In the territory of Hungary, there are about 40 settlements which are inhabited by Croats, whose linguistic competence embraces all the three dialects of Croatian (kaj, što and ča). However, their language varieties are a bit different from the Croatian language in Croatia, especially at the lexical level, as a result of being physically distant from the homeland. In this paper, the mental lexicons and speech productions of some representatives of the different Croatian communities living in Hungary will be discussed. The data are gained from examination of their narrative skills which also includes the study of their disfluencies during speech production. The main goal of these experiments is to find empirical data concerning the influences of the two languages on each other.

1. Introduction — the bilingual mental lexicon

The mental lexicon is a kind of internal dictionary which contains not only the ‘entries’ for each word a speaker knows but also all the linguistic information about the word: its semantic content, syntactic properties, phonological shape, etc. The semantic memory, which is reflected in the lexicon, is not strictly linguistic since it contains the mental representation of the individual’s knowledge of the world. This knowledge is represented in concepts and relations between these concepts (Appel and Muysken 1987). Studies of mental structure have never been easy to conduct, and descriptions of cognitive and linguistic development are always subject to inference and uncertainty. These difficulties are compounded when two languages occupy the linguistic domain of the mind (Bialystok 1998).

When studying the semantic representation of bilinguals, the structure of the bilingual mental lexicon and the connections of language, thought and
culture must also be taken into consideration. Languages reflect the culture, beliefs, values and identity of people who speak them (Baker and Jones 1998). Culture is something that everybody has. It is not only what is found in theatres, museums, universities, etc. The term refers to the property of a community which the members share and which might distinguish it from other communities (Hudson 1986).

In monolingual cases, building up the mental lexicon takes place while acquiring L1; however, it is subject to change all through the life. Recalling is a reconstruction of the very mental state when the given information was fixed (Gósy 1998). Different languages make different distinctions explicit and have different patterns of lexicalisation. Macroplanning is language-independent, microplanning is language-specific (Green 1993). In the case of bilinguals the relationship between a given L2 word and a given L1 word in the mental lexicon will vary from individual to individual, depending on how the words have been acquired and how well they are known, and also on the degree to which formal and/or semantic similarity is perceived between the L2 word and the L1 word in question (Singleton 1999).

Now the question is how the mind accommodates two linguistic systems and how reality is reflected through linguistic devices. The study of bilingual speech processing is tightly connected to the presumed storage theories. Attempts have been made to clarify whether bilinguals store information about a word and its associations separately for each language or they process words in terms of their semantic meanings and represent them in one memory store, independently of the language in which they appear. More recently, language fluency has been taken into account when considering the above question, and as a result, according to the hierarchical model of bilingual memory representation (Kroll and Stewart 1994), the single versus dual-store differentiation has been resolved. It has been found that less fluent bilinguals appear to have a dual-store, and the more fluent ones, a single-store conceptual representation. This model proposes that the conceptual store is connected to both L1 and L2 lexicons. However, the connections between the L1 lexicon and the conceptual store are strong and direct, whereas the connections between the L2 lexicon and the conceptual store are weak. Thus, the participant’s L1 is more likely to access the conceptual store directly than the participant’s L2. Heredia (1996) in his Second Revision (R-2) Hierarchical Model suggests using MDL (more dominant language) and LDL (less dominant language) instead of L1 and L2 based on the simple fact that in many cases L2 becomes more dominant than the earlier acquired L1. In this way, MDL has a stronger and more direct connection to the conceptual store regardless of whether it is the L1 or the L2.

The natural flow of speech is always checked by a monitoring mechanism (Levelt 1989), i.e. speakers can detect their own speech errors by
parsing their own inner and overt speech in the same way as they parse the speech produced by others. Thus monitoring is an integral part of the production process concerning lexical selection, phonological encoding, structural building, stylistic selections, resolving semantic ambiguities, etc. If an error is detected in the speech production process, further processing is cancelled, or at least suspended while the error is being repaired. According to the “hold-up” type of monitoring, all errors are repaired covertly, i.e. before articulation (Laver 1980). In contrast, Levelt (1989) argues for a flow-through monitor, which allows for the production process to go on, and thus, prearticularily detected errors may become overt. Levelt’s Main Interruption Rule (mir) poses that the flow of speech is interrupted as soon as error has been detected, and the time between the cutoff and the repair is used for planning the self-repair. Blackmer and Mitton (1991) observed no cutoff time before the repair, which implies that the planning of the repair must have occurred during the flow of speech. Van Hest (1996) demonstrates that in fast speech the monitor is able to keep up with the increased speed of conceptualizing. Certain disfluencies, e.g., filled pauses and repetitions, are regarded as by-products of covert repairs, resulting from errors detected at an early time when the error does not become overt.

