Gender Congruency and Cognate Effect in Bulgarian-English Bilinguals: Evidence from a Word-Translation Task

Armina Janyan (ajanyan@cogs.nbu.bg)
Central and East European Center for Cognitive Science, New Bulgarian University
21 Montevideo St., 1618 Sofia, Bulgaria

Marina Hristova (mhristova@nbu.bg)
Central and East European Center for Cognitive Science, New Bulgarian University
21 Montevideo St., 1618 Sofia, Bulgaria

Abstract

The present study aims to investigate the cognate and gender congruency effects in Bulgarian-English proficient bilinguals using a visual word translation task. The results of two experiments confirm the pattern generally observed in the area: we have found a strong cognate effect present in both translation directions (from the dominant L1 into non-dominant L2 and vice versa). In addition, we have obtained a gender congruency effect which suggests that bilinguals use gender information of their native language during translation.

Keywords: bilingualism, word translation, cognates, gender congruency, reaction time.

Introduction

Most researchers agree on two basic assumptions concerning bilingual language organization: first, that the concepts are shared between the languages and second, that lexical representations are segregated to a great extent. One such model of bilingual language representation is the revised hierarchical model (RHM) which assumes that lexical representations are language specific whereas conceptual representations are shared among the languages (Kroll & Stewart, 1994; Kroll & de Groot, 1997). The model specifies that language-dominant (L1) and language-non-dominant (L2) word forms are associated with each other and to a common concept to different degrees. The strength of these associations is a function of language proficiency. The RHM proposes that both L1 and L2 word forms activate common concepts but that L1 words activate concepts more strongly than L2 words. In addition, the RHM posits that L1 and L2 words can activate each other at the lexical level through translation. However, according to the model, the lexical links are stronger in the L2-L1 direction than in the opposite one. Thus, the RHM predicts facilitated translation from L2 to L1 which is confirmed by experimental research (e.g., Kroll & Stewart, 1994; Francis, Augustini, & Sáenz, 2003).

To address the issue of bilingual language representation, many studies compare processing of cognates and non-cognates in bilinguals. Cognates are words that have similar form and meaning in both languages (L1 & L2) such as English lamp and Bulgarian лампа [lampa]. Non-cognates have different forms in L1 and L2 but similar meaning such as elephant in English and слон [slon] in Bulgarian. The general finding is that cognates are produced, recognized and translated faster than non-cognates (e.g., Kroll & Stewart, 1994; Dijkstra, Grainger, & van Heuven, 1999; Costa, Caramazza, & Sebastian-Galles, 2000; Costa, Santesteban, & Caño, 2005). The faster production, recognition and translation of cognates are usually attributed to a common set of form-based representations (e.g. orthographic, phonological, morphological) that are used to process them in both languages. Another consideration in the model is that the advantage magnitude of cognates is also dependent on the language proficiency of an individual – the less proficient the bilingual is in his/her non-dominant language, the larger the magnitude (Kroll & de Groot, 1997) - which suggests that less proficient bilinguals may rely more on lexical translation strategy bypassing concept mediation.

However, in addition to various lexical and conceptual properties a word also has grammatical ones. One of these grammatical features widely researched in the area of lexical access is grammatical gender. Among the questions asked are where and when in the course of (monolingual) lexical access grammatical information is selected and what are the factors that influence that selection (e.g., Bates, Devescovi, Caño, 2005). The faster production, recognition and translation of cognates are usually attributed to a common set of form-based representations (e.g. orthographic, phonological, morphological) that are used to process them in both languages. Another consideration in the model is that the advantage magnitude of cognates is also dependent on the language proficiency of an individual – the less proficient the bilingual is in his/her non-dominant language, the larger the magnitude (Kroll & de Groot, 1997) - which suggests that less proficient bilinguals may rely more on lexical translation strategy bypassing concept mediation.

