Croatians grapheme encoding at the beginner level of learning Croatian as L2 by Farsi speakers - a case study

Sandra Marman[1] and Tanja Diklić[1]

[1]Faculty of Humanities and Social Sciences, University of Rijeka

April 23, 2024

Abstract

This study investigates grapheme encoding in Croatian as a second language among Farsi speakers after twenty hours of learning. Three phases of encoding tasks were administered: 1) dictation of individual phonemes, 2) dictation of words beginning with those phonemes, and 3) dictation of simple sentences with words from the previous phase. Respondents used "-" to denote unencoded items. Eleven Afghan respondents at the beginner level (A1 according to CEFRL), aged 18 to 63, were sampled conveniently. The study aims to assess: a) accuracy in encoding individual graphemes and words, b) problematic graphemes, and c) accuracy in encoding complete words. The results will illuminate initial decoding specifics for this group, confronting the added complexity of differing graphic systems between L1 and L2. Furthermore, implications for Croatian orthography acquisition as L2 will be discussed. Analysis of encoding by Farsi-speaking Croatian learners showed overall success with sentences but difficulty with individual graphemes, possibly due to reliance on lexical rather than phonological knowledge. Notably, struggles were observed with "nj," unlike with "c" as seen previously. This might be because "nj" is less common in Croatian, especially early on. Transfer errors from Farsi, like omitting short vowels, were evident. Instruction for Farsi learners should focus on specific grapheme errors such as "nj," "ć," "dž," and "đ," as well as consonant clusters and short vowel encoding in Croatian.

Sandra Marman
Faculty of Humanities and Social Sciences, University of Rijeka
sandra.marman@ffri.uniri.hr
Tanja Diklić
Faculty of Humanities and Social Sciences, University of Rijeka
tanja.diklic@ffri.uniri.hr

Croatian graphemes encoding at the beginner level of learning Croatian as L2 by Farsi speakers – a case study

This study investigates grapheme encoding in Croatian as a second language among Farsi speakers after twenty hours of learning. Three phases of encoding tasks were administered: 1) dictation of individual phonemes, 2) dictation of words beginning with those phonemes, and 3) dictation of simple sentences with words from the previous phase. Respondents used "-" to denote unencoded items. Eleven Afghan respondents at the beginner level (A1 according to CEFRL), aged 18 to 63, were sampled conveniently. The study aims to assess: a) accuracy in encoding individual graphemes and words, b) problematic graphemes, and c) accuracy in encoding complete words. The results will illuminate initial decoding specifics for this group, confronting the added complexity of differing graphic systems between L1 and L2. Furthermore, implications for Croatian orthography acquisition as L2 will be discussed. Analysis of encoding by Farsi-speaking Croatian learners showed overall success with sentences but difficulty with individual graphemes, possibly due to reliance on lexical rather than phonological knowledge. Notably, struggles were observed with "nj," unlike with "c" as seen previously. This might be because "nj" is less common in Croatian, especially early on. Transfer errors from Farsi, like omitting short vowels, were evident. Instruction for Farsi learners should focus on specific grapheme errors such as "nj," "ć," "dž," and "đ," as well as consonant clusters and short vowel encoding in Croatian.

Keywords: grapheme, phoneme, decoding, Farsi speakers, Croatian as a second language

1. Introduction

Individuals quickly acquire their native language through exposure and direct communication (listening and speaking). However, merely being exposed to a foreign language isn't sufficient for learning; one must acquire all language skills: listening, speaking, reading, and writing. Research suggests literacy impacts grammar acquisition, as some language forms are encountered primarily in written language (Dąbrovska, 2008; Jelaska, 2012). Unlike listening, writing requires direct teaching, especially in alphabetic systems where graphemes may represent phonemes differently (Jelaska, Musulin, 2011). Early writing in a foreign language can be challenging due to differences in phonetic and writing systems from one's native language. Croatian, with its shallow orthography, facilitates encoding what is heard for beginners (Grgić & Udier, 2012). Similarly, Persian (Farsi) also has a predictable relationship between graphemes and phonemes, though with some exceptions (Baluch, 2005).

Acquiring a foreign language's writing system becomes more complex when the graphic systems of the native language and target language differ. Farsi speakers, for instance, encounter added challenges when learning Croatian, which uses the Latin script. While all Latin scripts stem from classical Latin, they adapt to specific languages. For instance, Musulin and Jelaska (2011) note that Spanish and Croatian Latin scripts differ despite sharing the same origin. Such disparities are even more pronounced between writing systems from different sources. Unlike Croatian, Farsi uses a modified Arabic script written right to left, with no distinction between upper- and lower-case letters, mostly connected, with exceptions (Baluch, 2005; Alipour et al., 2019).

The aim of this study is therefore to analyse the difficulties in transferring phonemes into graphemes in relation to the respondents' native language. The results of the study will have direct implications for the teaching of Croatian as a second language to speakers whose native language is Farsi.

