

Journal of Universal Language 4
March 2003, 91-112

The Writing System of an Artificial Language: For Efficient Orthographic Processing

Miho Sasaki

University of Essex

Abstract

The major purpose of this paper is to explore the role of orthography in language and to propose new approaches to the development of an artificial language in relation to reading and spelling. Recent second language acquisition research has shown evidence that orthography is one of the most important features for acquisition of reading skills. For many years, effects of orthographic regularity on efficiency of language processing have been underestimated. This paper looks at the relationship between orthography and phonology across languages and examines what is the optimal writing system. The writing system of Esperanto is also investigated in terms of orthographic regularity to uncover possibility of a simpler writing system.

Keywords: orthography, writing systems, language acquisition and processing, reading

1. Introduction

Most languages in the world have the written form as well as the spoken form. A written language is essential in order to record or organise information; and because of this function, it seems to have some sort of priority or power to influence the reality. For example, it is often required to ‘read’ and understand terms and conditions, ‘sign’ a contract, or ‘fill in’ an application form. Knowledge of how to read and write is requisite for modern living and also writing itself is an important function of language.

When you write and read a language, you are not only using the written form of the language which is provided as a system but rather you are encoding speech to writing or writing to speech. In order to do so, the relationship between speech sounds and written symbols has to be recognised and processed simultaneously along with their meanings. This activity is understood as orthographic and phonological processing, and the skills to operate these processes, i.e., *processing skills*, must be acquired by a reader’s reading experience. These skills are necessary to use written language as efficiently as possible. Normal adult readers and spellers can read and spell relatively fast and correctly because they have acquired the processing skills. From a crosslinguistic point of view, recent reading research revealed that these processing skills might vary across languages since the conditions and rules of written forms, i.e., writing systems, vary (Wydell & Butterworth 1999, Perfetti & Tan 1998, Katz & Frost 1992). If the processing skills develop according to writing systems, there must be variety in the ease of acquisition and development of the skills across languages. This paper will discuss how people acquire and process a written language and what a simple writing system might be in respect to the acquisition and processing.

2. Language Evolution and Writing Systems

2.1. Primacy of Speech?

It is a fact that speech is historically prior to writing (Halliday 1987). According to Halliday, ‘primacy of speech’ indicates that speech is the primary medium of linguistic expression with regard to both the way language evolved and how children acquire a language. Speech is the most natural state of language that can be employed when people are within the hearing distance of each other’s voice. When humans first learned to communicate each other, they used sounds and gestures and such communications were limited in terms of space and time. There was no means for transmitting a message to another place or time, except for memorising and repeating it in person. Thus, it was natural that ancient people started to require a means of carrying and recording messages in a more stable form than speech. Ellis (1848) states “... memory of a go-between is too treacherous to be trusted, and we know how uncertain tradition is when tales are handed down from father to son for a few generations.” Orthography evolved in such a background, and until the invention of recording facilities of spoken language (e.g., recording machines in the nineteenth century) writing was the only means to record language.

Some linguists still strongly believe that speech is primary while writing is secondary for the reasons that children acquire speech first and illiterate people use language without writing and reading. However, at present, it is emphasized that writing and speech are different and equal manifestations of language (Crystal 1997). Writing is not the other state of speech: its strategies and processes of production and comprehension are rather autonomous from those in speech. In other words, writing has an independent role in language use.

2.2. A Common Script

The way that written language visually represents language is broad and diverse; there are various types of scripts and orthography. However, some languages often have a common script, such as Italian and Spanish, because of a same origin (i.e., Latin), or a strong influence they gave to or received from each other for geographical and historical reasons. Thus, in those languages with a common script, visual appearance can be similar, but the relationships between sounds and spellings vary from a language to another: Each language has a particular writing system. For example, English and French have words with the same spellings but different pronunciations, as shown in (1a-c).