A model that tries to take into account syntactic and grammatical features representation in the bilingual mind is the distributed feature model (Kroll & de Groot, 1997). First, the model assumes that lexical forms and meanings are integrated within unitary lexical and conceptual systems; moreover, it assumes varying degrees of overlap of conceptual and lexical features between L1 and L2 meanings. Thus, for example, the model predicts faster cognate translation not necessarily because the translation processing might be restricted to the lexical level but because cognates have a higher level of feature overlap on both the lexical and the conceptual levels (Kroll & de Groot, 1997). That is, the level of processing facilitation or interference is a function of the degree of form-meaning mapping consistency within and across languages. The model also takes into account the language proficiency level assuming weaker form-meaning mapping with less L2
proficiency. Finally, the model introduces a language-specific “lemma-level” which represents patterns of activation between word forms and meanings. In the absence of context, the lemma level may reflect only these form-meaning mappings. In the presence of context, the lemma may also reflect syntactic-level processes that selectively weight the activation of lexical and conceptual features (Kroll & de Groot, 1997). This means that in the presence of syntactic context (e.g., grammatical gender priming) the form-meaning mappings will get higher activation when the context is congruent with the mapping and the processing will be facilitated. On the other hand, if the context is incongruent, processing will be encumbered.

The present study has two main aims. The first is to confirm the cognate effect in Bulgarian-English bilinguals and the facilitated translation from L2 to L1. Both these effects are predicted by both models described above and widely established in other language pairs. Second, we ask if the bilinguals use gender information during lexical access when their native language does have gender marking1, and the non-native does not. Previous research on the role of the gender system in bilingual language production has obtained contradictory results. Costa, Kovacic, Frank & Caramazza (2003) asked a number of groups of bilinguals with different language pairs to name pictures whose names had either the same gender in both languages or different. The results showed no difference in these two conditions. The authors concluded that the two gender systems of a bilingual are functionally autonomous. On the other hand, Bordag (2004) conducted the same experiment with Czech-German and German-Czech bilinguals and obtained a robust effect of grammatical gender of L1 – pictures with congruent gender names were named faster in L2 than pictures with gender-incongruent names. Finally, a study by Guillemon & Grosjean (2001) employed auditory word naming preceded by either gender-congruent or gender-incongruent context. The results showed that gender priming produced inhibitory and facilitatory effects but only for early English-French bilinguals, late bilinguals were not sensitive to the gender marking. It seems that these results fit nicely with the distributed feature model (Kroll & de Groot, 1997).

Thus, we test the prediction of the distributed feature model (DFM) that context will lead to the selective weight of activation of some features over others. More specifically, we expect that gender-congruent context will facilitate translation especially in the case of cognates since they share a larger number of form features between languages than non-cognates. In addition, we expect that it will also enhance the effect of translation direction (faster translation from non-dominant gender-free language into dominant gender-inflected language).

In order to activate both languages and to approach bilingual mode as much as possible (Grosjean, 1997), we have used mixed language presentation that requires constant language switch: our language-proficient participants were presented with words in Bulgarian and in English and were required to translate from Bulgarian into English and from English into Bulgarian.

Experiment 1 was designed to replicate the cognate effect and facilitated translation in the L2-L1 direction in Bulgarian-English bilinguals.

**Method**

**Participants** 16 professional university teachers of English (4 males and 12 females) volunteered to participate in the experiment. They had an average age of 35.4 years (SD=6.7). All of them had normal or corrected to normal vision. All participants filled in a second language history questionnaire. Table 1 presents means and standard deviations of three self-reported measures.

**Stimuli and Design** 60 cognates and 60 non-cognates were selected from 2 data bases: one unpublished 400-word data base (containing subjective frequency and word imageability), and another data base with 520 pictures and their dominant responses in seven languages including English and Bulgarian (Szekely et al., 2004)2. For each picture the data base contains the dominant responses for English and Bulgarian. These dominant responses served as a kind of control of translation equivalents and as lexical items for the experiment. Items taken from the unpublished data base had no such control of translation equivalents; their dominant English translations were generated by a near-native speaker of English. Translation equivalents for both groups were then double-checked by a professional translator. Thus, each of the 60 cognates and 60 non-cognates had both English and Bulgarian equivalents. All words referred to concrete objects. Bulgarian cognate items were matched by subjective frequency with non-cognates.