In this paper we will examine how the Hungarian-Croatian bilingual mental lexicon is built and how it functions. Data analysed here are part of a larger scale study. In the study, 125 bilingual people were tested. Their languages were typologically different from Hungarian (a Finno-Ugric language). First, in a word association test, the participants were given the task of saying the very first word that came into their minds after hearing 188 Hungarian prime words one after the other. Secondly, a colour naming, and thirdly a picture naming test were carried out with them. The details, data analysis and results of Croatian-Hungarian bilinguals were previously published (Navracsics 2003). Here we will examine the speech disfluencies that can be observed in the speech production of Croatian-Hungarians in both their languages, which is the data gathered by the fourth test used in the study.

2. Participants and methods

2.1. Participants

Five participants were interviewed and tested. Details on participants are shown in Table 1. Three of them represent the Croatian national minority in Hungary, the other two were born in Croatia and now they live in Hungary. One of them (Julianna) is bilingual from birth, born in a Hungarian family near Osijek, the other (Janić) used to be a monolingual Croat, also
from the Osijek region. He settled in Hungary at the age of 25. He is a university student like all the others, except for one elderly woman (Marika), who is from Molnári (Mlinarce). She was born in a monolingual Croatian family in a Croatian village in Hungary, and she has been living there ever since. She started to learn Hungarian when she started school. She married a Croat, raised her children bilingually, but her grandchildren are all Hungarian monolinguals. She claims her language variety represents the kaj dialect. A student (Szabina) from Szentpéterfa (Petrovo Selo) was born in a Hungarian-Croatian mixed family, but the language used at home was only Croatian (the ča dialect), so the girl did not speak much Hungarian before she started school. Finally, we have a student (Zorán) from Hercegszántó (Santovo), who was raised in a Croatian family in a Croatian town in Hungary. He studied at the Croatian-Hungarian bilingual school there, but more recently he has lived in a monolingual part of Hungary where he is a university student. He speaks the što dialect of Croatian.

<table>
<thead>
<tr>
<th>name</th>
<th>age</th>
<th>onset of bilingualism</th>
<th>born</th>
<th>dialect</th>
<th>place of birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marika</td>
<td>70</td>
<td>age 6</td>
<td>HU</td>
<td>kaj</td>
<td>Molnári (Mlinarce)</td>
</tr>
<tr>
<td>Janić</td>
<td>30</td>
<td>age 25</td>
<td>CRO</td>
<td>kaj</td>
<td>Osijek</td>
</tr>
<tr>
<td>Julianna</td>
<td>20</td>
<td>birth</td>
<td>CRO</td>
<td>kaj</td>
<td>Osijek</td>
</tr>
<tr>
<td>Zorán</td>
<td>20</td>
<td>birth</td>
<td>HU</td>
<td>što</td>
<td>Hercegszántó (Santovo)</td>
</tr>
<tr>
<td>Szabina</td>
<td>18</td>
<td>birth</td>
<td>HU</td>
<td>ča</td>
<td>Szentpéterfa (Petrovo Selo)</td>
</tr>
</tbody>
</table>

