Computer-assisted Learning of Croatian Orthography concerning the *yat* reflex (CAL-COR)

Damir Boras, Dolores Grmača, Nikola Ljubešić
Department of Information Science, Faculty of Philosophy
University of Zagreb
I. Lučića 3, 10 000 Zagreb, Croatia
Phone: +38516002302 Fax: +38516156879
E-mail: dboras@ffzg.hr, dgrmaca@ffzg.hr, nljubesi@ffzg.hr

Abstract - The aim of this paper is to present two phases in constructing the Croatian orthography learning system concerning the *yat* reflex as well as its final form. During these two phases, we compared three different models of rules concerning the *yat* reflex - a basic, an intermediate and an advanced one. We also compared two different approaches for learning the *yat* reflex - an induction-deductive and an inductive one. All the data gathered during our research helped us in constructing a complete, powerful learning tool.

I. INTRODUCTION

The aim of this work is to develop a system for learning the *yat* reflex. Such a system belongs to the area of computer-assisted language learning (CALL), i.e. in case of high degree of interactivity and adjustment to the needs and capabilities of the learner, in the area of intelligent computer-assisted language learning (ICALL) [1]. This work is one of the first attempts to develop a system for learning a segment of the Croatian language as the mother tongue.

One of the biggest problems in the Croatian orthography is the *yat* reflex. Otherwise, the Croatian orthography is simple due to the high degree of the phonetization of the script. In the past, the Croatian language had the vowel *yat* (ě) that could be both long and short (like today's vowels). In contemporary Croatian language there are four possibilities for its rendering (*yat* reflexes) – *ije, je, e, i*. This can be best observed when words are compared from the ikavian, ekavian and jekavian dialects:

Ikavian: did, svit, uvīk
Ekavian: ded, svet, uvēk
Jekavian: djed, svijet, uvijēk

The reflex *ije* is found in long syllables, reflexes *e, i* in short syllables, and reflex *je* both in short and in long syllables. They are replaced due to the change in the quantity of the syllable which includes them.

The complexity of the reflexes of the phoneme *yat* has been indicated by our research where the students of the Faculty of Philosophy were tested. In the initial testing of their knowledge of the *yat* reflex, the examinees made mistakes in 16.66% of cases. The extent of the problem was proven by the fact that some of the tested students were those from the senior years of the Croatian language and literature study as well as other philological studies, who also had high percentage of mistakes, amounting to 12.30% and 14.36% respectively.

The research consisted of two phases. In the first phase, three learning models in the form of three different levels of rules complexity were tested. In the second phase, two learning models were tested, the one indicating the respective rule in case when the examinee makes a mistake, and the one indicating only the correct language form. Both phases contained the initial and the final knowledge test and one of the learning models in between. The initial knowledge tests in both research phases were identical, as well as the final ones. The obtained numerical results may seem negligible, but one has to take into consideration the complexity of the problem for the solving of which a single testing is certainly insufficient. There should rather be successive learning over a longer period of time. Learning according to certain models has also given negative results, the possible cause being the overburden to which the examinees were subjected by being exposed to an excessive quantity of specific data.

II. TECHNICAL PERFORMANCE

The model of the future system was developed in the form of dynamic web pages using HTML, ASP server-side scripting and JavaScript client-side scripting. The measured values were saved in a relational database. The examinees logged into the system model by means of the given usernames and respective passwords. After having logged in, the examinees were given 62 questions concerning the *yat* reflex. The examinees answered the questions by filling in the gaps selecting the correct reflex from the combo box. After having completed the initial knowledge test, the examinees were informed about their score in solving the problems and were given the correct answers to those questions that they had answered incorrectly. During the initial knowledge test, as well as during the final knowledge test, the time of test solving, the responses of the examinees and the correctness of the given answers were measured. These data are considered to be helpful in selecting the questions and their testing dynamics in the final form of the system. During the learning process, the time that an examinee spent on each slide was measured. The form of the final knowledge test is identical to the form of the initial knowledge test. Also, the same values were measured. At the end the examinee was informed again, for purely altruistic reasons, about the score achieved, as well as the correct forms of the incorrectly solved problems.

III. FIRST PHASE OF THE RESEARCH

In the first phase of the research three different models of the rules concerning the *yat* reflex were compared. The models differed according to the difficulty and volume. The number of examinees was 80. In both research phases the examinees were the students of the Faculty of Philosophy.