Since no subjective frequency was available for all English translation equivalents, objective frequency for words in both languages was taken as a common lexical characteristic. Objective word frequency data for Bulgarian words were derived from a 72-million data base (Simov et al., 2004),

---

1 Bulgarian is a Slavic language with three grammatical genders marked on nouns, adjectives, pronouns, (postpositional) articles, some numeral and verb-forms.

2 For details on participants, procedures of object naming experiment, and various psycholinguistic characteristics, visit an on-line data base at http://www.crl.ucsd.edu/~aszekely/ipnp/.

---

Table 1: Self-reported measures of second language history.

<table>
<thead>
<tr>
<th>SubProf M(SD)</th>
<th>AoA M(SD)</th>
<th>Period M(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5 (0.5)</td>
<td>9.3 (3.6)</td>
<td>25.5 (7.2)</td>
</tr>
</tbody>
</table>

**Note.** SubProf – subjective proficiency rating on a 6-point scale (6 – the most proficient), AoA – age of the second language acquisition (in years), Period – period of second language learning (in years).
converted into frequency score per million and then 10-base logarithm of the score was taken with one added to the score per million to avoid the undefined \( \text{Lg}(0) \). Objective word frequency data for English words were taken from an online data base of word-form characteristics (Balota et al., 2002). The English frequency data were based on Hyperspace Analogue to Language (HAL) frequency norms which consist of approximately 131 million words (Balota et al., 2002). The same procedure of conversion into one-million score and logarithm taking was applied to the English HAL data. In addition, the words were coded for word length measured in number of letters. Descriptive statistics is presented in Table 2. T-tests showed that English words were longer and more frequent than Bulgarian words and that cognates were longer than non-cognates (\( p<.05 \)). No differences were found only in word frequency within languages across word types.

Table 2: Means and standard deviations (in parentheses) of word characteristics for each language and each word type.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>En</td>
</tr>
<tr>
<td>Cognate</td>
<td>1.6 (0.6)</td>
</tr>
<tr>
<td>N-Cognate</td>
<td>1.4 (0.4)</td>
</tr>
</tbody>
</table>

The experiment had a 2 x 2 design: 2 (Word Type: Cognate vs. Non-cognate) x 2 (Language of presented stimulus: English vs. Bulgarian).

Procedure Four pseudo-randomized lists were constructed so that the same condition appeared no more than 3 consecutive times and so that there were at least two non-animate words between two animate. Each list contained 60 English and 60 Bulgarian words. Each stimulus was presented only once to each participant either in English or in Bulgarian.

Participants were tested in a sound-proof booth. The experimental session started with 8 practice trials, none of which appeared in the experimental part. Each trial started with a fixation cross “+” for a random time from 400 to 600 ms. After cross disappearance a word was displayed for 3000 sec or until a subject’s response was generated. Stimuli were presented at the center of the screen in 28-point Bulgarian Times. Participants were instructed to translate words as fast and as accurately as possible into the corresponding language (from Bulgarian to English and from English into Bulgarian). The intertrial interval varied randomly between 1.7 and 2.2 sec. Responses were recorded by the experimenter. Reaction time (RT) was measured from the onset of each stimulus. A serial response button box recorded voice onset RT. Stimulus presentation and response recordings were controlled by E-prime software (Schneider, Eschman, & Zuccolotto, 2002). The experiment took about 10 min.

Results and Discussion

The following types of errors were identified and excluded from the data prior to analysis. Trials on which no response was registered (3.3%) and trials with technical errors (3.0%) were excluded from the analyses. In addition, trials with task switching (word reading, 0.5%) and trials with different from the target equivalent production (3.1%) were discarded. Further, response times lying more than ±2.5 standard deviations from the mean were also removed, which resulted in the removal of 2.7% of overall responses.