2. Research on Decoding

According to ZEROJ, speakers at A1 and A2 levels can reproduce familiar words and short sentences but struggle to produce independent, coherent texts. Udier (2017: 212) describes A2 spelling competence in Croatian as recognizing and writing most sounds and short words accurately, but errors occur, particularly with sounds like č, dž, ć, and đ. Native speaker errors are

common, influenced by their first language (e.g., Cvikić, Bošnjak 2005). Difficulty varies depending on the learner's native language, especially with phonemes like /s/, /z/, and /c/. Writing, a crucial language skill, is acquired through learning, involving activities ranging from simple writing to composing complex texts (Yahya et al., 2012). Dictations are valuable for assessing language reception and decoding skills (Jelaska, Kekelj, & Šafarić, 2007). Errors stem from incomplete language acquisition or transfer from the native language (Jelaska & Barbaroša-Šikić, 2007). Orthographic competence in Croatian was researched by Grgić and Udier (2012) at the B1 level, showing vocabulary mastery's influence on orthographic acquisition. Jelaska and Musulin (2014) compared Croatian and Spanish phonological systems, finding similarities and complexities in mastering phoneme encoding for pronunciation and writing. Jukić, Diklić, and Prosenjak (2022) studied how French beginners learn Croatian graphemes. They found that learners struggled most with graphemes in words and sentences, but were successful with individual graphemes. Jukić (2022) focused on the transcription of the phonemes /c/, /z/, and /s/ in Croatian as a second language, specifically among Spanish-speaking beginners. The analysis revealed that the most difficulties occurred with the transcription of the /z/ phoneme, with the most common mistake being replacing it with /s/, indicating a lack of mastery in voicing contrast. Similarly, for the /c/ phoneme, the majority of incorrect transcriptions were written with /s/. Participants were most successful in transcribing the /s/ phoneme, with the most frequent incorrect substitution being with /z/.

3. Research on Foreign Language Acquisition by Farsi Speakers

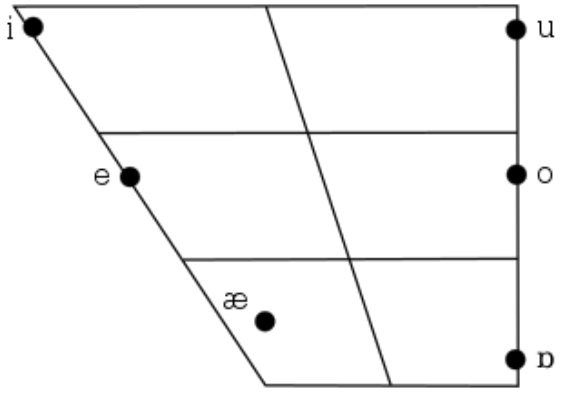
Research over the past twenty years has focused on the challenges Farsi speakers face in learning foreign languages, particularly English. Baluch (2005) discusses how Persian orthography impacts literacy, noting Farsi speakers' struggle with omitting short vowels in writing, leading to pronunciation errors and misunderstandings due to the significance of vowels in word meanings. Grapheme-phoneme inconsistencies in Farsi pose challenges in writing, affecting foreign language acquisition (Baluch, 2005). Gholamain and Geva (1999) found that English proficiency correlates with better Farsi acquisition among Persian immigrants' children, attributing this to cognitive and linguistic abilities rather than orthographic depth alone. Omidipour (2014) identifies Farsi speakers' errors in English acquisition, linking them to differences in phonemes and graphemes

between the languages, categorized as transfer or developmental errors. Khorasgani et al. highlight pronunciation challenges in English acquisition for Farsi speakers, such as the absence of consonant clusters at the beginning of syllables in Farsi. Alipour et al. (2019) discuss common errors, including epenthesis, vocalic transfer, and consonantal replacement, emphasizing the importance of understanding phoneme and grapheme differences between native and foreign languages for error reduction. Limited research exists on Farsi speakers' acquisition of Croatian as a second language (Diklić, 2022), focusing on phonological deviations in written discourse at the beginner level. The analysis reveals significant deviations, particularly in phoneme substitution, omission, and addition, often influenced by the phonological system of the speakers' native language.

4. Comparison of Vowel and Consonant Systems in Farsi and the Croatian Language

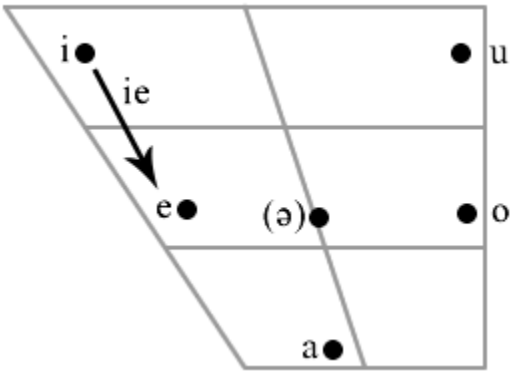
The Croatian and Persian languages, both belonging to the Indo-European language group, exhibit significant differences in their alphabet, vowel and consonant systems, and syllable structure. While both languages feature shallow orthography—where one grapheme corresponds to one phoneme - in Persian, multiple graphemes can represent a single phoneme (Baluch, 2005). Persian, also known as Farsi, falls within the Indo-Iranian language group and serves as the official language in Iran, Afghanistan, and Tajikistan. Although these variants share few linguistic distinctions, communication among speakers remains smooth. The Persian alphabet, derived from Arabic script, is used in Iran and Afghanistan, while Tajikistan employs a Cyrillic-based alphabet. With 32 letters—28 from Arabic and four additional consonants—the Farsi alphabet comprises eight basic letter forms, with variations formed by diacritic signs (Baluch, 2005; Alipour et al., 2019). Conversely, Croatian, a member of the Balto-Slavic language group, serves as the official language in Croatia and holds official status in Bosnia and Herzegovina and the European Union. Employing the Latin script, the Croatian alphabet consists of 30 letters, some of which have been added or modified. Regarding vowels, while Old Persian boasted eight vowels, modern Farsi has reduced them to six—three short (a, e, o) and three long (â, i, u). Short vowels are often omitted, except by inexperienced writers, resulting in phonemes lacking corresponding graphemes. Additionally, most linguists recognize two diphthongs in the system: /ei/ and /ou/ (e.g., Windfuhr, 1979; Khorasgani et al., 2015, as cited in Samareh, 2000; Baluch, 2005).