In the elementary model only the basic rules are described in simple words. These rules have been designed according to the course book for the eighth grade of the primary school [2]. The rules are organized in 9 short slides. Each slide ends, as well as in the other two models, by a task which checks whether the examinee has understood and mastered the described rule. The examinees do not go on to the next rule until they have answered correctly the question. As we had assumed, this model gave the poorest results precisely because of the insufficient explanation given on this problem which obviously requires a more detailed analysis of the rules and a more complex approach. The examinees, namely, made on the average in the final knowledge test 1.44% mistakes more than in the initial knowledge test.

The comprehensive model provides more extended rules, explained in great detail. The rules are designed according to the *Grammar of the Croatian Language, A Manual for the Basic Language Education* [3] and *Croatian Orthography Manual* [4]. They are presented on 14 slides. In spite of our expectations, this model did not yield the best results. The examinees made, namely, on the average 0.81% more mistakes in the final knowledge test than in the initial one.

The advanced model explains the rules by means of the most professional and most detailed level, and the rules are divided into concrete grammatical categories according to the replacement of the *yat* reflex. The rules have been developed according to *Croatian Grammar* by the Institute for Croatian language and linguistics [5] and the *Orthography Manual of Croatian Language* [6]. The rules are presented on 18 slides. This model has shown best results with an average increase in the accuracy of solutions in the final knowledge test of 1.68%.

The detailed data on the results of the first phase of research are presented in Table I and Table II.

TABLE I

DATA GATHERED DURING THE FIRST PHASE CONCERNING THE THREE MODELS

	elementary model	comprehensive model	advanced model
number of examinees	28	26	26
average improvement	-1.44%	-0.81%	1.68%
average improvement IKT* < 80%	-3.23%	1.61%	7.46%
average improvement IKT ≥ 80%	-0.60%	-2.09%	-0.72%

*IKT - initial knowledge test

The presented numerical results may seem negligible, but one needs to take into consideration the complexity of the problem that certainly cannot be solved over a short period of time such as this testing. The average duration of the testing in the first phase, for instance, took 16 minutes and 27 seconds. The reason why certain learning models gave negative results is, in our opinion, the excessive amount of rules over a short period of time that confused the examinees, as well as the overall burden to which the examinees were subjected by the mere testing process. This is also confirmed by the fact that for the solving of the final knowledge test, which is in its length same as the initial knowledge test, the students required on the average 21.529% less time than for the initial one. We are certain that the accomplished improvement would have been substantially greater had there been a break between learning and the final knowledge test. However, this was impossible for organizational reasons.

TABLE II

DATA GATHERED DURING THE FIRST PHASE IGNORING THE THREE MODELS

average improvement	-0.22%
average improvement IKT < 80%	1.61%
average improvement IKT ≥ 80%	-1.11%

The data also show that, except in the elementary model, the examinees with less foreknowledge achieved usually better results, as expected. On the contrary, the examinees with more foreknowledge failed to achieve any major improvement or even showed regression. We believe that the examinees with poorer foreknowledge in the elementary model showed regression because of the overgeneralization of the rules regarding the amount of foreknowledge.

The advanced model for this profile of the examinees proved to be the most successful with a great increase in case of examinees with poor foreknowledge and minimum decline of examinees with solid foreknowledge. The elementary and the comprehensive models obviously did not suit our examinees. It may be assumed that the elementary model would be the most suitable for the primary school age, and the comprehensive for the secondary school age.

IV. SECOND PHASE OF THE RESEARCH

The first phase, which showed that the advanced model yielded the best results, resulted in the second phase. The advanced model of the first phase was designed in two approaches: the induction-deductive and the inductive approach. The induction-deductive approach offers nine interactive exercises that consist of five questions each with examples of all the rules. In case of a wrong answer, the correct one is given as a part of the respective rule. In the inductive approach, the same nine interactive exercises are given, but here in case of a wrong answer only the

correct answer is supplied. In this research phase 91 students were tested. The results of this research phase are presented in Table III and Table IV.

TABLE III

DATA GATHERED DURING THE SECOND PHASE CONCERNING THE TWO APPROACHES

	Induction-deductive approach	Inductive approach
number of examinees	46	45
average improvement	0.04%	2.15%
average improvement IKT < 80%	1.27%	4.84%
average improvement IKT ≥ 80%	-0.50%	0.52%

This research phase continued the trend of slight improvement which once again indicates the complexity of this problem and the impossibility of solving it over a short period of time. The inductive approach yielded better results, which in turn indicates the burden that a rule imposes on the examinee. Regarding the distribution of improvement, the examinees with poorer foreknowledge achieved much greater improvement than those with greater foreknowledge.