Thus, a total of 87% of the originally collected RT data were included in further analyses. The reported results are based on both item and subject analyses. Since frequency and word length were not fully controlled, item data were analyzed with analysis of co-variance (ANCOVA) where word frequency and length served as covariates. Similarly, subject data were also run with ANCOVA with the three self-reported measures of language history (cf. Table 1) and subjects’ age as covariates. In such a way the influence of (possible) confounding contained in item and subject characteristics, was eliminated (or at least, highly reduced).

Table 3 presents item means and standard deviations for each condition in ms.

<table>
<thead>
<tr>
<th></th>
<th>Cognate</th>
<th>Non-cognate</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>1009 (106)</td>
<td>1071 (153)</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>987 (148)</td>
<td>1132 (151)</td>
</tr>
</tbody>
</table>

Figure 1: A Language by Word Type interaction.

A 2x2 ANOVA obtained a significant main effect of word type and significant interaction. Main effect of language was highly insignificant (\( p>.3 \)). Main effect of word type (\( F(1, 234)=27.18; p=.00 \); \( F(1, 56)=13.39; p<.001 \)) showed that cognates were translated faster (998 ms) than non-cognates (1101 ms). A language by word type interaction (\( F(1, 234)=5.86; p<.05 \); \( F(1, 56)=3.17; p=.081 \)) is presented in Figure 1. It can be seen that cognates significantly facilitated translation from Bulgarian into English. Cognate facilitation during translation from English to Bulgarian missed the significance level (\( p<.07 \)). In addition, the analyses showed a trend of easier translation from non-dominant English into dominant Bulgarian than from
Bulgarian into English \((p<.09)\). No significance was found between RT of cognate translation from both languages (see Figure 1).

To summarize, the results replicated the cognate effect (e.g., Kroll & Stewart, 1994) and showed a trend\(^3\) of facilitated L2-L1 translation only for non-cognates. Cognates were translated equally fast from both languages.

Experiment 2 employed the same design adding a gender congruency variable.

### Experiment 2

#### Method

**Participants** In addition to the same 16 subjects that participated in Experiment 1, 8 more professional university teachers of English volunteered for the experiment. Thus, 24 participants (4 males and 20 females) took part in the experiment. They had an average age of 36 years (SD=8.0). All of them had normal or corrected to normal vision. Table 4 presents means and standard deviations of three self-reported measures.

<table>
<thead>
<tr>
<th>SubProf</th>
<th>AoA</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>M(SD)</td>
<td>M(SD)</td>
<td>M(SD)</td>
</tr>
<tr>
<td>5.5 (0.5)</td>
<td>10.0 (3.4)</td>
<td>25.2 (8.1)</td>
</tr>
</tbody>
</table>

*Note.* SubProf – subjective proficiency rating on a 6-point scale (6 – the most proficient), AoA – age of the second language acquisition (in years), Period – period of second language learning (in years).

**Stimuli and Design** Another 60 cognates and 60 non-cognates were selected with the use of the same two data bases (see Experiment 1) and the same procedure was applied in establishing translation equivalents and in subjective frequency conversion. Descriptive statistics is presented in Table 2. Item characteristics (length and frequency) were matched with the characteristics of items of the Experiment 1 (cf. Table 2; no significant differences between the characteristics were found).

Table 5 presents means and standard deviations (in parentheses) of word characteristics for each language and each word type.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>En</td>
<td>Bg</td>
</tr>
<tr>
<td>Cognate</td>
<td>1.4 (0.5)</td>
</tr>
<tr>
<td>N-Cognate</td>
<td>1.4 (0.5)</td>
</tr>
</tbody>
</table>

Here again, t-tests showed that English words were longer and more frequent than Bulgarian words and that cognates were longer than non-cognates \((p<.05)\). No differences were found in word frequency within languages across word types (cf. Table 5).

