *et al.* (2004), where the effect of L1 on the learning of L2 gender systems was investigated. Also relevant is the study by Sabourin and Haverkort (2003), which investigates L2 ability with two different kinds of gender agreement constructions. While it seems to be the case that gender is difficult for L2 speakers, this may be more due to the fact that a lot of studies look at native speakers of English (a language with no grammatical gender) learning languages with grammatical gender. Different findings may be expected if L2 learners have grammatical gender in their L1. In fact, Franceschina, in an elicitation task, showed that while English speakers with a high proficiency in Spanish show persistent problems learning the Spanish gender system, Italian learners (with a gender system very similar to Spanish) do not. She claims that this is because speakers with

#### 4 *Transfer effects and grammatical gender systems*

English L1 have no underlying gender feature in their L1 grammars to map the Spanish gender feature onto. Italian speakers do have such an underlying feature in their L1 grammar.

Franceschina's findings support the failed functional features hypothesis (Hawkins and Chan, 1997) that states that learners are not able to acquire grammatical features that are not present in their L1. In another study looking at the production and comprehension of L2 gender (White *et al.*, 2004), it was found that both French and English learners of Spanish perform fairly accurately with the L2 gender system.<sup>1</sup> They conclude – based on the fact that no differences between the English and French groups were found – that presence or absence of gender in L1 is irrelevant. They further conclude that these findings support a full transfer/full access model of SLA (Schwartz and Sprouse, 1996), which states that second language learners have full access to the abilities to learn that L1 learners do (Universal Grammar; UG). This finding contradicts the findings of Franceschina (2001; 2002) where English speakers perform very poorly. These contradictory findings need to be investigated further. The current study tries to add to this body of literature in the hope of elucidating what may be going on in the above studies.

In a study investigating L2 ability with grammatical gender, Sabourin and Haverkort (2003) found that even if gender is similar across the L1 and the L2, some constructions may still cause problems for the L2 learner. Sabourin and Haverkort looked at German learners of Dutch on the acquisition of the Dutch noun phrase (NP). They found that the German participants could attain a native-like level only when the NP was definite. For indefinite NPs the German group did not perform well at all. It was suggested that the German group were able to use surface transfer to acquire the definite NPs, but for the indefinite NPs where the constructions in Dutch and German are less similar, they could not transfer the category gender to help them in the L2 process.

In the above studies the L1 of the L2 Spanish and Dutch participants was either English (a language with no grammatical gender) or a

<sup>1</sup>In this study many of the English participants had learned French at a young age in school and this may be skewing the results. White *et al.* do address this issue but the actual scores comparing English speakers with exposure to French and those without are not provided.

language with a gender system similar to the language being learned (French or Italian speakers learning Spanish, or German speakers learning Dutch). Thus, even in the study by Franceschina (2001) where a difference was found depending on whether the category gender is in the L1, it is not possible to determine if just having abstract grammatical gender in the L1 is enough to acquire an L2 gender system (deep transfer), or if the L1 and L2 must have similar morphological exponents (surface transfer). In the current study, looking at the adult L2 acquisition of Dutch, participants were selected from three different L1 backgrounds: German (with an abstract grammatical gender category and morphological exponents similar to Dutch), English (with no grammatical gender system) and Romance (languages with an abstract grammatical gender category, but whose exponents are quite different from Dutch).

The Dutch gender system currently has two gender categories: common and neuter. This system developed from an earlier three-gender system: the earlier masculine and feminine genders have collapsed into one common gender.<sup>2</sup> This is reflected in the fact that about two-thirds of nouns belong to the common gender class, while only one third belong to the neuter gender class. Evidence of the gender distinction is seen on the determiner: common gender nouns select the definite determiner *de*, and neuter gender nouns select the definite determiner *het*. Gender differentiation is also seen on relative pronouns where common gender items select the pronoun *die* and neuter gender items select *dat*. In L1 acquisition of Dutch the gender system does not seem to be fully in place even by age 3 years and 4 months (Gillis and de Houwer, 1998). Overgeneralizing to one of the genders also occurs in Dutch: Dutch children always choose the common gender determiner as the default.

The three different L1 language groups tested were chosen for the nature of grammatical gender (or lack thereof) in their L1. English has no grammatical gender and thus provides information on how participants without a gender distinction in their L1 handle such

<sup>2</sup>It must be noted, however, that a three-way gender distinction can be found in agreement with some pronouns. This distinction is used mostly in written language. The development of the Dutch gender system from three to two genders is discussed in van Hout (1996) and van Leuvensteijn *et al.* (1997).

distinctions in a language they are learning.<sup>3</sup> The Romance language group provides information about whether abstract properties of grammatical gender transfer into interlanguage grammars. Romance languages have a 2-gender system like Dutch but the systems show no correlation between the gender assignment to particular nouns/concepts in each language. Moreover, the agreement patterns in the Romance languages are different from those in Dutch. Finally, the German system, although it differs from the Dutch system in that it contains three genders, is similar in terms of assignment and agreement of gender; most German nouns and their Dutch equivalents have the same gender (masculine and feminine in German are mapped onto common in Dutch and German neuter is mapped onto Dutch neuter), and the elements that must agree with the noun's gender are very similar in the two languages. By looking at L2 learners of Dutch from different L1 backgrounds it may also be possible to tease apart different types of transfer. Surface transfer is represented by the direct transfer of morphologically similar gender realization between the L1 and L2. This type of transfer would only occur for the German group whose gender system is congruent to the Dutch system. Deep transfer, on the other hand, would be the transfer of the category gender (whether it is a congruent system or not) from the L1 to the L2.

It is predicted that after controlling for the level of L2 syntactic proficiency in a non-gender domain (for specific details, see Sabourin, 2003), an effect of L1 (agreeing with the results found by Franceschina) will be seen with the German group performing the best due to transfer of gender assignment properties from the L1 (surface transfer). It is also predicted that simply having the category gender will help (deep transfer), thus the Romance group should perform better than the English group, but that simply having gender, unless it has similar exponents, is not enough to acquire the L2 system as readily as when direct surface transfer is possible. The English group are expected to perform worst, since neither abstract nor concrete knowledge can be transferred from L1. The issue of deep vs. surface transfer also lends

<sup>3</sup>English marks gender in its pronominal system on the basis of semantic criteria (Corbett, 1991: 18). English does not, however, have gender-specified nouns or a system of gender concord between nouns and other elements in the extended nominal projection. Only the latter is relevant to the present study.

itself to a brief investigation of the role of UG in the SLA of grammatical gender. Support for deep transfer would suggest that UG is still available during SLA while support for only surface transfer may suggest that UG is not available in SLA, with a more general cognitive strategy of translation operating in this domain in SLA. We conclude with a discussion of whether L1 transfer reflects a surface effect at the level of explicit knowledge or whether the transfer is applicable to on-line processing.