(1)	<u>English</u>	<u>French</u>
a.	station /stéɪʃən/	station /stasjɔ̃/
	face /feɪs/	face /fas/
	nature /neɪʃər/	nature /natyʁ/
b.	steak /steɪk/	steak /stɛk/
c.	point /pɔɪnt/	point /pwɛ̃/
	instant /ɪnstənt/	instant /ɛ̃stɑ̃/

These words might have the same origins, but the pronunciations and meanings have been transformed to fit to each language system, without changing the visual appearance. Other languages like Chinese and Japanese share a script in the result of Japan's adoption of the Chinese orthography. The huge influence of the country led to the adoption of the script but there were extreme difficulties to adapt the Chinese written form to the Japanese language because Japanese and Chinese share neither the same linguistic origin nor the structure. Therefore, the syllabic Kana script was derived from the adopted Chinese characters, i.e., Kanji, to supplement the incongruity be-

tween the spoken and written languages. The current Japanese writing system consists of Kanji and Kana, and still the written languages of Chinese and Japanese are visually quite similar except for Kana, but the phonology is different.

(2) <u>Chinese</u> (traditional)		<u>Japanese</u>		
		<i>On-reading</i>	<i>Kun-reading</i>	
a. Character				
‘east’	東 /dong/	東	/too/ /ton/	/higashi/
‘spring’	春 /chun/	春	/shun/	/haru/
b. Word				
‘Tokyo’	東京 /dongjing/	東京	/tookyoo/	-----
‘England’	英國 /yingguo/	英國	/eikoku/	-----
b’. Word (inflectional)				
‘beautiful’	美麗 /meili/	美しい	(/bi/)	/utsuku-shiii/
‘to hit’	叩 /kou/	叩く	-----	/tata-ku/
c. Phrase				
‘Good morning’				
	早上好/zaoshanghao/	お早う	-----	/ohayoo/
‘Thank you’				
	謝謝 /xiexie/	ありがとう	-----	/arigatoo/
*The tone in Chinese is omitted.				

As you can see in (2a-c), a Japanese Kanji character usually has a multi-pronunciation: *On-reading* (Chinese pronunciation) and *Kun-reading* (Japanese pronunciation). However, *On-reading* is the pronunciation when the character was adapted to Japanese so it is rarely consistent to the current Chinese pronunciation. In addition, the representation of morphemes or higher level (i.e., phrases or sentences) generally requires Kana supplement to describe inflections and grammatical features (2b’, c). The Japanese writing system demonstrates that a borrowed script can be adopted but it may need a considerable alternation or support in order to accord with the lan-

guage structure.

2.3. Orthography and Phonology

A written language is governed by the writing system so that readers will know a sound from a spelling or similarly spellers will know a spelling from a sound. The writing system consists of the rules that represent relationships between spelling and sound. The data in (3a-c) and (4) show examples of the rules of the English writing system (Venezky 1970, Carney 1994).

(3) The letter string <gh> at the end of a word is pronounced /f/ or silence.

a. enough /ɪnʌf/

 cough /kɒf/

b. though /ðəʊ/

 through /θruː/

 to compare with

c. ghost /ɡɒst/

(4) A word does not end with the letter <v>.

a. have /hæv/

b. attractive /ətɹæktɪv/

Letters are presented in <>.

(3a-c) describes that the letter string <gh> has a one-to-many correspondence between the spelling and the sound. The data in (4) reveals that the pronunciation can end in /v/ but not the spelling. These rules can be taught at school but usually learnt from the experiences with the written language. There are further more hidden rules in English which do not represent actual sounds but provide visual cues for sounds, such as (5) and (6).

(5) A final <-e> marks the previous vowel as a long vowel.

- | | |
|------------------------|-----------|
| a. can <u>e</u> /keɪn/ | can /kæn/ |
| b. mat <u>e</u> /meɪt/ | mat /mæt/ |
| c. pin <u>e</u> /paɪn/ | pin /pɪn/ |

(6) A doubled consonant letter marks the previous vowel as a short vowel.

- | | |
|-----------------------|------------|
| a. latt <u>er</u> /æ/ | later /eɪ/ |
| b. rudd <u>er</u> /ʌ/ | ruder /u:/ |
| c. ripp <u>er</u> /ɪ/ | riper /aɪ/ |

As defined in (5), a final <-e> influences the quality of a preceding vowel in a word although it is not sounded itself. This type of spellings, i.e., a visual cue, is in fact commonly used in English as in (6). However, these context-sensitive correspondence rules for English spelling are extremely complex and may not guarantee a definite pronunciation. They are hidden cues which readers are not consciously aware of but visually and phonologically practiced. The English writing system includes several rules like these, which only apply to the English language. Although the English orthography is categorised as ‘alphabetic’ where letters basically describe how they sound, the English system is not always transparent when expressing phonology. This is because the correspondences between spelling and sound are rather vague and often inconsistent.