Gender primes were selected to be semantically neutral and to vary in their form and phrase structure starting from a single gender-inflected modifier \(\text{edm}a=\text{one}, \text{feminine, singular; njakakv}a=\text{some (kind of)}, \text{feminine, singular}\) to a verb phrase \(\text{tam se vizhdala}=\text{there was seen}, \text{feminine, singular; “tam se vizhdala edna/njakakv}a=\text{there was seen one/some}, \text{feminine, singular}\). The prime was considered congruent if it agreed in gender with the Bulgarian translation equivalent (e.g. “\text{ed}ma” = “\text{one}, \text{feminine, singular}, and “\text{sal}ata” = “\text{sal}ad”, \text{feminine, singular}) and incongruent if it did not (e.g. “\text{ed}na” and \text{kom}pas = “\text{compass}, \text{masculine, singular}). The five types of modifiers in each of the three genders (feminine, neuter, and masculine) were taken from the stimuli of another experiment conducted in the lab. The sound stimuli were recorded by a female native Bulgarian speaker in a neutral intonation. The words were digitized using the Macintosh SoundEdit system and placed in a sound file within the PsyScope experiment preparation shell (Cohen, MacWhinney, Flatt, & Provost, 1993).

The experiment had a 2 x 2 x 2 design: 2 (Word Type: Cognate vs. Non-cognate) x 2 (Gender priming: Congruent vs. Incongruent) x 2 (Language of presented stimulus: English vs. Bulgarian).

**Procedure** The procedure was the same as in Experiment 1 except for the following. Each of the four lists contained 50% primes that agreed in gender and 50% that disagreed in gender with either Bulgarian targets that were to be translated into English or Bulgarian equivalents of English targets that were to be translated into Bulgarian. The phrase structure of primes was distributed roughly equally in a list (i.e., each structure appeared approximately 1/5 of the time).

After cross disappearance the gender prime was presented auditorily, immediately after which a stimulus was displayed. The stimuli were presented in the center of the screen with 30-point CyrillclIISans font. The sound stimuli were presented using the Macintosh inbuilt sound system without headphones.

A Carnegie Mellon button box recorded voice onset RT and controlled stimuli presentation timing. A Power Macintosh 6400/200 equipped with the PsyScope software (Cohen et al., 1993) controlled order of presentation of the stimuli. The experiment lasted 15-20 min.

### Results and Discussion

Several types of errors were identified and excluded from the data prior to analysis. First, trials on which no response was registered \((5.8\%)\) and trials with technical errors \((6.3\%)\) were excluded from the analyses. Second, trials with task switching (simple word reading, 0.3\%) and trials on which a different from the target equivalent word was produced \((6.5\%)\) were discarded (in many cases participants tried to produce a gender congruent response in the incongruent condition, especially for a neuter gender prime followed by...
a masculine noun\(^4\)). Furthermore, response times lying more than ±2.5 standard deviations from the mean were also removed, which resulted in the removal of 2.15% of overall responses.

Thus, a total of 78.9% of the originally collected RT data were included in further analyses. The reported results are based on both item and subject analyses. Similarly to the analyses conducted in Experiment 1, both types of averaged data were analyzed with ANCOVA where in item analyses word frequency and length served as covariates (cf. Table 5) and participants’ age and the three self-reported measures of language history (cf. Table 4) - in subject analyses. Table 6 presents item means and standard deviations for each condition in ms.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th>Bulgarian</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognate</td>
<td>Non-cognate</td>
<td>Cognate</td>
<td>Non-cognate</td>
</tr>
<tr>
<td>Congruent</td>
<td>1031(171)</td>
<td>1187(166)</td>
<td>1061(147)</td>
<td>1310(164)</td>
</tr>
<tr>
<td>Incongruent</td>
<td>1101(178)</td>
<td>1249(189)</td>
<td>1098(180)</td>
<td>1308(190)</td>
</tr>
</tbody>
</table>