Table 1. Participants

2.2. The context of the study

The language of Croats living in Hungary is the standard variety of Croatian. However, when it is studied, it should be taken into account that Croats in Hungary do not represent a monolithic community. They settled sporadically in different parts of Hungary, which has influenced the formulation of their linguistic varieties, dialects and identity. The majority of the Croats live along the borders of Hungary with Austria, Croatia and Serbia. These were the devastated territories abandoned by Hungarians during and after the Turkish invasion, and they were suitable for settlement by the Croatians, also escaping from the Turks. At present, there are about 40 settlements with a Croatian population. The national self-government of Croats in Hungary comprises local self-government in 115 settlements. The Croatian National Self-Government was established in 1995, and its main objective is to ensure the representation of the minority’s interests in the country, the maintenance of language and culture, the establishment and maintenance of cultural and educational institutions. It founded the
Croatica Cultural, Information and Publishing Company in 1999. Thus the Croatian minority is the first one in Hungary to have their own publishing company. The main function of CROATICA is the printing and publishing of teaching materials for schools in the Croatian language. This is also the publisher of the weekly Hrvatski glasnik, the journal Pogledi, which deals with the cultural, political and religious life of Croats in Hungary. There are other literary publications, music CDs and cassettes, e.g. U blatu tragovi and Čestit Božić, which contains Christmas carols. Croatian was the fourth biggest national minority among the 13 officially acknowledged ones in Hungary in 2001. With the growing number of minorities, 15 620 people now claim that they are Croatians by nationality. It is important to note, at the same time, that according to the number of people claiming that their L1 is the nationality language, Croats are the third biggest minority in Hungary.

What is a ground for Croats in Hungary to identify themselves as Croats or to claim that their L1 is Croatian? In their interviews, two participants answered the question concerning their national identity by saying that they considered themselves Hungarians because they had always been living in Hungary. Once they are citizens of Hungary, they think they are Hungarians, even though they did not speak any Hungarian before they started school, and even now they speak only Croatian to each other in the village. However, they also admitted that in spite of the fact that they brought up their children bilingually, their grandchildren are monolingual Hungarians, and they accept this fact and do nothing to counter it. They prefer talking to their grandchildren in Hungarian. So in a number of cases, the fact that people were born and brought up in Hungary might influence their identity.

The linguistic dialects of Croatian in Hungary — All the three dialects of Croatian, i.e. što, kaj and ča are present in the Croatian linguistic minority in Hungary. Apart from the presence of all three dialects, some varieties have become more or less distant from the homeland dialects, and it can be witnessed mostly in the field of vocabulary. We will return to this later in the analysis of speech productions of Hungarian Croats.

Institutions with Croatian as the language of instruction — Santovo (Hercegszántó) has been one of the first settlements in Hungary to introduce Croatian as the language of instruction in schools and kindergarten. In the academic year 2006/2007 there were 22 children in the kindergarten and 106 students in the primary school. Apart from this town there are four more towns running educational institutions in Croatian: Martinci (Felsőszentmárton), Petrovo Selo (Szentpéterfa), where there is only preschool and primary education available in Croatian, but in Pécs and Budapest students
can go on with their secondary studies in Croatian. In another 35 towns and villages there is Croatian as a second language in the school curriculum, but there is no bilingual education in these schools.

2.3. Methods

The participants were shown four series of comic strips, each consisting of six pictures, and were asked to make up a story based on each series in each of their languages. The narratives were audio-taped, transcribed, and analyzed. Speech rate measurement was based on spoken syllables per second. Types of speech errors, disfluencies and repetitions were studied in both languages.

3. Results and data analysis

3.1. Speech rate

Table 2. shows the average number of syllables per second uttered in the two languages. Time was measured from the onset of the first syllable to the offset of the last syllable, and the number of syllables was determined. Speech rate was calculated as the number of syllables per second. Pauses were not excluded.

<table>
<thead>
<tr>
<th>Name</th>
<th>(age)</th>
<th>Hungarian</th>
<th>Croatian</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julianna</td>
<td>(20)</td>
<td>2.9075</td>
<td>1.5425</td>
<td>2.225</td>
</tr>
<tr>
<td>Szabina</td>
<td>(18)</td>
<td>3.1575</td>
<td>1.8925</td>
<td>2.525</td>
</tr>
<tr>
<td>Janić</td>
<td>(30)</td>
<td>2.82</td>
<td>2.57</td>
<td>2.695</td>
</tr>
<tr>
<td>Zorán</td>
<td>(20)</td>
<td>3.35</td>
<td>2.8925</td>
<td>3.12125</td>
</tr>
<tr>
<td>Marika</td>
<td>(70)</td>
<td>3.05</td>
<td>2.97</td>
<td>3.01</td>
</tr>
</tbody>
</table>