Figure 1 Farsi vowel system, IPA, 1999, pg. 125



The Croatian language has five vowels (a, e, i, o, u) and all of them can be either short or long. Unlike in Farsi both, short and long, vowels are written down.

Figure 2 Croatian language vowel system, Landau et al. 1999, pg. 67



As far as consonants are concerned, there are twenty-three consonants in Farsi consonant system. The syllables in the Farsi can have one of the following three forms: CV (consonant + vowel), CVC (consonant + vowel + consonant), and CVCC (consonant + vowel + consonant + consonant), which proves that a consonant cluster can never be at the beginning of the syllable. Two consonants can only be at the end of a syllable (Khorasgani et al., 2015; Alipour et al., 2019). Figure 3 shows the Farsi consonants regarding the manner and place of their articulation.

Figure 3 Farsi consonant system, IPA, 1999, pg. 124

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b		t d			k g	q	
Nasal	m		n					
Fricative		f	s z	ʃ ʒ		x ɣ		h
Affricate				t͡ʃ d͡ʒ				
Tap or Flap			ɾ					
Approximant	w				j			
Lateral approximant			l					

The Croatian language has 25 consonants. Unlike Farsi, Croatian consonant clusters can have up to six consonants, sometimes even seven. However, only four Croatian consonants in a cluster could be non-syllabic, the clusters that consist of five or more consonants always have one syllabic consonant: vibrant r (Jelaska, Lalli-Paćelat, 2014). If Croatian and Farsi vowel and consonant systems are compared, it can be concluded there are no letters *c*, *ć*, *đ*, *lj*, and *nj* in Farsi

Figure 4 Croatian language consonant system, Landau et al. 1999

		Labial	Dental-Alveolar	Alveolar-Palatal	Palatal	Velar
Stop	*Voiceless*	p	t			k
	Voiced	b	d			g
Affricate	*Voiceless*		ts	tʃ	tɕ	
	Voiced			dʒ	dʑ	
Fricative	*Voiceless*	f	s	ʃ		x
	Voiced	v	z	ʒ		
Nasal		m	n		ɲ	
Liquid			l r		ʎ	
Glide					j	

5. Research Methodology

The analysis of the conversion of spoken units into a written text in Croatian as a foreign language by Farsi-speaking respondents was conducted during the summer term of the academic year 2021/2022 in three phases. In the first phase, only the individual phonemes were dictated (only consonants: *b, c, č, ć, d,dž, đ, f, g, h, j, k, l, lj, m, n, nj, p, r, s, š, t, v, z, ž*), in the second, the words beginning with the phonemes dictated in the first phase (*baka*, *crkva*, *čarapa*, *ćirilica*, *djed*, *džem*, *đak, flauta*, *gitara*, *hrana*, *kava*, *lav*, *ljubav*, *majka*, *novine*, *njiva*, *pismo*, *robot*, *sestra*, *šuma*, *torba*, *vlak*, *zastava*, *žaba*)[1], and in the third phase, the simple sentences which contain words from the second phase of the research were dictated (i.e., *To je moja baka.*, *Tamo je crkva.*, *To je čarapa.*, *Ovo je ćirilica.*). The respondents wrote down the graphemes, words, and sentences one under the other, using the symbol minus (-) for the graphemes, words, or sentences they did not encode. The convenience non-probability sampling of eleven respondents from Afghanistan was used for this research, all being grown-up attendees at the beginning level of language acquisition (the A1 level according to CEFRL). The youngest respondent was 18, and the oldest was 63. They were informed that participation in the research was voluntary and that they could give up at any point. This research aimed to define the extent to which the respondents encode the Croatian graphemes correctly based on the listening to a dictated text. Concerning the research aim, the following research questions were formulated: What is the level of accuracy in encoding the individual graphemes and graphemes in words and sentences?; Which graphemes will cause the most problems in encoding?; What is the level of accuracy in encoding the whole word?. The analysis of the collected data follows the elements of direct interpretation proposed by Stake (1995) for data analysis in the case study research. The tables present individual data, simultaneously searching for similarities and differences between the individual cases (Creswell, 2007, p. 163).

[1] The words used in the research are part of the vocabulary for the beginner level.