A 2x2x2 ANCOVA obtained two significant main effects and one interaction. Main effect of word type (F\(_{1, 460}=150.10; p<.00; F_{(1, 180)=84.39; p=.00}\) showed that cognates were translated faster (1073 ms) than non-cognates (1264 ms). Main effect of gender congruency (F\(_{(1, 460}=7.12; p<.01; F_{(1, 180)=5.07; p<.05}\) revealed that congruent primes facilitated translation (1148 ms vs. 1189 ms). Although subject analysis showed significance in main effect of language, item analyses did not confirm it (F\(_{(1, 460)=0.09; p>.7; F_{(1, 180)=4.33; p<.05}\)\(^5\).

![Figure 2: A Language by Word Type interaction.](image)

A language by word type interaction (F\(_{(1, 460)=6.63; p<.05; F_{(1, 180)=3.85; p=.051}\) is presented in Figure 2. It suggests that cognates facilitated translation from both languages and that, predictably, translation from English into Bulgarian was easier in terms of reaction time than from Bulgarian into English. Finally, translation of cognates from both languages was equally easy (cf. Figure 2).

The results of this experiment replicated and strengthened those obtained in Experiment 1. The results suggest that cognate status has a robust facilitating effect on translation and that translation from L1 to L2 takes longer than translation from L2 to L1. These two effects are in agreement with the predictions of both types of bilingual models (RHM and DFM). Most importantly, the results showed the gender congruency effect on translation processing. Although no significant interaction between gender congruency and any other variable was obtained, Table 6 suggests that the congruency effect was present when translating into Bulgarian (the language with a gender system). Thus, gender congruent/incongruent primes seem to facilitate/inhibit translation when the target language has a gender system, but do not interfere with production in a gender-free language. A separate post hoc analysis on insignificant interaction between gender congruency and language (F\(_{(1, 460)=2.13; p=.15; F_{(1, 180)=2.80; p=.09}\) showed that indeed, translation from English into Bulgarian was affected by gender congruency (p<.05). Thus, the expectations were partially fulfilled. In particular, congruency was found to have an effect on translation in the L2-L1 direction, although in the present study this interaction did not reach significance level. This congruency processing advantage might emerge more clearly with the use of more sensitive measures such as EEG (e.g. Barber & Carreiras, 2003).

**Conclusion**

One of the purposes of this study was to replicate the cognate facilitation effect in word translation and translation facilitation in L2-L1 direction compared to L1-L2 in Bulgarian-English bilinguals. The results of two experiments clearly confirmed results of other studies (Kroll & Stewart, 1994; Francis et al., 2003) – L2-L1 translation was found to be faster, which could be taken as evidence in favor of stronger lexical links in that direction, as predicted by the RHM. In addition, the results suggest that translation facilitation in L2-L1 direction is valid only for non-cognates. Thus, in certain cases, cognate status will override the translation direction effect.

The second purpose of the study was to explore the gender congruency effect in a translation task. Studies on grammatical gender contribution to monolingual language processing suggest that in production, a robust gender congruency effect is usually observed in languages with a free-standing determiner system; however, it concerns languages where the determiner form can be selected on the basis of noun gender alone (as in German) and not those in which the determiner depends not only on the noun gender but also on the phonological context (as in Italian) (Miozzo et al., 2002; Schiller & Caramazza, 2003). Our study shows that the gender congruency effect may be detectable with bare noun translation task without producing freestanding gender-marked morphemes in addition to the noun (see Costa et al.,...
Our results are in agreement with the picture-naming bilingual study (Bordag, 2004) and extend the findings to the translation task and to bilinguals whose L1 has a gender system but whose L2 does not. They are also in accordance with the predictions of DFM that syntactic-level processes can increase/reduce the activation of lexical and conceptual features and consequently result in faster/slower selection of the appropriate translation equivalent.

Further research is needed to explore the gender congruency effect in bilinguals and its possible interaction with the cognate status of a word.

Acknowledgments

We would like to express our special gratitude to our volunteer participants for their patience and willingness to help.

References