TABLE IV

DATA GATHERED DURING THE SECOND PHASE IGNORING THE TWO APPROACHES

average improvement	1.08%
average improvement IKT < 80%	3.23%
average improvement IKT ≥ 80%	-0.03%

V. FINAL FORM OF THE SYSTEM

The database of the final system for learning the *yat* reflex will consist of exercises and corresponding rules that will include about 1000 most frequent words with the *yat* reflex. Furthermore, the data about the user's answers will be written in the same database, while the new exercises will be generated upon user's previous answers.

Users will solve the exercises by filling in the gaps selecting the correct *yat* reflex from the combo box. It was decided to use primarily the inductive approach because it yielded better results than the induction-deductive one and because the average number of mistakes made by students on the basic word forms was 15,97%, which is almost equal to the overall number of mistakes (16,66% in total).

The respective rule will be shown optionally to the user because our results showed that the rules did often confuse and overburden our examinees during the research. Furthermore, we should take into consideration the fact that this system will be used by individuals with average or even poor linguistic knowledge. These rules will be of great assistance in the database structure since it will be possible to find similar language forms concerning the *yat* reflex over some specific rule. We will be using the advanced model of rules from the first phase of the research because the system developed up to this point is intended to be used primarily by adults.

The final system will start the communication with the user asking him to answer the questions from all the areas concerning the *yat* reflex in order to locate the areas in which the user has the biggest problems. When designing this part of the system, we considered the results of our two knowledge tests that have shown some typical problems for most users in the area of the *yat* reflex (the list of the language forms and the percentage of correct answers is given in the annex). The system will always suggest the length of one learning period and the point in which the further learning should start.

Exercises given by the system will be primarily taken from the areas where the user makes most mistakes. The user will always be given more similar questions successively. Also, the similar exercises in succession will always be swapped by random exercises to assure that all the grammatical categories which include the *yat* reflex are covered.

The real proof of the system's quality will be the progress of the users using this system who have greater foreknowledge. These users, as shown in the research, are those who make minimal progress. The final form of this system will be constructed as a web application. Therefore, all the data gathered about users' habits and their success will be used to further optimize the system.

VI. CONCLUSION AND FURTHER DEVELOPMENT

Language learning is a complex process, which requires a specific amount of time. That is why we believe the learning progress shown by our examinees was so small. Also, we assume that one of the possible reasons (except the lack of time) would be the overburden caused by exposure of the examinees to an excessive quantity of specific data. Taking that into consideration, it is possible to conclude that the knowledge progress could be greater than the one shown by the given results, i.e. that this progress would be greater if the learning period was longer. That is why the user of this system should spend much more time learning iteratively. We believe that the final form of the system described in the previous chapter will eradicate all these problems.

Finally, the future of this system lies in its optimization that can be obtained analyzing the data gathered by monitoring the users' behaviour while using the system. Also, the further study should be done on the elementary and the comprehensive model from the first phase of the research since these two models are believed to be most suitable for the primary and secondary school age.

VII. LITERATURE

[1] R. Marinković, *Intelligent Systems for Education*, Hrvatska zajednica tehničke kulture, Zagreb, 2004. (in Croatian)

[2] Z. Lugarić, and N. Koharović, *Croatian Language 8: Course Book for the Eight Grade of Primary School*, p. 40-42, 108, Profil International, Zagreb, 2001. (in Croatian)

[3] S. Težak and S. Babić, *Grammar of Croatian Language: A Manual for the Basic Language Education*, p. 61-69, Školska knjiga, Zagreb 1992. (in Croatian)

[4] S. Babić, B. Finka and M. Moguš, *Croatian Ortography Manual*, p. 36-48, 6. ed., Školska knjiga, Zagreb, 2002. (in Croatian)

[5] E. Barić, M. Lončarić, D. Malić, S. Pavešić, M. Peti, V. Zečević and M. Znika, *Croatian Grammar*, p. 88-91, Školska knjiga, Zagreb, 1995. (in Croatian)

[6] V. Anić and J. Silić. *Orthography Manual of Croatian Language,* p. 124-131, Novi Liber and Školska knjiga, Zagreb, 2001.