Table 2. Average number of spoken syllables per person per second

As can be seen from the table, three participants’ speech rates are similar in both languages, which makes us believe that, in the case of bilinguals, it is also the individual rate that counts and which concerns the tempo of the speech production in both languages. However, significant differences can be observed in two participants’ speech rate — both much faster in their Hungarian than in their Croatian. In these cases we might think of language dominance though not in the expected way. Julianna’s speech rate in Hungarian is almost twice as fast as in Croatian despite the fact that she was brought up in Croatia. However, it is also true that she comes from a Hungarian minority family. On the other hand, Szabina comes from a Croatian family in Hungary and studies Croatian at college; still her dominant language seems to be Hungarian.
3.2. Disfluencies: pauses, repetitions and self-repairs

Like with the speech rate, disfluencies occur to the same extent and in the same way in the participants’ narratives. The most frequent type of disfluencies is the pause (both silent and filled), followed by self-repairs and finally repetitions. It should also be noted that only 6.2% of the uttered words included disfluencies. Figure 1. presents the percentages of the occurrences of speech disfluencies in the five participants’ speeches.

![Disfluency rates](image)

Figure 1. Disfluency rates

4. Discussion

4.1. Silent and filled pauses

The most frequently occurring disfluencies were silent or filled pauses. When bilinguals speak, both their languages are activated, and so the path the participants are to go along for the retrieval of the correct word, for phonological encoding or the selection of the required grammatical structure is much more complicated to get through. Along this path, monitors are distributed. If an error is detected during one of the stages of the speech production process, further processing is cancelled. Some disfluencies, like pauses and repetitions, are regarded as by-products of covert repairs, detected and corrected at such an early point that the error does not become overt (Oomen and Postma 2001). Another possible explanation is that disfluencies are the results of planning or temporal problems in speech production. At this point, it is hard to determine whether disfluency in overt speech has resulted from covert repairing activities or from other planning difficulties. The transcript of a narrative with silent and filled pauses (ööö = filled pause, // = silent pause) is given in (1).
4.2. Repetitions

When fluency breaks down in speech, speakers frequently repeat words. Au-Yeung, Howell and Pilgrim (1998) have shown that word repetition happens on function words, in general, and thus function words have a similar role to that of pausing. Fluency failures on content words may occur when the speaker commences saying the word when only the first part of the plan is ready (Howell et al. 1999).

Contrary to Au-Yeung, Howell and Pilgrim’s findings, in our data repetition concerns content words, too. In some cases, repetition includes a segment or other items as in (2).

(2) i. word-part, e.g. cu-cu caka, gostiteljica iš išla, pokaz pokazila, ras rasprima
   ii. word, e.g. i brzo otišao otišao kući, jako // ööö // // jako glasan, zato // ööö // // zato
   iii. function word, e.g. Do do na mjesta, ulazi u u kadu, kako se, kako se pas kupi

These stuttering-like repetitions on content words occur when the speaker is not totally ready with the planning process (Howell and Sackin 2001).

4.3. Word substitutions

The process of speech production consists of message generation or conceptualization, formulation and articulation (Levelt 1989). Lexicalization occurs in two stages. First lemmas are accessed, which are an intermediate level of abstract lexical items. Only then can the phonological forms be accessed. At present, there is no consensus as to whether the two stages of processing are discrete or whether they overlap. Errors may occur at both stages, and whole-word substitution speech error studies (Harley and MacAndrew 2001) distinguish two types of errors: one where the target and the intrusion are related in sound and another where they are related in meaning. Lexical access may fail and may result in a word substitution in two ways. An incorrect branch of the semantic net may be chosen, and so a semantic associate of the target is erroneously selected, leading to a semantic paraphasia.
Alternatively, if the final pointer to the target phonological form slips, a nearby item will be substituted, leading to a phonological paraphasia. If the error emerges as early as accessing lemmas from a semantic-conceptual input, semantic errors will have to be repaired or substituted. Harley and MacAndrew (2001) distinguish between associative and shared-feature semantic substitutions. Shared-feature errors are those where the target and the intrusion are from the same semantic field or are hierarchically related. Associative errors are those where there is only an association between the target and the intrusion. Word substitution errors are important because they provide a window on the processes of lexicalization.

In our participants’ narratives the following disfluencies, listed in (3), (4), (5) and (6), were observed concerning the lexicalization process.

