Perceived lexical similarities between L2 Italian and L3 English in the reading comprehension of Croatian-Italian bilingual EFL learners

Ana Bradičić
Italian High School – Rijeka, Croatia
Teacher of English and German Language and Literature
ana.bradicic1@ri.t-com.hr

Abstract: Recent studies investigating transfer in language acquisition have shown that not only the knowledge of a first language, but also all other languages known to a person may facilitate the acquisition of a new language. This is also the case with languages belonging to different language families such as English and Italian, which, nonetheless, have many conspicuous lexical similarities. Whereas studies have concentrated on language production and error analysis, it is necessary to acknowledge the importance of investigating third language comprehension as well. The present study examines the perception of lexical similarities in written text comprehension by Croatian-Italian bilinguals who are at two distinct levels of English proficiency. A form involving similarity judgments for lexical items varying in the degree of formal and semantic similarity has been designed drawing on real language use as provided by corpora. The results obtained are compared to objective formal similarity as provided by a string matching algorithm, the normalized Levenshtein distance. Results suggest that the ratings of lexical similarity perceived by the learners are related to formal and semantic word similarity. They also indicate that in the case of semantically similar words older students rely more on previously acquired lexical knowledge, whereas younger learners tend to give more uniform ratings relying more on formal similarity. We suggest that an explicit approach to raising the learners’ awareness of language similarity and to promoting transfer as a learning strategy would improve the third language learning process and its outcome.

Key Words: Lexicon, crosslinguistic influence, similarity perception

Introduction

For a number of years studies of crosslinguistic influence (CLI) focused on the role of the first language in the acquisition of the second language, and the analysis of errors in the learners’ second language production had the main role in establishing the predictive force of transfer (Larsen-Freeman and Long, 1991). The results of more recent studies have shown that all languages acquired after the first language might be the source of CLI when learning a new language (Cenoz, 2001; Dewaele, 1998; Ringbom, 1987; Selinker and Baumgartner-Cohen, 1995; Williams and Hammarberg, 1998). Several factors have been identified as relevant in this process, among which crosslinguistic similarity, language proficiency, context of communication, age, recency of language use and the “foreign language effect” or the L2 status. In the case of the second language, its ability to function as a source language for CLI is not determined as much by the order of acquisition as by other factors such as proficiency in the source language, frequency and recency of use, and the degree of similarity to the recipient language. Since the majority of studies deal with language production, it is necessary to investigate CLI effects in perception and comprehension in order to gain a deeper understanding of the process of third

Crosslinguistic influence, the term proposed by Kellerman and Sharwood Smith in 1986, is used in the literature as the equivalent of the term transfer or interference, and it includes the wide range of psycholinguistic phenomena which result from the contact of two languages, including preference, positive and negative transfer, avoidance and borrowing (Jarvis and Pavlenko, 2008).

The phenomenon called linguistic distance, typological similarity, psychotypology and crosslinguistic similarity refers to the degree of similarity between the source language and the recipient language. It determines the possibility of the appearance of transfer (Kellerman, 1977; Ringbom, 1978; Wode, 1976; Cenoz, 2001; Ringbom, 2001, 2007).
language acquisition. While in language production the speaker starts from the preverbal intention and assigns it a linguistic form, comprehension has its starting point in the lexical form the reader or listener gives meaning to by linking the language form to the semantic knowledge he possess (Ringbom, 2001). Therefore learners try to find similarities between the languages they know in order to establish a relationship of equivalence between them, and in this process formal similarities have a greater importance in comprehension than in language production (Ringbom, 2007). However, it is important to distinguish between objective and subjective similarity, but also between objective and subjective difference. Objective similarities are often not the same as subjective similarities since it is the subjective perception of the speaker to determine the degree of transferability of certain language features from the source language to the recipient language (Kellerman, 1978). The subjective crosslinguistic similarity influences the degree to which the learner relies on the source language in learning and using the target language, whereas objective similarity affects the likelihood that transfer will be positive or negative (Odlin, 1989). However, objective similarities and differences and subjective differences don't lead to transfer, but it is the crosslinguistic similarities that the learners perceive or assume to exist, that form the basis of interlingual identifications that generate most types of transfer (Jarvis and Pavlenko, 2008). Perceived language distance is therefore considered the main factor influencing crosslinguistic similarity, and formal similarity between all language systems known to the learner a constant factor of crosslinguistic influence (De Angelis and Selinker, 2001).