VIII. ANNEX

INITIAL KNOWLEDGE TEST RESULTS

	Percentage of correct answers	Form
1.	7,018%	sazreti
2.	48,538%	rasiječem
3.	49,708%	pogriješiti
4.	53,801%	dijela
5.	59,064%	osvijetlivši
6.	59,064%	zastarjeti
7.	60,234%	presjedati
8.	61,988%	premještati
9.	67,251%	zalijepljena
10.	67,836%	sasjeći
11.	70,760%	sazrijevanje
12.	71,345%	rješenje
13.	73,099%	presjedanje
14.	73,684%	svjetlom
15.	74,854%	razumjeti
16.	76,023%	razumijevanje
17.	76,608%	pripovijeda
18.	77,193%	zastarijevati
19.	78,947%	djelić
20.	78,947%	presječen
21.	80,117%	čovjek
22.	80,702%	zvjezdanom
23.	84,211%	korijen
24.	84,795%	iskorjenjiv
25.	84,795%	riješiti
26.	84,795%	rješavanjem
27.	86,550%	dio
28.	86,550%	ljepilom
29.	87,135%	premjestiti
30.	87,135%	razumijevati
31.	88,889%	iskorijeniti
32.	88,889%	polijevati
33.	89,474%	naljepnica
34.	89,474%	rječica
35.	89,474%	sazrijevati
36.	89,474%	zvijer
37.	90,058%	umjetnik
38.	90,643%	bjedniji
39.	90,643%	najbjedniji
40.	91,228%	prijenos
41.	91,228%	svjetlucaju
42.	92,982%	pobjedu
43.	92,982%	politi
44.	92,982%	prilaz/prijelaz
45.	92,982%	pripovjedačica
46.	93,567%	djed
47.	93,567%	živjeli
48.	94,152%	korjenčić
49.	94,737%	bijedan
50.	94,737%	najljepše
51.	94,737%	zvjerad
52.	94,737%	živjela
53.	95,322%	prenositi
54.	95,906%	djedov
55.	95,906%	zvjezdurine
56.	96,491%	ljepše
57.	96,491%	vidjeli
58.	97,076%	rijeka
59.	97,076%	živio
60.	97,661%	pogr(j)eška
61.	98,246%	lijepo
62.	98,246%	zvijezde

FINAL KNOWLEDGE TEST RESULTS

	Correct anwser	Form
1.	43,275%	brježić
2.	46,199%	obavijesna

	Correct anwser	Form
3.	59,649%	zasijeci
4.	60,234%	sagorjeti
5.	60,819%	namještati
6.	66,667%	smjer
7.	67,251%	bdjeti
8.	67,836%	donijela
9.	68,421%	ispovjedim
10.	70,760%	donijeli
11.	71,345%	umjeti
12.	71,930%	zapovijeda
13.	74,269%	presjeći
14.	76,608%	rječnički
15.	76,608%	sagorjevši
16.	77,193%	odsjekavši
17.	77,778%	namještanje
18.	77,778%	povijesno
19.	77,778%	umijeće
20.	79,532%	polijetati
21.	79,532%	umijeće
22.	80,117%	namjestiti
23.	80,117%	uspjeti
24.	81,871%	br(j)egovi
25.	82,456%	ispovijed
26.	82,456%	mijenjaš
27.	83,041%	polijetanje
28.	85,380%	nalijevati
29.	85,965%	ocjene
30.	86,550%	dragocjene
31.	86,550%	mjenjačnici
32.	86,550%	smjernica
33.	86,550%	zapovjednik
34.	87,719%	naliti

	Correct anwser	Form
35.	87,719%	promjenjive
36.	89,474%	rječit
37.	90,643%	cijene
38.	90,643%	ispovjednik
39.	90,643%	rječnik
40.	90,643%	zamijeniti
41.	91,813%	probdi(je)m
42.	92,398%	htjeli
43.	92,398%	zamjenjivati
44.	93,567%	htio
45.	93,567%	primjedba
46.	94,152%	donio
47.	94,737%	prijepis
48.	95,322%	htjela
49.	95,322%	poletjeti
50.	95,906%	bjelje
51.	96,491%	predložiti
52.	96,491%	prijedlog
53.	96,491%	prijevod
54.	96,491%	svjetovi
55.	96,491%	uspijem
56.	97,076%	najbjelje
57.	97,076%	prepisati
58.	98,246%	bijelo
59.	98,830%	prevoditi
60.	98,830%	svijet
61.	98,830%	vri(je)
62.	99,415%	riječ