6. Research Results and Discussion

6.1. *Accuracy in Encoding Individual Graphemes*

It is evident from Table 1 that the foreign language learners at the beginner level had no difficulties in encoding individual graphemes *b*, *f*, *z*, and *ž* because they all wrote them down correctly. The high accuracy percentage is also evident in encoding the individual graphemes *c*, *nj*, and *lj*, even though they do not exist in Farsi. That could mean the respondents became aware that those phonemes and graphemes do not exist in their native language but can recognize them in the Croatian language, primarily when dictated individually. Respondents had significant problems with graphemes *đ* and *dž*, whose percentage and frequency are pretty low. Those results are consistent with previous research, confirming the highest accuracy level of encoding the phonemes which exist in both languages and have similar pronunciation. More problems arise with phonemes that do not exist in the native language (Musulin, Jelaska, 2014; Omidipour, 2014).

Table 1 Accuracy in Encoding Individual Graphemes

		Accurately encoded grapheme	**Inaccurately encoded grapheme**	**Grapheme not encoded**
B	f	11	0	0
	%	100	0	0
F	f	11	0	0
	%	100	0	0
Z	f	11	0	0
	%	100	0	0
Ž	f	11	0	0
	%	100	0	0
D	f	10	1	0
	%	91	9	0
J	f	10	1	0
	%	91	9	0
K	f	10	1	0
	%	91	9	0
N	f	10	1	0
	%	91	9	0
P	f	10	1	0
	%	91	9	0
S	f	10	1	0
	%	91	9	0
Š	f	10	1	0
	%	91	9	0
T	f	10	1	0
	%	91	9	0
L	f	10	0	1
	%	91	0	9
C	f	9	2	0
	%	82	18	0
NJ	f	9	2	0
	%	82	18	0
V	f	9	2	0
	%	82	18	0
M	f	9	1	1
	%	82	9	9
R	f	9	1	1
	%	82	9	9
H	f	8	3	0
	%	73	27	0
LJ	f	8	3	0
	%	73	27	0
G	f	7	3	1
	%	64	27	9
Č	f	6	5	0
	%	55	45	0
Ć	f	6	5	0
	%	55	45	0
Đ	f	3	8	0
	%	27	73	0
DŽ	f	2	9	0
	%	18	82	0

6.2. *Accuracy in Encoding Graphemes within a Word*

While encoding the graphemes in words, the respondents had minimal difficulties with graphemes *b* and *f*. However, Table 2 shows that all respondents correctly encoded graphemes *d*, *j*, *k*, *m*, *n*, *p*, *s*, and *t* as well, which was not the case with the individual graphemes. Those are all phonemes and graphemes that exist in the respondents' native language. Most difficulties arose when encoding graphemes in words containing grapheme *dž* since only one respondent wrote it down correctly. At the bottom of the table is the grapheme *nj*, which nobody wrote down correctly. In contrast to Table 1, when *nj* as an individual grapheme did not cause significant difficulties, writing it down within the whole word became challenging. That can point to the inability to segment specific phonemes in context and be connected to orthographic depth, i.e., the fact that learners in word recognition within the language with shallow orthography rely on phonology. It means that foreign language learners whose native language has shallow orthography will be more successful in encoding the individual phonemes rather than words (Katz & Frost, 1992)[2]. Additionally, the phoneme and grapheme nj are not so frequent in words at the beginner level of learning[3] Croatian, which can affect the inability to acquire that phoneme and grapheme in different contexts, i.e., completely.

Table 2 Accuracy in Encoding Graphemes in Words

		Accurately encoded grapheme	**Inaccurately encoded grapheme**	**Grapheme not encoded**
B	f	11	0	0
	%	100	0	0
D	f	11	0	0
	%	100	0	0
F	f	11	0	0
	%	100	0	0
J	f	11	0	0
	%	100	0	0
K	f	11	0	0
	%	100	0	0
M	f	11	0	0

[2]In languages with deep orthography, the speakers (in reading and writing) use their lexical knowledge and word meaning to read or encode a specific word successfully (e.g., Ellis et al., 2004).

[3]For example, in the vocabulary bank at the end of the textbooks Hrvatski za početnike and Dobro došli 1, which are intended for foreign language learners at beginner level, there are only a few words under entries beginning with the grapheme nj. In the textbook Hrvatski za početnike these are the following words: *njegovati*, *nježnost* and *njega*, while in Dobro došli 1 these are the possessive adjectives *njegov*, *njezin* and words *njega*, *njegovati*, *njihati*, *njihaljka* and *njiva*.

	%	100	0	0
N	f	11	0	0
	%	100	0	0
P	f	11	0	0
	%	100	0	0
S	f	11	0	0
	%	100	0	0
T	f	11	0	0
	%	100	0	0
Č	f	10	1	0
	%	91	9	0
H	f	10	1	0
	%	91	9	0
V	f	10	1	0
	%	91	9	0
Z	f	10	0	1
	%	91	0	9
L	f	9	2	0
	%	82	18	0
LJ	f	9	2	0

	%	82	18	0
R	f	9	1	1
	%	82	9	9
Š	f	9	1	1
	%	82	9	9
Ž	f	9	1	1
	%	82	9	9
C	f	7	4	0
	%	64	36	0
Đ	f	7	4	0
	%	64	36	0
G	f	6	5	0
	%	55	45	0
Ć	f	3	7	1
	%	27	64	9
DŽ	f	1	10	0
	%	9	91	0
NJ	f	0	11	0
	%	0	100	0

6.3. *Accuracy in Encoding the Whole Word*

Table 3 shows that while encoding the whole words, all respondents encoded word *baka* correctly. They were less successful in encoding the words *čarapa* and *jakna* (91%). One respondent encoded the word *čarapa* as *ćarapa*[4],and *jakna* was encoded as *jekna*. Nine respondents (82%) correctly encoded the words *kava, ljubav*, and *zastava*. As for errors, two respondents encoded the word *kava* as *kawa,* possibly indicating the influence of some other language systems. Word *ljubav* was encoded inaccurately as *lubav* and *lijubav*, indicating that respondents did not acquire the written form of diphthong *lj*. Word *zastava* is encoded inaccurately as *zastave*, while one respondent did not encode it at all.