The importance of perceived crosslinguistic similarity has been attested in all areas of language knowledge, from phonology, morphosyntax to the discourse, but it is mostly evident in the field of the lexicon. Crosslinguistic similarity is most obviously perceived on the basis of formally similar or identical individual items or words. The similarities may also be functional or semantic. Formal similarity is perceived first, in that getting the word form precedes getting the word meaning, and most cases involving the transfer of formal properties seem to reflect perceived similarities, while many cases of semantic transfer seem to occur merely on the basis of assumed similarities and often despite observable differences. Moreover, in the case of formal transfer the source language tends to be a closely related language, that is a source language that the learner perceives as being closely related to the recipient language, whereas semantic transfer strongly tends to come from a language in which the learner is highly proficient, most commonly the L1 but also an L2 in cases where the learner is highly advanced in the L2 (Ringbom, 1987, 2001; Cenoz, 2001).

Although perceived similarity mainly facilitates learning, especially when cognate languages are concerned, it can also lead to errors in production and comprehension, as in the case of false friends, i.e. lexemes that belong to different languages, but because of their formal similarity lead to wrong conclusions about their meaning (Gabryś-Barker, 2006, Ringbom, 2007). The first encounter with cognates often leads to an approximate understanding and it doesn't require much cognitive effort on the part of the learners since they easily connect the core meaning of words with the same or similar form. However, on more advanced levels of language learning, it is necessary to invest a greater cognitive effort to acquire the full meaning of lexical units, their stylistic and contextual components (N. Ellis, 1994, 1997; Nemser, 1998; Singleton, 1999).

Despite the fact that the research of lexical transfer has received much attention, due to the complexity of this phenomenon that includes several dimensions of lexical knowledge, it is necessary to investigate the implications of this process in multilingual speakers where lateral transfer is manifested, i.e. the transfer from the L2 to the L3 (Jarvis and Pavlenko, 2008).

**Method of the Study**

The main goal of this study is to gain an insight into the perception of lexical similarities between L2 Italian and L3 English in the reading comprehension of Croatian-Italian bilingual speakers. With regard to objective and subjective similarity, the aim is to establish if the subjects perceive crosslinguistic lexical similarity in reading comprehension, and if so, to which degree they rely on semantic, i.e. formal similarity. Furthermore, the aim is to establish the existence of a relationship between the perception of lexical similarities of L2 Italian and L3 English with the subjects’ characteristics (first language, Italian language learning experience, context and frequency of use of L2 Italian, length of studying and language proficiency in L3

---

26 For the purpose of this paper, it is important to underline the difference between the second (L2) and the third language (L3), since when acquiring the L2 the learner has at his disposal only the knowledge of his L1, whereas when acquiring the third language (L3), he already possesses the knowledge of the previously learnt languages (Hammarberg, 2001). Therefore, an adult learner who starts learning an L3 or an L4, possesses metalinguistic knowledge and learning strategies that a learner of the L2 doesn't have (Hufeisen, 1998).
English, self-rated language knowledge of English and Italian and formal evaluation in L2 and L3 at the end of the school year) and the features of the lexical items included with reference to word class membership, historical relatedness and semantic and formal similarity. The study of the perception of lexical similarity between L2 English and L3 Swedish (Utgof, 2008) represents the methodological starting point and the basis of this work.

**Sampling**

The present study includes 69 subjects, students of a high school in Croatia where the language of instruction is Italian. 35 of them are students of the first class and have been studying English for 7 years and 34 of them are students of the fourth class and have been studying English for 10 years. All informants are bilingual speakers whose L1 is mostly Croatian and the L2 is Italian, which is mainly used only as the language of instruction. The subjects differ also on the following points: Italian language learning experience, context and frequency of use of the Italian language, self-rated proficiency in L2 and L3 and the formal mark they have in the subjects Italian language and literature and English language at the end of the school year 2009/2010.