Besides looking at the effects of L1 in the experiments presented here, both L2 knowledge of the Dutch gender that must be assigned to a given noun (gender assignment: Experiment 1) and L2 knowledge of the grammatical rules dictating Dutch gender agreement between the noun and the relative pronoun (gender agreement: Experiment 2) are discussed. This knowledge is tested by the use of off-line tasks, so that possible limitations on cognitive capacity that might arise in oral production should not interfere with the results. The participants are allowed to see the entire sentence together, which should decrease working memory load. Participants can look back at earlier parts of the sentences to verify items used. Investigating the knowledge level of L2 participants is very important in determining whether L2 gender distinctions and agreement can be learned and whether there is a difference between assignment and agreement.

## **II Experiment 1**

This first experiment looks at whether the L2 participants can assign the correct gender to a list of nouns. Do three different L2 groups know the gender of nouns and is there a difference to be found depending on their L1?

The aim of the experiment is to show whether there are transfer effects of an abstract category 'gender'. If so, does transfer depend simply on the existence of a category gender in the L1 (in this case both the German and Romance speakers but not the English speakers are expected to show positive transfer effects), or does it require not only that the category exist in the L1 but that its morphophonological realization is similar in the L1 and the L2 (in this case only the German speakers should show positive transfer effects). It must be noted,



however, that if no effects of L1 are found in the current study, which focuses on learners who have achieved a reasonable level of proficiency, that this does not provide evidence for overall absence of language transfer in SLA. It is possible that transfer effects exist at an earlier stage of L2 development, but that eventually these effects are neutralized. Similarity in performance of speakers from different L1s at advanced levels of L2 proficiency is consistent with both initial L1 transfer and with no initial L1 transfer (White, 2000).

The effect of familiarity on the knowledge of lexical gender is also investigated by comparing nouns of different frequencies. Nouns were selected from the CELEX database (Burnage, 1990); half the nouns are high frequency nouns while the other half are considered to be middle frequency. This is important: the higher the frequency of a noun, the more experience and familiarity a speaker should have with it, which may result in better learning of both the item and the types of information, such as gender, that are linked to it. We will also be able to determine if participants make use of a default gender in this task. In Dutch, the common gender occurs two thirds of the time and thus if participants are sensitive to this fact they may choose to assign the common gender to more nouns.

### *1 Method*

*a Participants:* In this experiment 70 adult L2 speakers of Dutch were tested. There were 25 participants with German as their L1, 21 with a Romance language as their L1 and 24 with English as their L1. As the grammatical aspect being investigated is gender, no participants were included who had learned (as a child) a second language that could be considered closer to the Dutch gender system than their native language.<sup>4</sup> For example, as the German system is very close to the Dutch system, it did not matter if they had learned another language with gender as a child; on the other hand, Romance speakers were excluded from participation if they had learned a Germanic gender

<sup>4</sup>Any knowledge of other languages that participants had was noted. Although most English participants did not have any other languages, it was often the case that German and Romance participants had learned English. Because of the impossibility of finding L2 speakers of Dutch with no knowledge of English this information was noted but participants were not excluded if they had knowledge of English.

language as a child and English participants were excluded if they had learned any gender language as a child.

In view of the goal of testing only relatively advanced L2 speakers, participants were required to have a high level of proficiency. Only participants living in the Netherlands for at least three years were considered.<sup>5</sup> Another inclusion criterion was that participants needed to show a high level of accuracy on a proficiency task which concentrated on agreement phenomena within the verb phrase. This was done to have a measure independent of grammatical gender, while still testing agreement phenomena. No L2 group showed significant differences in the proficiency scores when compared to a native speaker group. Also, all the L1s (of the L2 participants) have relatively similar systems in this domain, so that the effect of L1 is not expected to be large. Only participants who obtained higher than 90% on this test were included in the current experiments (for details on this test, see Sabourin, 2001; 2003). Information on the tested participants as well as information on the proficiency testing scores can be seen in Table 1.

*b Materials:* In total, 160 nouns were tested in this experiment. In order to test gender assignment knowledge fairly, nouns were chosen that all L2 participants were expected to know. In order to investigate the effect of familiarity of the knowledge of the gender of nouns, half of the items were of high frequency, while the other half were middle frequency: high frequency log between 1.96 and 2.98 (average 2.28); middle frequency log between 1.11 and 1.49 (average 1.31) according

**Table 1** Participant information: the number of participants included for each language group along with amount of exposure and their proficiency scores

First language	Exposure to Dutch	Proficiency score
German ( <i>N</i> = 25)	Range: 2–49 yrs Average: 11.6 yrs	Range: 92–100% Average: 97%
Romance ( <i>N</i> = 21)	Range: 3–33 yrs Average: 13.6 yrs	Range: 91–100% Average: 96%
English ( <i>N</i> = 24)	Range: 2.5–50 yrs Average: 14.4 yrs	Range: 90–100% Average: 96%

<sup>5</sup>There were two exceptions: one German and one English participant each had less than 3 years exposure to Dutch. The German participant had started learning Dutch while still in Germany and this time was not included in the calculation of length of exposure to Dutch. The English participant obtained one of the highest overall scores and thus the decision to keep this participant in was taken.

to the CELEX database (Burnage, 1990). The middle frequency items were still of reasonably high frequency in order to increase the chance that L2 participants would recognize these items. Although all items were recognizable, there is probably still an effect of frequency as participants would have been exposed less to the middle frequency items.) Half of the nouns are of common gender (*de* items) and the other half are of neuter gender (*het* items).

*c Procedure:* This experiment was carried out as the second section of a test sequence that also contained Experiment 2. Participants were asked to make a *de* or *het* judgement for each noun, even if they were not sure of the gender. An exception was made for words the participants did not know. If participants had never heard the words before, they were asked to circle the item (this was never more than 5 nouns for any participant). This was done to ensure that all items analysed were recognizable to the participants. These items were coded neither as correct nor incorrect but as unanswered (thus they did not factor into the averaging of the score). Nouns were presented in a random order. Two different presentation orders were given.