Eight respondents (73%) encoded the words *pismo*, *sestra*, *šuma*, and *žaba* accurately. Word *pismo* was encoded inaccurately as *pesmo, pesmoj* and *pisma*; *sestra* was encoded as *sestar*, *sestva*,and *sistra*; *šuma* was encoded as *šoma*, Šuma, while one respondent did not encode it at all; *žaba* was encoded as *žapa*, *zaba*, and only one respondent did not encode it at all.

[4]As well as in some earlier research on writing letters č and ć by native and non-native speakers of Croatian (Jukić, 2018; Đurđević, 2020), the results of this research point to the frequent substitutions of those two phonemes in writing.

Seven respondents (64%) encoded the words *hrana*, *majka*,*novine* and *vlak* accurately.Word *hrana* was encoded as *herana* (N=2), *harne*, and *karana*; *majka* as *majke*, *mauka*, *mayka*,and *maika*; *novine* as *novina* (N=2), *novena*, and *noveni*, and *vlak* as *blak*, *velak*, *vulak*, *velika*. Six respondents (55%) encoded the words *lav* and *robot* accurately, others encoded word *lav* as *ljva*, *love* (N=2), *lov*, and *ljv*, and *robot* as *rrubot*, *yobut*, *rubot* (N=2), while one respondent did not encode it at all. Five respondents (46%) encoded the words *đak* and *crkva* accurately. *Đak* was encoded as *jak* (N=3), *đack* (N=2), and *đzak*, and *crkva* as *serkova* (N=2), *srcov*, *carkva*, *cerkva*, *serckva*. Four respondents (36%) encoded the words *flauta*, *gitara*, and *torba* accurately. *Flauta* was encoded as *flavta*, *plota*, *flu*ta, *fluota*, *flawta*, *flouta* (N=2); *gitara* as *ketera*, *kitara*, *qitara* (N=2), *gittara* (N=2), and *getara*; word *torba* was predominantly encoded as *turba* (N=2), and one respondent encoded it as *tauba*.

The words *djed* and *džem* are at the bottom of the table, with one accurate encoded entry (9%) and the words *ćirilica* and *njiva*,which were not encoded correctly by any of the research respondents. Word *djed* was encoded as *dijet* (N=4), *did*, *djad*, *djet*, *dejd*, *djat*, and *dead*; word *džem* was encoded as *jame*, *džam*, *june*, *qym* (two respondents), *đzem* (two respondents), *djem*, *game*, *đem*. Word *ćirilica* was encoded as: *čirelatica, ćeirnlijce, čeriltca, čjerilica, ćrilica* (N=2), *čerilica*, *čereljća*, *čirilica*, and *čeriljtca*, and one respondent did not encode it at all. Word *njiva* was predominantly encoded as *niva* (N=8), the second variant being *neva* (N=3). Other challenges in encoding words *ćirilica* and *njiva* are consistent with the results of previous research, which prove that the absence of some phonemes and graphemes in the native language causes specific errors.

Furthermore, in the analysis of phonological deviations of Farsi speakers at the beginner level of learning Croatian, Diklić (2022) also provides examples where it is evident that the respondents have not fully mastered the grapheme "lj" in words such as "prijatelj" (preitel, piryte, prejatel, prijatil, preitel, preyjette) or "obitelj" (obetil, obiteji, obitlj, obitel). Diklić (2022: 213) suggests that the respondents may have first heard these words without seeing them written down, mastering pronunciation but not the written form of the words. This is supported by research results indicating that "nj" as a separate grapheme did not pose significant difficulties, but when writing words containing the "nj" grapheme, it did cause difficulties (Jukić and Diklić, 2023).

Table 3 Accuracy in Encoding the Whole Word

		Accurately encoded word	Inaccurately encoded word	Word notencoded
baka	f	11	0	0
	%	100	0	0
čarapa	f	10	1	0
	%	91	9	0
jakna	f	10	1	0
	%	91	9	0
kava	f	9	2	0
	%	82	18	0
ljubav	f	9	2	0
	%	82	18	0
zastava	f	9	1	1
	%	82	9	9
pismo	f	8	3	0
	%	73	27	0
sestra	f	8	3	0
	%	73	27	0
šuma	f	8	2	1
	%	73	18	9
žaba	f	8	2	1
	%	73	18	9
hrana	f	7	4	0
	%	64	36	0
majka	f	7	4	0
	%	64	36	0
novine	f	7	4	0
	%	64	36	0
vlak	f	7	4	0
	%	64	36	0
lav	f	6	5	0
	%	55	45	0
robot	f	6	4	1
	%	55	36	9
crkva	f	5	6	0
	%	45	55	0
đak	f	5	6	0
	%	45	55	0
flauta	f	4	7	0
	%	36	64	0
gitara	f	4	7	0
	%	36	64	0
torba	f	4	7	0
	%	36	64	0
djed	f	1	10	0
	%	9	91	0
džem	f	1	10	0
	%	9	91	0
ćirilica	f	0	10	1
	%	0	91	9
njiva	f	0	11	0
	%	0	100	0