**Materials and Analysis Processes**

The questionnaire used to gather the data consists of a first part involving similarity judgements for 28 word pairs varying in the degree of formal and semantic similarity. Formal similarity was computed by means of the normalized Levenshtein distance, which is a string edit distance for measuring the amount of difference between two strings of characters taking into consideration their length. Since similarity and difference are expressed by a reciprocal measure, the measure of formal similarity is given precedence in consonance with the theoretical assumption that learners tend to rely more on similarities than on differences (Ringbom, 2007). The measure of semantic similarity is given by the similarity ratings of five experienced Croatian-Italian bilingual teachers of English. The included lexical items differ also in their grammatical category, etymological origin, and out of 28 word pairs, there are 23 word pairs that belong to the category of deceptive cognates. Although English belongs to the Germanic language family, a significant number of lexemes has its origin in Latin, so in terms of its lexicon it can plausibly be regarded as a Romance language (Singleton, 2006), which has important implications for the psychotypological effect. It is therefore obvious that a learner who has a knowledge of Italian will encounter a whole array of English words whose meaning he can associate with a previously acquired language. However, deceptive cognates can lead to wrong conclusions since their meaning in English doesn't match the meaning of a similar word in Italian. The importance of context lies in the possibility of lexical and conceptual processing of ambiguous lexemes and in aiding comprehension (Gianico and Altarriba, 2008). All test items have therefore been contextualized and included into sentences provided by the British National Corpus for the English sentences and by *Corpora e lessici dell’italiano parlato e scritto* and *Corpus dell’italiano scritto contemporaneo* for the Italian sentences. The British National Corpus is an exemplary representative corpus, used in many previous studies, while the corpora of Italian have been chosen for ease of access and their content including spoken and written language as well as literary texts. The sentences were skimmed and chosen at random, and some were edited for length to fit in the form. The included items were checked against the vocabulary list included in the student’s book used in grade 1 and 4 (Soars, L. and J., 2005, 2009) and chosen according to the researcher’s judgment. To exemplify the mode of analysis, it has to be mentioned that the similarity ratings have been given on a scale ranging from 10 (totally different) to 100 (completely the same) as in the study of lexical similarity perception between Swedish and English by Utgof (2008). The usage of ciphers was abandoned in hope that the students would place a mark after their intuition rather than concentrate on mechanically choosing a number. The respondents were asked to place a cross on the line to reflect how similar they considered the word pairs to be.

Language biography is the second part of the questionnaire supplying information about the biographical characteristics of the informants, context of acquisition and use of L2 Italian and L3 English.

**Findings and Discussion**

In order to establish and differentiate the effects of formal and semantic similarity on the crosslinguistic similarity ratings and verify the effects of the grade the subjects attend, i.e. the level of L3 proficiency, a three-way ANOVA with repeated measurements on 2 factors was performed followed by post hoc Scheffé tests where appropriate (i.e. if the F score was statistically significant) (Table 1.). The word pairs have been divided into two categories according to the degree of their formal and semantic similarity with values ranging from 0 – 0.50 for
low similarity and $0.51 - 1$ for high similarity. The alpha for achieving statistical significance was set at .05. Additionally, effect sizes were calculated to examine the strength of the relationship between the variables.

**Table 1. Effects of formal similarity, semantic similarity, grade and their interaction on English-Italian word pair similarity judgements**

<table>
<thead>
<tr>
<th>Factor</th>
<th>$F_{1,67}$</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>formal similarity</td>
<td>19.23**</td>
<td>.22</td>
</tr>
<tr>
<td>semantic similarity</td>
<td>25.51**</td>
<td>.28</td>
</tr>
<tr>
<td>Grade</td>
<td>0.70</td>
<td>.01</td>
</tr>
<tr>
<td>formal similarity * semantic similarity</td>
<td>0.01</td>
<td>.00</td>
</tr>
<tr>
<td>formal similarity * grade</td>
<td>2.56</td>
<td>.04</td>
</tr>
<tr>
<td>semantic similarity * grade</td>
<td>4.46*</td>
<td>.06</td>
</tr>
<tr>
<td>formal similarity * grade</td>
<td>0.19</td>
<td>.00</td>
</tr>
</tbody>
</table>