## 2 Results

A 3-way analysis of variance (ANOVA) was performed with item frequency (high and medium) and item gender (common and neuter) as the within-participants factors and L1 (German, Romance and English) as the between-participants factor. Main effects of frequency ( $F(1,67) = 186.59, p < .001$ ), gender ( $F(1,67) = 66.58, p < .001$ ) and L1 ( $F(2,67) = 21.47, p < .001$ ) were found. In Table 2 the average

**Table 2** Average accuracy (and range) for each language group by item type

	High frequency		Medium frequency	
	<i>de</i> items	<i>het</i> items	<i>de</i> items	<i>het</i> items
German	Average: 99.3% (95–100%)	Average: 96% (87–100%)	Average: 96% (90–100%)	Average: 90.1% (76–100%)
Romance	Average: 94.8% (90–100%)	Average: 92.8% (72–100%)	Average: 93% (50–100%)	Average: 76.5% (52–98%)
English	Average: 86.8% (65–100%)	Average: 86.5% (62–98%)	Average: 88.6% (70–98%)	Average: 69.8% (45–93%)

accuracy rates and the range for each category can be found. In general, the findings show that high frequency items were responded to more accurately than middle frequency items, and that the common gender items were easier than the neuter gender items. One exception to this is the fact that the English group showed their best score for the low frequency *de* items.

All 2-way interactions are significant. The interaction of frequency and L1 ( $F(2,67) = 6.28, p = .003$ ) shows that each L1 group performed better on high frequency items, but that a larger difference between high and middle frequency is seen for the Romance speakers. There was also a significant interaction between L1 and item gender ( $F(2,67) = 3.76, p = .028$ ). Here there was a smaller difference between common and neuter gender for the German speakers than for the other groups. The significant interaction found between frequency and gender type ( $F(1,67) = 91.74, p < .001$ ) clearly shows that, while for high frequency items there is only a small advantage for common gender items, at middle frequency the common gender items have a much higher accuracy rate. The 3-way interaction between frequency, gender type and L1 was also significant ( $F(2,67) = 15.40, p < .001$ ). This interaction is depicted graphically in Figure 1. Figure 1 illustrates that the interaction of L1 with gender type is mainly due to the middle frequency condition.

### 3 Discussion

The first striking result is that all groups performed on average above an 80% accuracy rate when scores are collapsed across all four

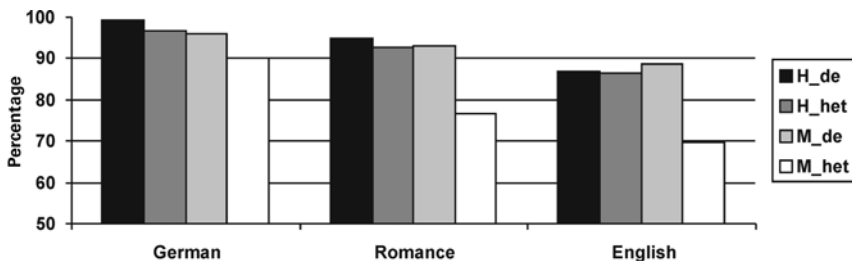


Figure 1 Accuracy scores for Experiment 1

categories, which is well above chance level. Clearly, L2 speakers can assign gender correctly to nouns. Thus, learning gender at the lexical level does seem to be possible for all L1 groups. It must be noted, however, that for each category there is quite a large spread of scores (especially for the English and Romance groups). It is clear from this experiment that transfer is not necessary to learn to assign the correct gender to a word, since even the English group who have no gender in their native language can assign gender well above chance level. However, transfer does appear to play a role as seen in the significant effect of L1. It appears that both surface and deep transfer can give an advantage to the learner, since both Romance and German L1 speakers do better than the English group. However, surface transfer can be even more useful, as seen in the better performance of the German group, particularly on middle frequency items.

An effect of familiarity was found for all groups with better scores for the nouns that the participants have more exposure to (the higher frequency nouns). The German group did very well overall with an average score of above 90% in all categories. This was not very surprising since there is a high amount of congruency between the Dutch and German systems.<sup>6</sup> Although the German participants performed very well overall, they performed relatively poorly on the middle frequency neuter gender items (90.1%). In all other categories the average was above 96% correct. The Romance participants also did very well: they had an almost 90% average accuracy rate, however, they did perform worse than the German group. Comparable to the German participants, the Romance participants also obtained their worst scores on the middle frequency neuter items (76.5%). Although the English group, as expected, performed the worst (with a 83% average), they still performed quite respectably on the high frequency nouns of both genders. The English group followed the same pattern as the other L2 groups by obtaining the worst score on the middle frequency neuter items (69.8%).

<sup>6</sup>This congruency between the German and Dutch system was investigated. For high frequency items it did not matter whether the items were congruent (the German group performed well), but for the lower frequency items they did better on the congruent items (Sabourin, 2003). This suggests that the German group is making use of a direct translation strategy, which they can overcome with items they are more familiar with (high frequency items).

As mentioned above, all L2 groups have their lowest score on the middle frequency neuter items. While it seems that none of the participants have a lot of difficulty with gender assignment of high frequency nouns of either gender, all groups show an effect of assigning a default gender for the middle frequency items. As approximately two thirds of Dutch nouns are of common gender there is a very easy (and obvious) use of a default gender strategy that can be applied by L2 learners of the Dutch language: if the gender of a noun is not known, then assign the noun to the common gender (*de*). While the application of the default gender appears to occur in all groups, it is clearer for the Romance and English speakers and appears to be present even for high frequency nouns in these groups. The German group, able to apply direct surface transfer, does not need to use this strategy to the same extent. On the less familiar items (the middle frequency items) the English group uses this strategy to the extent that they apparently just assign the default *de* determiner, leading to apparent high accuracy on the *de* words and many mistakes on the *het* words.

The effect of frequency and its interaction with L1 raises some important issues about the nature of learning gender in an L2 and the nature of transfer effects. This difficulty with the middle frequency items could also simply mean that in order to actually 'set' a gender for items, participants need to have enough exposure to the item. Frequency effects do not seem to be nearly as prominent in L1 language learning, where single exposures may be enough to set the gender. Experiment 1 showed that all L2 groups can assign gender to the Dutch nouns (albeit with varying degree). In Experiment 2, the ability to go beyond simple gender assignment and use their knowledge for gender agreement will be investigated.