In the process of encoding isolated words, the attention should be drawn to encoded variants of the word crkva, which were *serkova*, *carkva*, *cerkva*, and *serckva*. The absence of consonant clusters at the beginning of syllables in Farsi can explain such a result. Adding an extra vowel aims to break the consonant cluster in the foreign language, which was proved by earlier research (e.g.,Khorasgani et al., 2015; Alipour et al., 2019; Diklić, 2022). Similarly, the encoded entries for the word hrana were *herana, harne*, and *karana*, and the word *vlak* was encoded as *velak* and *vulak*. The word *djed* was often encoded with the initial grapheme *t* instead of *d*, which proves the substitution of voiced consonants by voiceless (Alipour et al., 2019), i.e., indicates the failure in voiced-voiceless distinction acquisition. It is consistent with previous research results, which indicate that the foreign language beginners first acquire voiceless and subsequently voiced consonants (Khorasgani et al., 2015). Even six respondents encoded the word *torba* as *turba*. The transfer between the native language, which encodes only long vowels, most probably caused such results (Alipour et al., 2019; Diklić, 2022). Respondents had heard the long vowel in dictated text

but did not encode it as the vowel *o*, which is not encoded in Farsi since it is short, but as a long and encodable vowel *u*.

6.4. *Accuracy in Encoding Graphemes in Words in a Sentence*

When encoding the graphemes in words in sentences, all respondents encoded as many as eleven graphemes accurately, which is a significantly higher accuracy percentage in relation to encoding the individual graphemes. That can mean the respondents rely more on lexical knowledge than on phonology, which was proved in Baluch's research (2005). Graphemes *dž* and *nj* are at the bottom of the table, with only two respondents who encoded them accurately.

Table 4 Accuracy in Encoding Graphemes in Words in a Sentence

		Accurately encoded grapheme	**Inaccurately encoded grapheme**	**Grapheme not coded**
B	f	11	0	0
	%	100	0	0
D	f	11	0	0
	%	100	0	0
F	f	11	0	0
	%	100	0	0
J	f	11	0	0
	%	100	0	0
K	f	11	0	0
	%	100	0	0
M	f	11	0	0
	%	100	0	0
N	f	11	0	0
	%	100	0	0
P	f	11	0	0
	%	100	0	0
R	f	11	0	0
	%	100	0	0
S	f	11	0	0
	%	100	0	0
T	f	11	0	0
	%	100	0	0
Č	f	10	1	0
	%	91	9	0
H	f	10	1	0
	%	91	9	0
L	f	10	1	0
	%	91	9	0
Z	f	10	1	0
	%	91	9	0
V	f	10	0	1
	%	91	0	9
Š	f	9	2	0
	%	82	18	0
Đ	f	8	3	0
	%	73	27	0
Ž	f	8	3	0
	%	73	27	0
G	f	7	4	0
	%	64	36	0
LJ	f	7	4	0
	%	64	36	0
C	f	6	5	0
	%	55	45	0
Ć	f	5	6	0
	%	45	55	0
DŽ	f	2	9	0
	%	18	82	0
NJ	f	2	9	0
	%	18	82	0

6.5. *Accuracy in Encoding Words in a Sentence*

Table 5 shows the accuracy in the encoding words in a sentence. All respondents (100%) encoded the words *baka* and *kava* accurately. Nine respondents (82%) accurately encoded the terms *čarapa*, *hrana*, *jakna*, *majka*, and *sestra*. *Čarapa* was encoded as *čerapa* and *đarapa*; *hrana* as *hrne* and *karana*; *jakna* as *jekna* and *jaka*; *majka* as *mjka* and *majaka*; *sestra* as *sistra* (N=2).

Eight respondents (73%) encoded the words *pismo* and *zastava* accurately. *Pismo* was encoded as *pesmo* (N=3) and *zastava* as *zastva*, *zastara*,and *zaztava*. Seven respondents (64%) accurately encoded the words *đak*, *šuma*,and *vlak*. The encoded entries for the word *đak* include the following: *đzek*, *đek*, *đeik*, and *džak*. *Šuma* was encoded as *šoma* (N=2) and *soma* (N=2), and *vlak* as *vlek*, *velak*,and *vilak*, while one respondent did not encode it at all.

The words *gitara*, *lav*, *ljubav*,and *robot* had five accurately encoded entries. *Gitara* was encoded as *gittara* (N=2), *kitera*, *qitara* (N=2), and *getara*; *lav* was encoded as *ljav*, *lov* (N=4), and*lave*; *ljubav* as *liubav* (N=3), *ljubev*, *ljobov*, and *lijobav*; *robot* as *rubot* (N=2), *robute*, *robat* (N=2), and *rodot*.

Four respondents (36%) encoded the words *crkva* and *flauta* accurately. *Crkva* was encoded as *sirkve*, *srikva*, *srkva*, *serkva* (N=2), *cerkva*, and *cerkova*; *flauta* as *flute*, *faluta* (N=2), *fluta* (N=2), *flavota* and *flouta*.