3. The Lexicon

The relations between linguistic symbols (i.e., sounds or spellings) and meanings are arbitrary, except for pictograms or onomatopoeias. Even so, a word is always passed in a form such as spoken, written or sign language. It is very difficult to ‘guess’ the meaning from a form which you do not know because there is no systematic

rule to connect form and meaning. Indeed, learning the relation between a form and its meaning is an important part of language acquisition. One can say that what children or second language learners need to learn about a new language is the relationships between forms and meanings which are currently agreed in the language's society.

3.1. The Structure of the Lexicon

The lexicon stores the information about words as well as elements of words (Harley 2001). It is organized reflecting orthographic, phonological, semantic and syntactic features. When people encounter a word in a spoken or written form, the phonological or orthographic information will access the lexicon and only when the semantics are successfully retrieved, they understand the meaning. At the same time the lexicon is involved with production of language. A semantic representation in the mind will search for a word by its phonological or orthographic information then it is transformed into an appropriate physical form such as sound or spelling to submit to the other. The lexicon is the store of common knowledge that allows people communicate.

What structure does the mental lexicon have? Psycholinguistic studies have submitted several models of the lexicon regarding its processes and access. In terms of how many lexicons there are, the core question is if there is only one lexicon for speaking, writing, listening and reading, or there are four lexicons for each activity. It is also argued that there are two lexicons for input and output or alternatively for written and spoken language each. Most of the experimental data has shown evidence against a single lexicon governing both input and output; therefore, there may be two separate lexicons used for recognition and production (Harris & Coltheart 1986, Harley 2001). For example, Shallice, McLeod, & Lewis (1985) asked their subjects to detect a name in the auditory stimuli while

they were reading visually presented words aloud. They found no significant difference in the performances of the combined task compared to those of each single task only. The results indicated that the speech perception and production processes involved in the two tasks were functionally separate. According to previous empirical studies, it seems that the modality of language involves the structure of lexicon.

In fact, however, it is difficult to distinguish 'storage' from access. Some researchers believe that there are multiple semantic memory systems as discussed above; on the other hand, Allport & Funnell (1981) suggested that there are just different access pathways to one lexicon instead of separate lexicons. Their argument is that it is not economical to have four lexicons that contain the same concepts. Instead, language processes could split early in processing, in other words, in different access pathways. Thus, the neuropsychological data that support the existence of four different lexicons can be explained by four different access mechanisms for each modality; spoken and visual word recognition may use different mechanisms and also there may be distinct input and output with some sub-lexical mechanisms (Harley 2001). The structure of the lexicon is a complex issue because it is hard to test experimentally. Nonetheless, the possibility of different stores is less likely compared with the hypothesis of processing models by different access mechanisms.

3.2. Orthographic Information in the Lexicon

Because of the arbitrary relationship between an orthographic (or phonological) form and a meaning, the lexicon must involve some stored entry for the orthographic information. The types of orthographic information for English, for example, are considered to be those features of *grapheme-phoneme regularity* and *orthographic structure* (Venezky 1995). A phoneme is a smallest sound unit dis-

tinguished in a language such as /θ/ and /æ/ in English; and a grapheme is an equivalent visual unit to a phoneme like <th> and <a>. The grapheme-phoneme correspondence patterns can be categorised by three groups in terms of the regularity: *consistent* (e.g., f, j, ph, q, sh, v, wh, z), *inconsistent but predictable* for most of the other consonants and a few vowels (e.g., <gh> for /f/, silent, or /g/, as described in (3a-c)), and *inconsistent and unpredictable* (e.g., *yacht*). The spelling group with 'inconsistent but predictable' pronunciations has features which determine the pronunciation, such as position within a word, adjacent letters and sounds, and position relative to the main word stress. These patterns in all the three groups should be stored as orthographic information.