### III Experiment 2

The goal of this second experiment was to investigate L2 speakers' knowledge of grammatical gender. As opposed to Experiment 1, which explored gender assignment, this experiment looked at gender agreement between the noun and a relative pronoun. In Experiment 1, it was demonstrated that the learners were fairly well able to access and assign a gender to a noun. In the current experiment we asked participants to do an off-line ungrammaticality detection and correction

test to determine whether the L2 learners were able to use this knowledge to process grammatical agreement in sentence context.

In Dutch, the form of the relative pronoun used for singular nouns depends on the gender of the noun being relativized. If the noun is a common gender noun, then the relative pronoun is *die* in both the definite and indefinite cases. In the neuter gender case, the relative pronoun *dat* is used. In the plural, not used in this experiment, the genders are collapsed and *die* is used as the relative pronoun with nouns of both genders. Sentences were presented in which the relative pronoun had the appropriate gender agreement or an inappropriate gender agreement, to determine whether the L2 groups could identify the ungrammaticality.

We chose to manipulate relative pronoun agreement because a study by Myles (1995) suggested that structural distance is an important variable in the use of gender agreement rules. Myles showed that L2 ability with grammatical gender agreement is correlated with the structural distance (defined in terms of embeddedness) between noun and agreeing element: the greater the structural distance between the agreeing element and the noun, the more difficult the task was. In this sense, Experiment 1, on gender assignment, focused on gender agreement with minimal structural distance between the determiner and noun (the nodes are sisters within a phrase) while in the present experiment knowledge of gender agreement must be applied across several nodes, although the physical distance is the same. If the greater structural distance noted for relative pronouns does cause more problems for the L2 learner, then all L2 groups should have more trouble with agreement between noun and relative pronoun than they showed for simple assignment of gender as in Experiment 1.

As in Experiment 1 above, effect of L1, familiarity and of default gender will be investigated. Further, in this experiment, we also examine whether the presence of an overt gender marker has an effect on gender agreement.

### *1 Method*

*a Participants:* The same L2 participants as in Experiment 1 were also tested in the current experiment. Thirty-four native speakers of Dutch were also tested.

*b Materials:* This test contained 80 target sentences, each of which included a critical ‘noun–relative pronoun’ sequence. All sentences are of the restrictive relative clause type with some involving object relativization and some involving subject relativization. In half of these items, the sequence included a definite determiner before the head of the relative clause. In this condition the participants thus always had the correct gender information reinforced by the presence of the determiner and only had to determine whether the accompanying relative pronoun was correct or not. The other half of the items were sentences with an indefinite determiner before the head of the relative clause. Here, no overt gender information is provided with the noun, since the same indefinite form is used with both genders. For all 80 sentences a grammatical and ungrammatical version was made. Examples of these sentences can be seen in Examples 1 and 2 below.

- 1) De baron        die/\*dat        in het kasteel woonde, is overleden. (RP–definite)  
 The baron<sub>com</sub>    that<sub>com/\*neut</sub>    in the castle lived,    has died.  
 The baron that lived in the castle has died.
- 2) Een lichaam    dat/\*die        slap is,        heeft training nodig. (RP–indefinite)  
 A body<sub>neut</sub>        that<sub>neut/\*com</sub>    flabby is,    has    training necessary.  
 A body that is flabby needs training.

The critical nouns used in the sentences of this experiment were 80 of the nouns tested in Experiment 1. Of these 80 nouns 19 were human reference nouns (e.g. historian and painter), which almost always fall into the common gender class. Five nouns began with the prefix *ge-* which usually falls into the neuter gender class and one noun had the suffix *-te* which usually falls into the common gender class. The frequency and gender of the critical items used were manipulated as described above for Experiment 1. This resulted in a 2 (grammatical by ungrammatical) by 2 (definite vs. indefinite) by 2 (high vs. medium frequency) by 2 (common vs. neuter) design with 16 conditions.

Two lists were created to which the grammatical and ungrammatical version of each sentence were assigned according to a Latin square design; each list contained 5 items per condition randomly distributed across the list. Besides the 80 sentences of interest there were also 200 filler sentences.<sup>7</sup> The order of presentation of sentences in the

<sup>7</sup>The filler sentences included the sentences used in the proficiency measure (explained above and presented in Sabourin, 2003) as well as other filler sentences.



grammaticality test was such that in no case were sentences from the same condition presented one after the other. Two more lists were generated by splitting the lists just described in half and reversing the order.

*c Procedure:* Each participant received a test questionnaire consisting of two sections. The first part was the grammaticality judgement task, in which the participants were first asked to go through the test making a yes/no decision as to the grammaticality of each sentence. They were asked to complete this task in approximately 30 minutes. If they were not yet done after the 30 minutes, they were asked to mark the sentence they were at and then continue with the test. This was done to check that all participants completed the task in approximately the same amount of time.

After judging the grammaticality of all the sentences, they were asked to go back to the beginning and for every sentence they had marked as ungrammatical, they were asked to make a correction so that the sentence would become grammatical. This was done to ensure that participants had judged a sentence as ungrammatical for the correct reason and not, for instance, because they felt that an incorrect preposition or incorrect word order had been used.

The grammaticality judgement responses were only considered as correct answers if they were both correctly judged as grammatical or ungrammatical and if the appropriate correction was made. For example, in the ungrammatical version of the sentence in Example 1 above, it may be the case that a participant correctly judges the sentence as ungrammatical, but, in the correction of the sentence only changed the tense of the verb. If this was the case, the sentence was scored as incorrect. Similarly, if the sentence was supposed to be marked as grammatical, but the participant rated the sentence as ungrammatical, and then made a correction that was unrelated to the condition being looked at, the sentence was scored as correct. For example, if one of the above sentences had been in its grammatical form but the participant still said it was ungrammatical due to the tense of the verb, then the error would have been disregarded in the scoring. This was done so that all participants were analysed on the same error types, those being the error types of interest, gender errors.

## 2 Results

Because of the presence of a large number of multi-factorial interactions in the analysis including all factors, the responses for this experiment will be discussed as analysed in two 3-way ANOVAs with definite vs. indefinite and common vs. neuter gender as within-participant factors and L1 group (native, German, Romance and English) as between participants factor. Grammatical items are not analysed here as a strong ‘yes’-bias effect was found (participants’ tendency to respond ‘yes’, that items were grammatical, if they were unsure of the answer) and thus scores on grammatical items do not accurately reflect the participants’ knowledge of grammatical agreement (for more information on the ‘yes’-bias, see Sabourin, 2003). One ANOVA focused on the high frequency items while the other looked at the middle frequency items.<sup>8</sup> As pointed out in the *Participants* section above, the same participants took part in this experiment as in Experiment 1. Correlation analyses were also performed to compare participants’ scores in this experiment to their scores from Experiment 1.