Two respondents (18%) accurately encoded the words *džem*, *novine*, *torba*, and *žaba*. *Džem* was encoded as *đzem* (N=2), *đem* (N=4), *geam* and *džem* (N=2). When encoding the graphemes *đ*and*dž*,it is evident that the respondents do not distinguish them completely, so they substitute them in some cases even in encoding. The word *novine* was encoded as *novina* (N=3), *navine*, *nobina*, *novene* (N=2), and *noveni* (N=2); the predominant encoded entry for the word *torba* was *turba* (N=8) again and *torboa*; *žaba* was mostly encoded as *žava* (N=5), *zaba* (N=2), *žave*,and *žara*.

One respondent (9%) encoded the word *ćirilica* and *djed* accurately. The encoded entries for the word *ćirilica* were *ćereica, čerlsa, ćjrier, čavrlica, čerlica, čerilica* (N=2), *ćereica*, *čiirilica*, *ćerileca*, and *djed* was encoded as *dijete* (N=2), *dijet* (N=3), *djet* (N=4), and *deat*.

The word *njiva* is at the bottom of the table. Not one respondent encoded it accurately, as was the case with the same word while encoding individual words. It was mainly encoded as

niva(N=5), followed by *neva* (N=2), *njva* (N=2), *nijva*, and *nieva*. These results also prove that the respondents did not acquire the written form of *ć* and *nj*,which do not exist in L1.

Table 5 Accuracy in Encoding Words in a Sentence

		Accurately encoded word	**Inaccurately encoded word**	**Word not encoded**
baka	f	11	0	0
	%	100	0	0
kava	f	11	0	0
	%	100	0	0
čarapa	f	9	2	0
	%	82	18	0
hrana	f	9	2	0
	%	82	18	0
jakna	f	9	2	0
	%	82	18	0
majka	f	9	2	0
	%	82	18	0
sestra	f	9	2	0
	%	82	18	0
pismo	f	8	3	0
	%	73	27	0
zastava	f	8	3	0
	%	73	27	0
đak	f	7	4	0
	%	64	36	0
šuma	f	7	4	0
	%	64	36	0
vlak	f	7	3	1
	%	64	27	9
gitara	f	5	6	0
	%	45	55	0
lav	f	5	6	0
	%	45	55	0
ljubav	f	5	6	0
	%	45	55	0
robot	f	5	6	0
	%	45	55	0
crkva	f	4	7	0
	%	36	64	0
flauta	f	4	7	0
	%	36	64	0
džem	f	2	9	0
	%	18	82	0
novine	f	2	9	0
	%	18	82	0
torba	f	2	9	0
	%	18	82	0
žaba	f	2	9	0
	%	18	82	0
ćirilica	f	1	10	0
	%	9	91	0
djed	f	1	10	0
	%	9	91	0
njiva	f	0	11	0
	%	0	100	0

When encoding individual words, the encoded entries for the words *crkva*, *hrana*, and *vlak* reveal that some respondents add an extra vowel to break the consonant cluster. When encoding the words in a sentence, the encoded entries for the word *flauta*, among which was *faluta*, reveal the same underlying principle. Eight respondents encoded the word *torba* as *turba*, which confirms the substitution of the short vowel o with the long u. There were difficulties in encoding the word *gitara*, when respondents often substituted *g* with *k* and *q*. The fact that Farsi has graphemes *g*, *k*, and *q*, which are similar according to the place of their articulation, could cause such encoding deviations.

7. Conclusion

Analysis of encoding by Farsi-speaking beginner learners of Croatian revealed generally successful encoding, particularly with sentences. However, they struggled most with individual graphemes, potentially indicating reliance on lexical rather than phonological knowledge. Difficulty was pronounced with graphemes absent in their native language, notably "nj," unlike earlier findings regarding "c." This discrepancy might be due to the infrequency of "nj" in Croatian, especially at early stages. However, learners did become aware of "c" during alphabet acquisition, absent in Farsi. Transfer errors from L1 were evident, such as omitting short vowels in Farsi. Focus in Farsi learner instruction should target specific grapheme errors, especially "nj," "ć," "dž," and "đ," and address consonant clusters and encoding of short vowels in Croatian, divergent from their L1. Future research should explore further phoneme-grapheme relationships beyond initial word phonemes.

References

Alipour, Maryam; Salehuddin, Khazriyati; Stapa, Siti Hamin (2019) „An Overview of the Persian EFL Learners' SpellingDifficulties.“*International Journal of Multicultural and Multireligious Understanding* 6/2012, 127–158.

Baluch, Bahman (2005) „Persian Orthography and Its Relation to Literacy.“ In *Handbook of Ortography and Literacy* 1st edition. ed. R. Malatesha Joshi, P. G. Aaron, 365–376.

Barešić, Jasna (2010) *Dobro došli 1 – udžbenik i rječnik za učenje hrvatskoga jezika za strance*.Školska knjiga, Zagreb.

Čilaš-Mikulić, Marica, GulešićMachata, Milvia, Pasini, Dinka, Udier, Sanda Lucija (2013), *Hrvatski za početnike 1 – udžbenik i rječnik*. Hrvatska sveučilišna naklada, Zagreb.