The second type of orthographic information, orthographic structure, is represented by the regularity of letters or letter sequences, i.e., frequency. For example, a simple statistic analysis using a corpus reveals that the letter <z> does not occur as often as <t> in English. The positional frequency of letters can be also analysed in this statistical regularity. Further analysis is demonstrated by the rule-governed regularity that is established on whether a sequence violates English phonological and orthographic conventions. For example, /sk/ is an allowable initial consonant cluster with possible spellings such as <sk> <sc> or <sch>; on the other hand, /tl/ or /dl/ do not occur at the initial position so there is no corresponding spellings <tl> or <dl> at the word initial position. The rule-governed regularity allows producing pronounceable and structurally regular English nonsense words such as *voam*, *woach*, *trenq*, *derl*, etc., where the words, however, contain positional letter sequences which do not occur in English words (i.e., *voa-*, *woa-*, *-eng*, *-erl*). Similarly, *clav* and *ckib* are both phonologically acceptable but orthographically illegal because of final <v> and initial <ck> (Venezky 1995: 119). Thus, letter or letter sequence regularity (i.e., frequency) is important information stored in the lexicon and used to recognise and to produce written form.

It is assumed that orthographic information is available in sub-systems along with phonological or semantic information in order to access lexical items. However, the types of orthographic information and the importance may vary across languages. For English, the grapheme-phoneme correspondences and the orthographic structure are the important information. One of the many reasons could be because English orthographic structure possesses more than one potential spelling for almost every sound. Thus, the information is crucial knowledge to read and write English efficiently. It is important to note that the structure of language and its processing are not separate issues and from the perspective of language acquisition, learning a language is equivalent to acquiring relevant information and moreover developing its processing skills. Therefore, the simpler and more consistent a language structure is, the easier to collect information and develop the specific processing skills.

4. Diversity of Writing Systems and Processes

Early word recognition research suggested that phonological information is always necessary to access the meaning of a word (Rubenstein, Lewis, & Rubenstein 1971). However, subsequent research has proposed that both visual and phonological methods could exist and operate in parallel. This *parallel processing* model explains how logographic systems such as Chinese Hanzi and Japanese Kanji access meaning directly from a visually presented word. It also supports the notion that logographic systems have phonological elements in their processing. In fact, Kanji processing is not independent from phonology; Wydell, Patterson, & Humphreys (1993) provided the evidence that Kanji readers showed hesitation and more errors on the same sounded Kanji words (i.e., *homophones*) in a semantic judgment task even though the semantic differences were obvious from the character's visual representation

(the *homophone effects* on Kanji).

(7) <u>Kanji homophone pairs</u>	<u>English homophone pairs</u>
/genshi/ 原始 原子	/sʌn/ son - sun
/kisha/ 記者 汽車	/teɪl/ tail - tale

The homophone effect is a typical phenomenon in phonologically based writing systems such as English (Van Orden 1987).

In terms of the orthographic representation, Katz & Frost (1992) claimed that the degree of dependence of phonology on alphabetic writing systems is diverse across languages; therefore they proposed the *Orthographic Depth Hypothesis*, which posits effects of the consistency between graphemes and phonemes on processing written words. According to the Orthographic Depth Hypothesis, Spanish and Italian have regular or shallow orthography, on the other hand, English and Dutch are categorised as languages with irregular or deep orthography, as compared in (8a, b). Logographic scripts can be also categorised here as an extreme example.

(8) <u>Spanish</u>	<u>English</u>
a. 'i' /i/ mi, siete	'i' /ɪ/ mint, sin
	'i' /aɪ/ idea, pint
b. 's' /s/ causa, mismo	's' /s/ case, summer
's' /z/ raise, misery	
'z' /z/ azul, vez	'z' /z/ zoom, seize

Katz & Frost (1992) proposed that a shallow orthography is more easily able to support a word recognition process whereas a deep orthography may require reference to morphology via a word's visual orthographic structure. Thus, word recognition processes can vary depending on the degree of orthographic regularity across languages.

There is argument as to whether there is a variety of favoured strategies for reading in different types of orthography. For example,

Japanese and Korean employ logographic and phonological combined writing systems. Do they use different processing for those two different orthographies? It is likely that they mix orthography-independent access and orthography-dependent access. According to the result from Wydell et al. (1993) discussed above, Kanji processing can use either of the two processing routes in accessing the properties of a lexical item. Moreover, although Japanese syllabic Kana script has a simple one-to-one connection between letters and phonemes, familiar Kana words are often accessed via orthography. Therefore, it is presumed that different processing strategies are used within the same language, although there might be a favoured strategy for a script such as a phonological access for Kana script. In summary, it is likely that they have alternative strategies depending on the familiarity and regularity of the words.