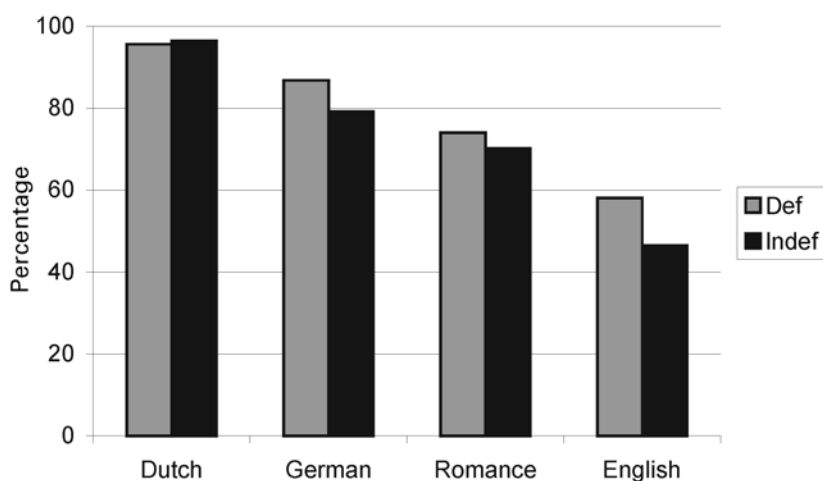
*a High frequency items:* In this first 3-way ANOVA looking at the responses to ungrammatical items for the high - frequency items, the main effects of definiteness ( $F(1,100) = 14.28, p < .001$ ), gender ( $F(1,100) = 4.35, p = .04$ ) and L1 ( $F(3,100) = 21.43, p < .001$ ) were all significant. The definite items, with gender explicitly marked in the determiner, were responded to more accurately than the indefinite items (79% vs. 73%). Common gender items were responded to more accurately than the neuter gender items (78% vs. 74%). Using pairwise comparisons to analyse the main effect of L1, it was found that the native speakers performed significantly better than all L2 groups. Furthermore the English group performed worse than both the German (mean difference = .308,  $p < .001$ ) and Romance (mean

<sup>8</sup>It must be noted that the native speakers did not perform at 100% accuracy. The errors made by native speakers were incorrectly saying an ungrammatical sentence was correct. These errors likely just represent mistakes, possibly due to unconscious correction. When native speakers were asked about the incorrect sentences afterwards, they all made the correct judgment and just said they read over the error.

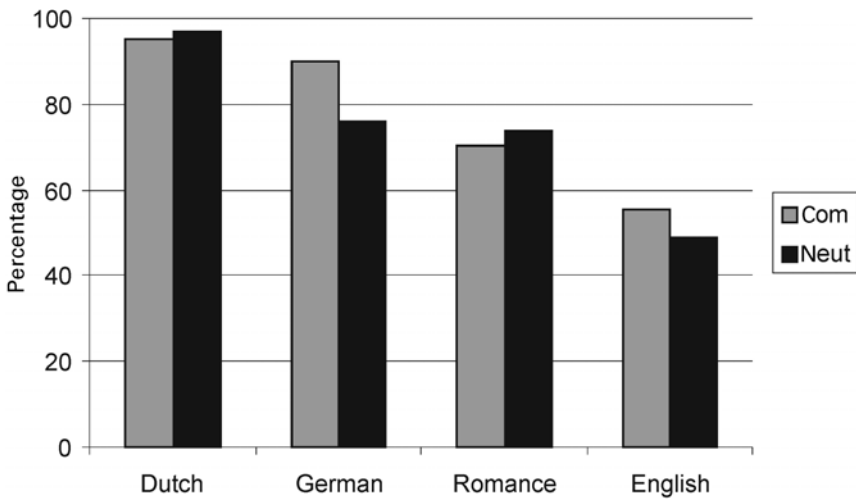
difference = .199,  $p = .002$ ) groups. The difference between the German and Romance groups was not significant (mean difference = .109,  $p = .082$ ), although a trend towards an effect can be seen.

In total three significant interactions were found for this analysis. There was a significant interaction between definiteness and L1 ( $F(3,100) = 3.7, p = .014$ ). This interaction can be seen in Figure 2. As can be seen in this figure, although all L2 participants perform better on definite items, this difference is greatest for the English group. The 2-way interaction between gender and L1 was also significant ( $F(3,100) = 4.48, p = .005$ ). This interaction is depicted graphically in Figure 3. In this interaction, the Germans show the largest difference between common and neuter gender items, with a very high score on the common gender items. A significant 2-way interaction was also found between definiteness and gender ( $F(1,100) = 8.26, p = .005$ ). This interaction is depicted in Figure 4. Here it can be seen that, while gender makes no difference on the performance of definite items, once indefinite items are considered, the participants, on average, perform better on common gender items.

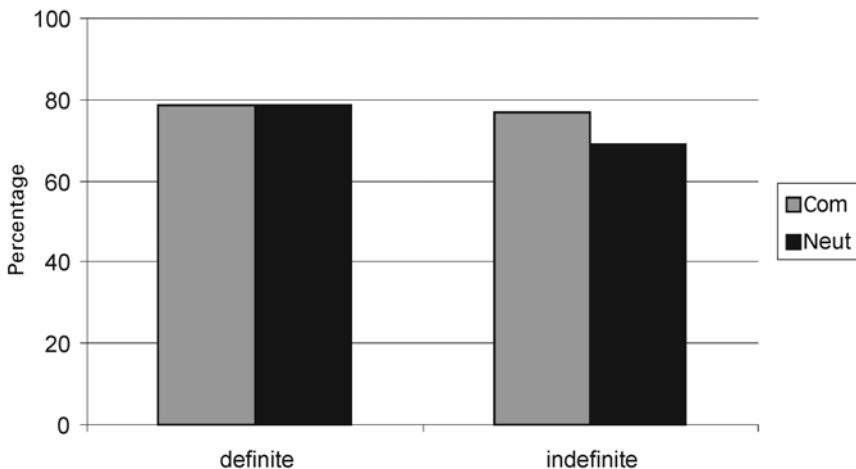
*b Middle frequency items:* In the second ANOVA, looking at the middle frequency ungrammatical items, the same effects as above were