Dąbrowska, Ewa (2008) „The Effects of Frequency and Neighbourhood Density on Adult Speakers' Productivity with Polish Case Inflections: an empirical test of usage-based approaches to morphology."*Journal of Memory and Language*, 58 (4), 931–951.

Diklić, Tanja (2022) "Fonološka odstupanja govornika farsija u pisanome diskursu na početnoj razini učenja hrvatskoga kao inoga jezika". *Zeszyty Łużyckie*, 57, 203-217.

Đurđević,Antonela (2020) „Ovladanost pravopisnom normom glasova /č/ i /ć/."final paper, Odjel za kroatistiku, Sveučilište u Zadru, Zadar, 2020.

Ellis, Nick idr. (2004) „The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts." *Reading Research Quarterly* 39(4):438–468.

Gholamain, Mitra; Geva, Esther (1999) „Ortographic and cognitive factors in the concurrent development of basic reading skills in English and Persian."*Language Learning* 49,183–217.

Grgić, Ana; Udier, Sanda Lucija (2012) „Pravopisna kompetencija na razini B1 u hrvatskome kao inome jeziku."*Lahor*, *14/2012,* 204–220.

Jelaska, Zrinka; Barbaroša-Šikić, Mirela (2007) „Raščlamba diktata kao pokazatelj ovladanosti jezikom."In *2. Međunarodni skup Hrvatski kao drugi i strani jezik (HIDIS 2): Knjiga sažetaka.ed*.Novak-Milić. J et al. Zagreb.

Jelaska, Zrinka;Kekelj, Martina; Šafarić, Ines (2007) „Tumačenje izraza hrvatskih riječi u diktatima."In *VI. zn. skup Istraživanja govora: Knjiga sažetaka*, ed.Varošanec-Škarić et al. Zagreb.

Jelaska, Zrinka (2010) „Fonološka sastavnica u jezičnoj obradi."In *Proizvodnja i percepcija govora*. ed.Mildner, Vesna; Liker, Marko, 275–296. Zagreb: FF press.

Jelaska, Zrinka; Musulin, Maša (2011) „Slovo i slovopis španjolskoga i hrvatskoga jezika."*Lahor* 12/2011,211–239.

Jelaska, Zrinka (2012) „Ovladavanje materinskim i inim jezikom."In *Inojezični učenik u okviru hrvatskoga jezika: okviri za uključivanje inojezičnih učenika u odgoj i obrazovanje na hrvatskome jeziku*, ed.Češi, M., Cvikić. L., Milović, S., 19–35. Zagreb: Agencija za odgoj i obrazovanje.

Jelaska, Zrinka; Lalli Paćelat (2014) “Zatvornički sljedovi u sredini hrvatskih i talijanskih riječi - mogući izvor poteškoća govornika talijanskoga jezika u ovladavanju inim jezikom”. Jezikoslovlje 15.1, 67-89

Jukić, S. (2018) Frazem kao sastavnica diktata u nastavi hrvatskoga kao inoga jezika. Strani jezici, 47 (1-2), str. 79–94.

Jukić, S. (2022). Ovladanost pisanjem fonema /c/, /z/ i /s/ na temelju diktiranoga teksta u hrvatskome kao inome jeziku. U: Marković, I.; Nazalević Čučević, I.; Gligorić, I. M. Riječi o riječi i Riječi: Zbornik u čast Zrinki Jelaska. 269–284. Zagreb: Disput.

Jukić, S.; Diklić, T.; Prosenjak, B. (2022). Pisanje u hrvatskome kao inome jeziku – prikaz slučaja francuskih početnika. Izlaganje na 14. Stručnom i znanstvenom savjetovanju za lektore hrvatskog kao drugog i stranog jezika (SIH): Opis kao temelj poučavanja. Zagreb: Filozofski fakultet.

Khorasgani, Amir Toghyani; Khorasgani, Ali Toghyani; Aray, Narges Keshti (2015) „A survey on several potentially problematic areas of pronunciation for Iranian EFL learners.“*Indonesian EFL Journal* 1(2),189–198.

Landau, E. Lončarić, M., Horga, D., Škarić, I. (1999). Croatian. U: Esling, J. (ur.), *The Handbook of the International phonetics association: A guide to the use of the international phonetic alphabet*, 66-69. New York: Cambridge University Press.

Musulin, Maša (2011) „Grafemkaovišeznačnica.“ *Lahor* 12/2011,240–247.

Musulin, Maša; Jelaska, Zrinka (2014) „El sistema fonológico en el habla y en la lectura: español y croata.“*SRAZ LIX*, 77–94.

Omidipour, Mahmoud (2014) „An Analysis of Errors in Writing Among Adult Persian Learners of English.“*International Journal of Language Learning and Applied LinguisticsWorld*vol.5 (3)/2014, 176–187.

Windfuhr, Gernot Ludwig (1979) Persian. In*The World’s Major Languages* ed. B. Comrie, 523–547. Oxford UP.

Yahya, Azizi; Ishak, Harrison BT.; Zainal, Zaidah; Faghat, LadanJavdan; Yahaya, Noordin (2012) „Error Analysis of L2 Learners' Writings, a Case Study.“*International Conference on Language, Medias and Culture* 33/2012: 114–118.