5. Second Language Acquisition with Different Writing Systems

Recent second language (L2) research has presented some evidence in reading and word recognition regarding writing systems. Muljani, Koda, & Moates (1998) examined if there would be facilitation of an alphabetic first language (L1) orthography on an L2 alphabetic system compared to logographic L1. They found the facilitation for intermediate-level learners: Indonesian learners of English were better than Chinese learners in their lexical decision task. Gairns (1992) also found that alphabetic L1 readers rely on phonological information much more than logographic L1 readers using a lexical decision paradigm: the alphabetic L1 readers made more errors on reading a nonsense word 'snoe' for 'snow'. Thus, it is assumed that the L1 orthography has an important role in L2 acquisition. In fact, L2 reading literature suggests that the L1 efficiency of

word recognition facilitates L2 reading and comprehension (Nassaji & Geva 1999, Koda 1996). Therefore, it is plausible that alphabetic L1 readers may have an advantage on reading an alphabetic L2 from the beginning. Moreover, Sasaki (2002)'s crosslinguistic experiment on L2 word recognition showed that the Italian advanced learners of English responded faster than the native English speakers in a word recognition paradigm. It indicates that the Italian learners could make use of their first language processing skills for the regular script in reading irregular English words. Thus, the strategies for L1 and L2 appear to be transferable and compatible.

L2 reading is essentially quite different from L1 reading. When we look at the backgrounds of L2 readers, L2 readers have prior reading experience in their L1s and L2 reading is crosslinguistic involving two or more languages. According to Koda (1996), the nature of L2 processing can be illustrated by the factors in (9).

- (9) a. L2 reading experience
 - b. Structural similarity or non-similarity between L1 and L2
 - c. Transfer of L1 processing experience

In relation to the crosslinguistic processing mechanisms, there are two predominant theoretical hypotheses to explain the process of word recognition in different orthographies (Gholamain & Geva 1999): the central processing (universalist) hypothesis and the script dependent hypothesis derived from the orthographic depth hypothesis. The central processing hypothesis posits that basic reading skills in all languages are influenced primarily by common underlying cognitive and linguistic factors such as verbal memory and rapid automatic naming. In contrast, the script dependent hypothesis posits that the development of word reading processes in different languages might vary due to orthographic regularity (Katz & Frost 1992). In short, accurate word recognition skills develop more slowly in languages with an irregular orthography, i.e., English. This

will also relate to the question why the patterns of reading disability in different languages vary, e.g., the high appearance of dyslexia in English speakers than in Italian speakers (Snowling 2000). Thus, L2 acquisition and the types of writing systems including the combination of L1 and L2 writing systems are strongly related each other in terms of cognitive processing skills. Therefore, it is important to consider the effect of orthography on language learning efficiency.

6. The Most Efficient Writing System

This paper has been emphasizing that a writing system should vary from one language to another. As we have seen, behind the writing system, there are language users who develop processing skills from their language. Written language processing skills are established strongly affected by the writing system. Therefore, the choice of a simple orthography which allows a most consistent writing system is indispensable for an artificial language.

For example, the 28 letters of the Esperanto alphabet, presented in (10), are based on a simple principle: unlike English, each letter corresponds to one sound and each sound corresponds to one letter (Eco 1995). In short, every letter must be pronounced. Therefore, it is supposed to be a regular (or shallow) orthographic language and it can be found easy by any readers of a Roman alphabet language. However, it is not quite visually simple or familiar to many of the Roman alphabet readers because of the addition of circumflexed letters, *ĉ, ĝ, ĥ, ĵ, ŝ, ŭ*. Further, it is not phonologically easy to learn because those pairs of two letters represent similar sounds such as <c> and <ĉ> or <h> and <ĥ> in (11a, b).