**Figure 2** Interaction between definiteness and L1 for the ungrammatical high frequency items

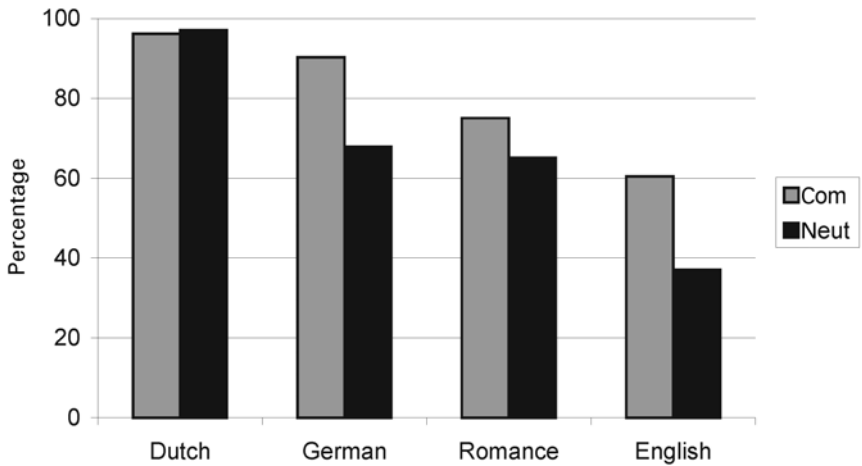


**Figure 3** Interaction between gender and L1 for the high frequency grammatical items



**Figure 4** Interaction between gender and definiteness for the high frequency items

analysed. The main effects of definiteness ( $F(1,100) = 9.8, p = .002$ ), gender ( $F(1,100) = 43.7, p < .001$ ) and L1 ( $F(3,100) = 27.75, p < .001$ ) were all significant. As with the high frequency items, the definite items, with gender explicitly marked in the determiner, were responded to more accurately than the indefinite items (77% vs. 71%). Common gender items were responded to more accurately than the



**Figure 5** Interaction between gender and L1 for the middle frequency items

neuter gender items (81% vs. 67%). Only one 2-way interaction reached significance; this was the interaction between gender and L1 ( $F(3,100) = 8.88, p < .001$ ). This effect can be seen in Figure 5. This significant interaction is best explained by the fact that unlike the native speakers, all L2 groups performed much better on the common gender items. It is especially important to note that in a comparison of the high and middle frequency items, the English group performed better on the middle frequency common items than on the higher frequency common items. Both the German and Romance groups performed better on the higher frequency items. This will be discussed in the conclusions section below.

*c Correlation analyses:* Correlation analyses are presented to investigate how closely participants' use of gender in detecting gender agreement errors agrees with their ability to retrieve gender information (gender assignment). In overall correlation analyses, accuracy rates on both the scores on high frequency items ( $r = .74, p < .001$ ) and scores on the middle frequency items ( $r = .76, p < .001$ ) correlated with the gender assignment scores. Correlations were also performed for each L2 group separately. For the German group, accuracy on the gender assignment task correlated with both accuracy on the high frequency items ( $r = .66, p < .001$ ) and middle frequency items ( $r = .52,$

$p = .007$ ). The same pattern was found for the Romance group; gender assignment correlated with accuracy on the high frequency items ( $r = .56, p = .009$ ) and middle frequency items ( $r = .63, p = .002$ ). The English group also showed this pattern; gender assignment scores correlated with accuracy on the high frequency items ( $r = .47, p = .02$ ) and middle frequency items ( $r = .45, p = .028$ ).

Because all groups showed this same pattern, a separate MANOVA was performed, taking the scores from Experiment 1 as a covariate to see if a significant effect of L1 could still be found. In the overall 5-way MANOVA the main effect of L1 was not significant although it tended toward significance ( $F(3,100) = 2.31, p = .081$ ). Given multiple interactions with L1, MANOVAs were also done looking at high and middle frequency items separately as above. For the high frequency items no significant effect of L1 was found once gender assignment was factored out ( $F(3,100) = 1.07, p = .367$ ), however a significant effect of L1 was found for the middle frequency items even when the difference in gender assignment ability was considered ( $F(3,100) = 3.82, p = .012$ ).

### 3 Discussion

Unlike Experiment 1 where all L2 groups could perform gender assignment, the ability to do gender agreement is clearly dependent upon the L1. Overall the German group performed fairly well, the Romance group also did reasonably well though not as well as the German group, while the English group seems to only perform at chance suggesting that they were unable to perform gender agreement. However, the correlation between assignment and agreement in the scores of the English group does suggest that this inability to perform agreement is not an across the board phenomenon, and that sufficient ability with gender assignment can be applied to gender agreement even when there is no gender in the L1. This needs to be investigated further in order to determine whether the English group can in fact acquire gender agreement given enough exposure to the L2.

One pattern that can be seen in the grammaticality judgment task is the higher accuracy scores for sentences containing higher frequency critical items for all L1 groups. This is the same pattern that was found

in the gender assignment task of Experiment 1. This suggests the presence of a learning gradient: nouns that are more familiar to the participants are more likely to be dealt with grammatically. This gradient is further supported by the fact that when including the scores on gender assignment as a covariant, the effect of L1 is only significant for the middle frequency items. This suggests that for items that participants in all groups are more familiar with and that they know the gender of, it is possible to perform at a native speaker level in terms of gender agreement. This is the case at least in an off-line task, but once items that are less familiar are used, whether or not the learner knows the gender of the item, there is still an effect of L1 and therefore of transfer.

Since only the middle frequency items show an effect of L1 when the score on the assignment test is taken into account, we limit our discussion of transfer to these data. On middle frequency items, little difference was found between the German and Romance speakers; both these groups of participants are better than the English speakers. This lack of difference between the German and Romance group is not due to the Romance speakers' mastering the use of the relative pronoun, but in fact seems to be due to the German speakers' relatively poor performance on gender agreement as opposed to how well they did on gender assignment in Experiment 1. This suggests that surface transfer is relatively less useful for grammatical agreement processes, although it helps speakers to access the gender of individual items. Deep transfer, however, seems to be more important for agreement type processes as both groups with gender performed better than the English group and no significant difference overall was found between the Romance and German group.