*p<.05; ** p<.01

As can be seen from Table 1., in analyzing the role of formal similarity and the role of English proficiency level, i.e. grade, and their interaction on English-Italian word pairs similarity judgments, the only statistically significant effect is that of formal similarity. What is interesting is that there is a negative relationship between formal similarity and similarity judgments, i.e. formally more similar word pairs are perceived as less similar. Based on the size of the effect, 28% of the answers’ variance can be explained on the basis of semantic similarity and 22% on the basis of formal similarity. As to the effects of formal similarity, English language proficiency level, i.e. grade, and their interaction on English-Italian word pairs similarity judgments, there is a statistically significant effect of semantic similarity. In other words, highly similar lexical items have received higher similarity ratings. Grade hasn’t been found statistically significant, whereas there is a statistically significant interaction of semantic similarity and grade on word pairs similarity judgments although the effect size is relatively small. While the ratings of the grade 1 subjects tend to be more uniform, grade 4 students rely more on semantic similarity when expressing their similarity judgments. Although the strength of the effect of the interaction between semantic similarity and age on similarity ratings is relatively small, it is interesting to note that the effects of semantic similarity on similarity ratings are different with age. As shown in Picture 1., the range of semantic similarity ratings is larger in the ratings of older students (grade 4) than of the younger ones (grade 1), which is in accordance with the finding that crosslinguistic similarity perception on more advanced levels is based on previously acquired lexical knowledge. The first grade students’ ratings, whose lexical knowledge is on a lower level compared to that of the fourth grade students, are similar to those of the older students, since they have given low similarity ratings for semantically different words and higher ratings for semantically more similar words, but their ratings are in a smaller range, around the middle values on the rating scale, probably due to their fear of giving the “wrong” rating.
In order to test the correlation between similarity ratings and the subjects' first language, a one-way ANOVA was performed, whereby no significant effect of the first language on the overall similarity rating was found (Table 2.). The subjects gave uniform ratings of crosslinguistic similarity regardless of their L1 since the majority of them have Croatian as their L1 and use Italian as an L2 in the schooling context. However, even when they speak Italian in their family, they use Croatian in everyday communication in the broader community. In view of the language specificity of our subjects, this finding is in line with the results of previous studies indicating semantic transfer taking place from the first language or, in the case of high proficiency, from the second language.

Table 2. ANOVA results for the effects of L1 on lexical pairs similarity ratings.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$F_{1,67}$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>first language</td>
<td>1.11</td>
<td>.02</td>
</tr>
</tbody>
</table>

In the analysis of variance of the crosslinguistic similarity ratings and the experience of the learning of the Italian language, no significant effect of the L2 learning experience on the ratings of the similarity of lexical pairs was found (Table 3.). All informants evaluated their experience of learning Italian as very positive or neutral, and no one expressed any negative experiences connected to the acquisition of their L2. Out of 69 students, 29 of them rated the experience of learning Italian as very positive, 20 positive and 20 neutral.

Table 3. ANOVA results for the effects of L2 learning experience on lexical pairs similarity ratings.

<table>
<thead>
<tr>
<th>Factor</th>
<th>$F_{2,66}$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian language learning experience</td>
<td>0.46</td>
<td>.01</td>
</tr>
</tbody>
</table>

The correlation between the ratings of crosslinguistic similarity and the use of the Italian language in the formal (school, church) and informal (family, friends) context has been analyzed using the linear regression analysis. As can be seen from Table 4., the model can't be regarded as statistically significant. The formal context is generally irrelevant with respect to the similarity ratings, while the use of L2 Italian in the informal context is...
context is a negative predictor, although on the limit of statistical significance, of similarity ratings. In other words, the subjects rate lexical pairs as less similar the more they use Italian in informal situations. This could be explained by the fact that those speakers that use Italian with family and friends have Italian as their dominant language, whereby there is no effect of the second language which has been recognised as one of the most influential factors of crosslinguistic influence, and a high language proficiency implies a greater attention given to semantic, and not to formal features of a word.

Table 4. Regression analysis results for the crosslinguistic similarity ratings and the context of use of L2 Italian (formal and informal context).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>Informal</td>
<td>-.26</td>
<td>.05</td>
</tr>
</tbody>
</table>

R²=.6; F_{2,66}=2.12, p>.05

As shown in Table 5., no statistically significant correlation between lexical pairs similarity rating and self-rated knowledge of Italian was found. However, it is significantly connected to the mark at the end of the school year in that subject. In fact, L2 proficiency has been recognised as one of the relevant factors in crosslinguistic influence.

Table 5. Connection between similarity ratings, self-rated knowledge in Italian and the mark in the subject Italian language and literature at the end of the school year.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. similarity rating</td>
<td>.15</td>
<td>.28*</td>
<td></td>
</tr>
<tr>
<td>2. self-rated knowledge of Italian</td>
<td></td>
<td>.33**</td>
<td></td>
</tr>
<tr>
<td>3. formal evaluation in Italian</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<.05; ** p<.01

Similarity ratings don't show a correlation neither with the self-rated knowledge of English nor with the mark at the end of the school year in that subject (Table 6.). However, it is interesting to note that the correlation between self-rated knowledge and the mark at the end of the school year is bigger in the case of L3 English than L2 Italian. This can be explained by the fact that Italian is regarded as the students’ first language, so the teachers set higher expectations and the students themselves are more self-critical with respect to their knowledge of Italian and rate it lower than is the case with English.