(3) Filling in the missing word in the repaired production
   i. Da bi se // da bio kupa se
   ii. cuc // svoju cucku

(4) Associative substitutions
   i. gostiteljica — an associative error, the word ‘restaurant’ is substituted for the word ‘waitress’; see example in (1)
   ii. jednomu // drugomu človjeku (= a man, another man)
   iii. Čača kani se šjetat ali zabila je zabio je doma cucku // kad je kad vidi // jednomu drugomu človjeku cuc // svoju cucku onda ide domoa ali već cuck je // ustaji i ona jako // paživ.

(5) Shared-feature semantic substitutions
   i. Dobro ako me razumeš onda sve u redu, evo, prva situacija, dječko tjera je svoju kera, i gleda da // kera psa i čovek veli da bi se, da bio kupa se u kadi i nešto // pokaz pokazila, i za time se ras rasprim a se i ulazi u u u kadu, i dotle sve gledaju i dječak i kera gledaje kako se ovaj stari kupa za time jedan akvarij uglane ovaj ovaj pas i skoči u tom akvarij i gera htio pojesti ribe koji su bile unutri, i zatim čita ova obitelj gleda kako se, // kako se pas kupi u akvariju. (standard word ‘pas’ vs. dialect word ‘kerja’)
   ii. sa prikolicom, kampkučicom (both vehicles are attached to the car but the first one on the side, the second — at the back)
   iii. Naslov izlet u // lošem vremenu, porodica ide na izlet, sa mama, tata i sve djece s jednim psom, // ustupali su jedan auto, trčići u auto svi sluši samo su // izaborali psa žaluci vratio se // gazda // potiše // stojeći na putu prolazi auto // sa prikolicom // // / kampkučicom gdje bio je pas gledajući kako je ima mjesto i bolje mu // ide.
Due to feedback connections between the phonological and lexical units, phonological similarity between two words which do not show any semantic relationship may trigger phonological paraphasia (Askari 1999), as in the examples in (6). The Croatian word ‘kuća’ means ‘house’, and its phonological shape is very similar to the Hungarian word ‘kutya’, which means ‘dog’.

This example of cross-language word substitution supports the single-store and, partially, the hierarchical model of bilingual language representation. Both lexicons are tapped, and the substitution happens not because of the semantic, but rather due to phonological similarity. There seems to be a common semantic and phonological system in balanced bilinguals, which could be tapped independent of language.

4.4. Errors at the formulation stage — self-repairs

In our data, self-repairs were the largest number of filled disfluencies. Self-repairs are made in order to correct errors detected by the monitor at the conceptualization, formulation stages of production. During the formulation stage, some competing structures might interfere, causing errors in the surface structures, as in the examples (7) through (12).

Number agreement problems — Substitutions for number agreement are not evidence of grammatical incompetence; they are rather slips of the tongue, as in (7).

(7)  
   i. Gdje je bila, gdje su bili
   ii. za stenovanje primetili su da primetio je da je pas

Gender related problems — Problems related to gender might be due to the influence of the Hungarian language (which does not differentiate between genders only at the lexical-semantic level in some words), as in (8).

(8)  
   i. zabila je zabio
   ii. jedan, jedna // žena
Case-marking errors — Errors related to case-marking come up in the speech of bilinguals in whose language suffixation is a productive way of combining words into sentences. They may show some lack in the grammatical competence of the speakers, as in (9). In these examples there are faulty agreements between nouns and numerals or adjectives.

(9)  

i. *Pod jednom, pod jednim*  

ii. *u drugim u drugom*  

iii. *Jednom, jednime prikljucama*

Analytical structures — Preference for the use of analytical structures in Hungarian instead of the correct synthetic ones, as in (10). In Hungarian, the analytical structure *ember kutyával* is not acceptable, it should be ‘kutyás ember’ (doggy man).

(10) *lát egy másik ember kutyával, aki ugrik utána, nagyon boldogok*  

  (he sees another *man with a dog*, who jumps after him, they are very happy)

Hypercorrection — The base grammatical forms are put together in Hungarian under the influence of Croatian in (11), instead of the correct ‘ez alkalommal’.