An overall effect of using the common gender as a default is also observable in the data. However, as can be seen in the interactions between gender and L1, the size of the effect tends to be different for the different L1 groups. There is not much evidence to indicate that the English group is making use of a default on high frequency items, only on middle frequency items. In fact, the English group, with only about 50% accuracy on the ungrammatical items in general, seems to be performing at chance. Only the German group shows a consistent strong effect of using a default gender.

Overall, grammatical gender seems to be more difficult to use in agreement processes than in assignment *per se*. Nevertheless, some interesting patterns are seen in the data. When all the relative pronoun data is considered there is an effect of L1 for at least middle frequency items. This suggests that at least in the initial stages of L2 acquisition transfer plays a large role in helping the learner along, with both surface and deep transfer playing an important role in the process of agreement.

#### IV General discussion

These experiments addressed several issues. The first was the extent and nature of transfer effects from the first language. In particular this was tested in L2 Dutch, making use of a gradient in the degree of similarity with regard to gender among the L1s of the participants: English has no gender, Romance has different gender and German has similar gender. In addition the experiments presented allowed us to look at frequency effects, the use of a default gender and L2 ability in assignment versus agreement. Each of these points is discussed below.

##### 1 Effects of L1

Taken together, the two experiments discussed above indicate that grammatical gender agreement poses a problem that is independent of the level of general syntactic proficiency; the relatively high level of overall proficiency in Dutch L2 of all participants did not correlate with an equally high proficiency in the specific domain of grammatical gender. On the contrary, performance on grammatical gender was affected by the L1 of the participants. For gender assignment, the German group performed best compared to the other two L2 groups; next, in terms of accuracy are the Romance speakers with an average score about 10% lower than the German group. This difference was, however, definitely smaller for the gender agreement experiment.

Doing worst on both gender experiments was the English group with no gender in their L1. This would suggest that deep transfer of the gender category plays a strong role in L2 acquisition. Further, surface transfer (direct transfer of the L1 gender system: a gender translation strategy) seems to be more helpful in learning lexical gender for gender



assignment than deep transfer, as seen in the difference between the German and Romance groups in Experiment 1.

The advantage of surface transfer over deep transfer was much less apparent for gender agreement, where the German group only performed better than the Romance group for high frequency items. Since no effect of L1 was found for high frequency items when scores on Experiment 1 were taken as a covariate, further research must be done to ascertain whether this reflects a real effect of only surface transfer playing a role or whether this is really a German advantage due to a direct gender translation strategy.<sup>9</sup>

This finding is in contrast to the study on gender production by White *et al.* (2004) that seemed to indicate that English participants can learn grammatical gender in Spanish and use it correctly. On the other hand, the current findings converge with the findings of Franceschina (2001; 2002), who found that the English speakers had persistent problems in learning the Spanish gender system. In fact, the English group seems to only perform at chance for gender agreement, though their scores for gender assignment are relatively good.

The difference between the English and Italian groups in Franceschina's study led her to conclude that learners could transfer abstract features (deep transfer) from L1 to L2. However, the relatively good performance of the German group compared to the Romance group suggests that while transfer of abstract features plays a role in the L2 acquisition of gender, direct surface transfer of a congruent system is even more helpful, at least for acquiring gender assignment. Surface transfer, thus, probably accounted for some part of the advantage shown by the Italian group in her study.

Deep vs. surface transfer, however, still needs to be further investigated. Sabourin and Haverkort (2003) found differential transfer effects within a German group acquiring Dutch: surface transfer effects were found in the absence of deep transfer effects among the same participants. This suggests that surface and deep transfer not only differs across L2 groups but that it can also differ within an L2 group.

<sup>9</sup>However, scores on agreement within the noun phrase (Sabourin, 2003) suggest that this is a true effect of surface transfer, with the gender category not transferring to allow the Romance participants to acquire the Dutch gender system.

## 2 Frequency effects

The results that were found in both experiments suggest, not surprisingly, that higher frequency nouns (which could also be considered the more familiar words) were easier for L2 participants in both gender assignment and gender agreement. This is consistent with previous results by Andersen (1984) who suggested that the amount of use of a language determined the level of gender acquisition. This could indicate that a certain amount of input is necessary before participants are able to fully assign a gender category to nouns. This is true even for the German speakers who can use a strategy of direct translation.

## 3 Use of a default strategy

Though all L2 participants did well on the gender assignment they all showed their lowest scores on assigning the correct determiner to the middle frequency neuter gender items. This pattern can be explained on the basis of two factors. First, learners make use of a default strategy in which they assign the more frequent gender (common gender) when they do not know the gender of a particular noun. Second, participants are more likely to know the gender of more frequent items. Thus, for the lower frequency items, the use of a default gender was more likely to occur. This pattern is particularly clear for the English group where the middle frequency *de* items appeared most accurate; the English group assigned the *de* determiner to almost all items they were not sure of, thus giving an appearance of high accuracy for common gender items.

This use of the default determiner was also seen in higher accuracy scores on common gender items of Experiment 2 for the English and German groups. The Romance group, on the other hand, made more use of the neuter gender form of the relative pronoun. Why this is the case is not obvious. It may be due to the high frequency of homophones of the neuter gender relative pronoun *dat*. *Dat* is used in many other constructions and seems to occur more often than *die*. Also, according to the CELEX database (Burnage, 1990), *die* (only as a relative pronoun) is less frequent than *dat* (appearing as both a complementizer and as a relative pronoun). This offers a possible explanation for why the Romance speakers overgeneralize to the use of the neuter gender

pronoun. Additionally it may explain a smaller effect of using a default gender in Experiment 2 as compared to Experiment 1 for all three L1 groups.

#### *4 Assignment vs. agreement*

The last issue addressed here was the extent to which there is a difference between the ability to learn gender assignment and the ability to learn gender agreement, although clearly the latter cannot efficiently be learned without the former. Higher scores on assignment as compared to the agreement scores were apparent for all of the L2 groups. This likely reflects the fact that while all L2 participants can learn the gender of specific items in isolation, making use of a noun's gender in an agreement context is much harder to do. In effect, this suggests that while all groups seem to be able to learn the gender of the L2 items, even if there is no gender whatsoever in their L1, after sufficient exposure, they differ in their ability to use this information in context.