Table 6. Correlation between similarity ratings, self-rated knowledge in English and the mark in the subject English language at the end of the school year.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. similarity rating</td>
<td>-.08</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>5. self-rated knowledge of English</td>
<td></td>
<td>.70**</td>
<td></td>
</tr>
<tr>
<td>6. formal evaluation in English</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p<.01
In order to test the effects of word class similarity, etymological similarity and formal and semantic similarity on English-Italian word pairs’ similarity judgments, a series of one-way ANOVAs with repeated measurements on one factor was performed. Word pairs have been divided into two groups according to grammatical category (the same – different), genetic relatedness (the same – different), and membership of the category of deceptive cognates. Effects of all the mentioned variables have been found statistically significant. Words belonging to the same word class are rated as more similar than words belonging to different grammatical categories. Likewise, a statistically significant effect of etymological similarity has been established and in this case the effect size is the largest. True cognates are perceived as more similar than words having a different etymological origin. The effect of the deceptive cognates is also statistically significant. Words of high formal similarity but semantically completely or partially different are rated as less similar (Table. 7.). This finding can be brought into relationship with the first finding (Table 1.) regarding low similarity ratings for formally similar items indicating that the subjects rely more on semantic than formal similarity which is in line with the results of the study we took as our starting point (Utgof, 2008).

Table 7. Effects of word class similarity, etymological similarity and formal-semantic similarity on English-Italian word pairs similarity judgements

<table>
<thead>
<tr>
<th>Factor</th>
<th>$F_{1,68}$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>word class similarity</td>
<td>14.97**</td>
<td>.18</td>
</tr>
<tr>
<td>etymological similarity</td>
<td>61.54**</td>
<td>.48</td>
</tr>
<tr>
<td>deceptive cognates</td>
<td>13.84**</td>
<td>.17</td>
</tr>
</tbody>
</table>

** $p<.01$

The results of this study indicate a statistically significant correlation between crosslinguistic similarity and the experience of learning the Italian language, the context and frequency of use of the Italian language, the length of learning and proficiency in English, self-rated language proficiency in L2 and L3 and formal evaluation. Words belonging to the same word class and of the same etymological origin have been rated as more similar. Whereas in the case of word class membership results can be brought into connection with the lexemes being contextualized and thus facilitating comprehension, explicit knowledge of the etymological origin isn’t available to high school students since at this level the learning and teaching of both English and Italian is approached from a synchronic view, and the mechanism of forming interlingual associations doesn’t depend on the origin of a lexical item but on the perception of lexemes as analogous in the mind of the language user (Olujić and Bošnjak Botica, 2007). However, what should be remembered is that transfer is an internal phenomenon that exists in the minds of individual language users and springs from the interaction of languages stored and processed within the same mind. Perceived crosslinguistic similarity also depends on individual learning styles and the characteristics of each individual learner (Jarvis and Pavlenko, 2008).

Conclusions

Based on the findings of this study, the conclusion is that Croatian-Italian bilingual speakers perceive crosslinguistic similarities between the English and the Italian language, and that their perception is related to both formal and semantic crosslinguistic similarity. As to the words which coincide in both meaning and form, the first grade students rate these words as less similar than fourth grade students since it is necessary to acquire a certain level of language proficiency for transfer to occur and crosslinguistic influence is manifested on more advanced levels of language learning when the learner has acquired additional competences and recognizes similarities between the source language and the recipient language (Wode, 1976; in Jarvis and Pavlenko, 2008). In the case of deceptive cognates the students on the more advanced level of language proficiency rely more on their lexical knowledge and rate deceptive cognates as more different although they are words of high formal similarity, as opposed to first grade students who rely more on formal similarity.

In order to facilitate L3 acquisition, an explicit approach to raising the learners’ awareness of language similarity should be adopted so that the knowledge of all previously acquired languages might get activated with the purpose of promoting positive transfer, enhancing learners’ receptive strategies for inferring word meanings and developing metalinguistic awareness. On the other hand, in the case of deceptive cognates and partial...
translation equivalents there is the need to emphasize semantic and conceptual crosslinguistic differences and to use interlingual comparisons so as to make the students aware of negative transfer. Therefore, CLI implicitly as a reciprocal, multi-directional influence of all the language systems possessed by an individual, and explicitly as a learning strategy, facilitates the process of learning not only the English language, but also for those bilingual speakers whose dominant language is Croatian, the learning of Italian vocabulary by means of lexical similarities with English.

Since this study has included a limited sample it is clear that the findings have a limited generalizability and there is a need for replication that would address a wider range of contexts and include appropriate monolingual controls, especially monolingual speakers whose L1 is Croatian and who study both Italian and English in a formal context as foreign languages and are similar to the subjects of this study on a range of sociodemographic variables.
References