(11) *ezzel az alkalommal nem szeretné magával vinni a kutyát*  

  (this time he doesn’t want to take the dog with him)

As a result of the intruding competing structures in the formulation phase, contaminations may happen, as in example (12) where the structures ‘túl hangosra meg idegesítőre sikerült’ and ‘túl hangossá meg idegesítővé válik’ were mixed.

(12) *csak ez a koncert kicsit túl hangosra, meg idegesítővé válik*  

  (but this concert has become too loud and annoying).

The above examples prove that in the bilinguals’ speech productions there may be both intralingual and cross-linguistic errors which may or may not be repaired by the speakers. As in the course of fast speech, speakers do not realize that some speech errors spoil their speech production in one language, they will not recognize the cross-linguistic errors either, and this is the most natural evidence of the compound character of their mental lexicons.

5. Conclusions

The general tendency observable in Hungary as concerns minorities policy is valid for the Croatian national minority as well. While the number of people belonging to the Croatian minority has been increasing in terms of
their national identity in the past 30 years, the number of those claiming
their first language is Croatian has been decreasing or stagnating. Thus,
linguistic assimilation is well underway. However, the existence of educa-
tional institutions with Croatian as the language of instruction, may give
some hope for the national minority to become strong and be successful in
their language maintenance efforts.

In the speech production of Croatian-Hungarian bilinguals who use both
their languages in their everyday lives many features of bilingual speech can
be observed. These features are unknown or unrecognizable to monolinguals,
and so they do not know the reason why Hungarian Croats speak differently.
It remains an open question whether disfluencies in overt speech result from
covert repairing activities or from the two languages competing with each
other in the planning processes. However, our data show differences in terms
of repetitions. Some studies reveal that repetition concerns mainly function
words. In our study, there is evidence that repetitions occur when uttering
content words, too. This may confirm our suggestion that the speaker is not
ready with the planning processes.

Cross-linguistic influences and their transfer during speech production
are the natural by-products of bilingualism. The person who has two lin-
guistic systems (or one shared?) in his or her mind has much more compi-
lcated paths to go along while processing speech. Semantic and phonological
paraphasia, formulation errors, self-repairs — all provide evidence of the
compound character of the bilingual mental lexicon. The quality of the
speech production of bilinguals is very close in both languages. The speech
rate is almost identical (if not, dominant bilingualism is presumed) in an
individual’s production of the two languages. Disfluencies, such as pauses,
repetitions and re-starting are represented equally in the production of both
languages.

In Europe and North America, monolinguals have had a contemptuous
view of bilinguals since the beginning of the last century, as a result of faulty
psychological tests, unfair circumstances of language testing of bilingual
working-class children and their comparison to monolingual middle-class
ones, etc. (cf. Grosjean 1992). It should be the task of linguistic institutions
and teacher training colleges to explain the essence of bilingualism. The
result should be that monolinguals would not feel comfortable looking down
upon others whose linguistic configuration is different. If we overcome this
bias of ours, the number of bilingual people will grow all across the continent,
and the revival of national minorities in different countries will be more
successful.
6. References


Hrvatsko-mađarski umni rječnik

U Mađarskoj postoji četrdesetak mjesta naseljenih Hrvatima, čiji govori obuhvaćaju sva tri hrvatska narječja (kajkavsko, štokavsko i čakavsko). Međutim, uslijed prostorne udaljenosti od domovine njihovi su idiomi donekle različiti od hrvatskoga jezika u Hrvatskoj, posebno na leksičkoj razini. Građa za istraživanje njihova jezika prikupljena je psiholingvističkim pokusima poput testa asocijacije riječi, testa imenovanja boja, testa imenovanja slika i raščlambe njihovih pripovjedačkih vještina. Glavna je svrha tih pokusa pronaći empirijske podatke koji bi pokazali utjecaj jednoga jezika na drugi. U ovome radu raščlanila govorna proizvodnja nekoliko predstavnika različitih hrvatskih zajednica koje žive u Mađarskoj, uključujući njihove zastoje tijekom govorne proizvodnje. Na njezinu su se temelju donijeli zaključci o obilježjima umnoga rječnika ispitanika.

Ključne riječi: umni (mentalni) rječnik, govorna proizvodnja, hrvatski jezik, mađarski jezik

Key words: mental lexicon, speech production, Croatian, Hungarian