The German group shows less advantage of surface transfer for gender agreement than for gender assignment. Taken together with the evidence that the German group made use of the default strategy like the other groups, this suggests some limitations on the extent to which surface transfer can help. The results of the two experiments presented here provide an indication of a difference in L2 knowledge and the L2 use of that knowledge in the grammatical system. It seems that certain aspects of grammar (e.g. gender) can be learned to a high degree of accuracy but that using this knowledge remains a problem for L2 learners. For the L2 data presented here, the difference between assignment and agreement seems to suggest that acquiring lexical gender knowledge (assignment knowledge) is possible even for those whose L1 does not include this feature, but that more syntactic-level agreement knowledge is considerably more difficult. However, the L2 data on agreement presented here does correlate with L2 ability on assignment. Beyond this effect though, an effect of L1 remains for the middle frequency items. This shows that agreement is still more difficult than assignment. Further, the English group that performs well on the assignment task only performs at chance on the

agreement task. This suggests that at least for a group with no gender in their L1, gender agreement is very difficult and may be impossible to acquire.

In summary, the data presented here suggests that surface transfer of the L1 lexical gender values does occur. Whether or not deep transfer of the grammatical category gender occurs still needs to be further investigated. It does seem to be the case, though, that L2 acquisition requiring deep transfer is more difficult. If the syntactic category gender could be transferred then it would have also been expected that the Romance speakers could also perform well on the tasks, as they have an L1 gender category that could have potentially been transferred to the L2. Their performance on the agreement task is above chance, and it is clearly worse than the Germans for middle frequency items. However, this hypothesis would need to be further tested by more on-line tasks and by the use of neuro-imaging where on-line processing can be measured and directly compared to native processing. Differentiating between the linguistic ‘setting’ of an item’s gender and just memorizing it (via a more general cognitive strategy) in off-line assignment tasks is very difficult solely using behavioural methods. This is where techniques that can measure on-line processing rather than conscious decision-making would be helpful to determine if linguistic or general cognitive strategies are being used. It is possible that gender assignment can be done based on more general cognitive skills (simply learning or memorizing what gender goes with what item) and thus can, with enough experience, be learned by any L2 learner. In contrast, gender agreement, relying on more linguistic strategies, can only be learned in the L2 if the same strategies are present in the L1.

### *Acknowledgements*

We are grateful to two anonymous *Second Language Research* reviewers for their comments, and to Marco Haverkort and Wander Lowie for discussion of the research presented here. The research for this study was conducted while the first author was a PhD candidate at the School of Behavioral and Cognitive Neuroscience, Department of Linguistics at the University of Groningen.

## V References

- Andersen, R.** 1984: What's gender good for anyway? In Andersen, R., editor, *Second languages: a cross-linguistic perspective*. Newbury House, 77–99.
- Burnage, G.** 1990: A guide for users. Nijmegen, CELEX Centre for Lexical Information.
- Corbett, G.** 1991: *Gender*. Cambridge University Press.
- Dewaele, J.-M.** and **Véronique, D.** 2000: Relating gender errors to morphosyntax and lexicon in advanced French interlanguage. *Studia Linguistica* 54, 212–24.
- Finneman, M.D.** 1992: Learning agreement in the noun phrase: the strategies of three first-year Spanish students. *IRAL* 30, 121–36.
- Franceschina, F.** 2001: Against an L2 morphological deficit as an explanation for the differences between native and non-native grammars. In Foster-Cohen, S. and Nizgorodcew, A., editors, *EUROSLA Yearbook*. Volume 1. John Benjamins, 143–58.
- 2002: Case and  $\Phi$ -feature agreement in advanced L2 Spanish grammars. In Foster-Cohen, S., Ruthenberg, T. and Poschen M.-L., editors, *EUROSLA Yearbook*. Volume 2. John Benjamins, 71–86.
- Gillis, S.** and **de Houwer, A.** 1998: *The acquisition of Dutch*. John Benjamins.
- Hawkins, R.** 1998: Explaining the difficulty of French gender attribution for speakers of English. Paper presented at EUROSLA 8, Paris.
- Hawkins, R.** and **Chan, Y.C.** 1997: The partial availability of Universal Grammar in second language acquisition: the failed features hypothesis. *Second Language Research* 13, 187–226.
- Myles, F.** 1995: Interaction between linguistic theory and language processing in SLA. *Second Language Research* 11, 235–66.
- Rogers, M.** 1984: On major types of written error in advanced students of German. *IRAL* 22, 1–29.
- 1987: Learners' difficulties with grammatical gender in German as a foreign language. *Applied Linguistics* 9, 219–35.
- Sabourin, L.** 2001: L1 effects on the processing of grammatical gender in L2. In Foster-Cohen, S. and Nizgorodcew, A., editors, *EUROSLA Yearbook*. Volume 1. John Benjamins, 159–69.
- 2003: *Grammatical gender and second language processing: an ERP study*. GRODIL: Groningen Dissertations in Linguistics, 42.
- Sabourin, L.** and **Haverkort, M.** 2003: Neural substrates of representation and processing of a second language. In van Hout, R., Hulk, A., Kuiken, F. and Towell, R.J., editors, *The lexicon–syntax interface in second language acquisition*. John Benjamins, 175–95.
- Schwartz, B.D.** and **Sprouse, R.** 1996: L2 cognitive states and the full transfer/full access model. *Second Language Research* 12, 40–72.

- Shelton, J.** 1996: Second language acquisition of grammatical gender and agreement in French. PhD thesis, University of Essex.
- Vainikka, A.** and **Young-Scholten, M.** 1996: Gradual development of L2 phrase structure. *Second Language Research* 12, 7–39.
- van Hout, R.** 1996: Waar kwamen hun onze taal binnen? [Where did they enter our language?] In van Hout, R. and Kruijssen J., editors, *Taalvariatie [Language variation]*. Foris, 143–56.
- van Leuvensteijn, J.A.** with **Ruijsendaal, E.**, **Mooijaars, M.A.**, **Geirnaert, D.J.G.** and **Temperlaars, R.** 1997: Vroegnieuw-nederlands [Early Modern Dutch] (circa 1550–1650). In van den Toorn, M.C., Pijnenburg, W.J.J., van Leuvensteijn J.A., and van der Horst, J.M., editors, *Geschiedenis van de Nederlandse taal* [History of the Dutch language]. Amsterdam University Press, 227–359.
- White, L.** 1985: The pro-drop parameter in adult second language acquisition. *Language Learning* 35, 47–62.
- 2000: Second language acquisition: from initial state to final state. In Archibald, J., editor, *Second language acquisition and linguistic theory*. Blackwell, 130–55.
- White, L.**, **Valenzuela, E.**, **Kozłowska-Macgregor, M.** and **Leung, I. Y.-K.** 2004: Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics* 25, 105–33.