(10) The Esperanto alphabet

a, b, c, ĉ, d, e, f, g, ĝ, h, ĥ, i, j, ĵ, k, l, m, n, o, p, r, s, ŝ, t, u, ŭ, v, z

(11) a. <c> /ts/ as in <i>rats</i>	cent-o	hundred
	cerb-o	brain
<ĉ> /tʃ/ as in <i>church</i>	ĉiu	each one
	ĉef-o	chief
b. <h> /h/ as in <i>his</i>	hav-i	to have
	honor-o	honour
<ĥ> /x/ as in <i>loch</i>	Hin-a	Chinese
	ĥemi-o	chemistry
(12) a. <g> /g/ as in <i>go</i>	gem-o	gem, jewel
	Geno-o	Genoa
b. <ĝ> /dʒ/ as in <i>large</i>	ĝarden	garden

One might find difficulty to learn the relationship between unknown letters and sounds at first, and the visual similarity such as <g> and <ĝ> can also affect processing efficiency at later stages. It is understandable that the circumflexed letters were suppressed in an improved version of Esperanto and Ido, i.e., Novial by Otto Jespersen (Jacob 1943). He even excluded a few sounds, e.g., palatised sounds (/ð/ or /ñ/) for establishing the *phonetic* writing system, and also used special letters, e.g., <f> for /ph/.

Furthermore, the inconsistency with existing pronunciations may cause confusion, considered that learners of Esperanto would know other languages such as English. As seen in (12a), the Esperanto word 'gem' is pronounced as /gem/ while English 'gem' sounds /dʒem/. It would be easier if it was spelled as 'ĝem' so that the pronunciation is consistent to that of the English. Conversely, the English 'garden' is spelled 'ĝarden' in Esperanto indicating the sound /dʒarden/ (12b), although this sounds like the French word 'jardin' /ʒɑrdɛ̃/ for 'garden'.

One of the problems of the Esperanto writing system is derived from its orthographic structure where the spellings seem to have been established with mixed reference to orthography and phonology of base languages. The types of inconsistent reference can be

categorised into three groups as follows. Examples with each group of inconsistency with English are shown in (13a-c): 1) the orthographic representation in one of the base languages was adopted but not the phonology (13a); 2) the phonology was adopted but not the orthographic representation (13b); and 3) the orthographic and phonological forms were similar to those of the base language but they are slightly different for some reason (13c).

(13) <u>Esperanto spelling</u>	<u>Esperanto sound</u>	<u>English spelling</u>
a. scienc-o	/stsientso/	science
psikolog-io	/psikologio/	psychology
psalm-o	/psalmo/	psalm
ĉap-o	/tʃapo/	cap
b. ŝi	/ʃi/	she
ĉarm-a	/tʃarma/	charming
ĝin-o	/dʒin/	gin
c. poŝt-o	/poʃto/	post
ŝton-o	/ʃtono/	stone
gast-o	/gasto/	guest

Those words categorised in (13a) are easy to read visually but difficult to realise the sounds, while those words categorised in (13b) are easy to both pronounce and spell as long as the letter-phoneme correspondences in Esperanto have been mastered. The words categorised in (13c) are very similar to the English equivalents so that their minor differences could be difficult to learn.

The typical language has 5 to 7 vowels and the average 23 consonants. Therefore, the ideal writing system may have 30 letters or letter strings for 30 corresponding phonemes. It will be useful to add special but distinctive letters for some phonemes such as /θ/ /ð/ and /tʃ/ if these sounds are included in a new language. The choice of graphemic shapes according to the relationship with phonemes is essential in order to allow readers to develop simple processing skills. As for letter strings, for example, the letter <h> can be al-

lowed to overlap in the letter string <sh> for the phoneme /ʃ/ as long as the letter string represents only this sound. One of the features that makes the English orthography complicated is letter strings. The letter string <ch> has multiple sounds such as *character* and *chance* as well as *cliché*. Another feature is morphologically oriented spellings that do not associate with the current phonology illustrated in (14).

- (14) *heal* /hi:l/ – *health* /helθ/
 child /tʃaɪld/– *children* /tʃɪldrən/

- (15) Variation of the sounds of plural –s
 a. /s/ books
 b. /z/ pens, friends
 c. /ts/ hats

The plural –s (15) or past tense –ed are morphemes which have specific functions and meanings while the pronunciations vary according to the context in English. The morphological feature is useful for conveying the meaning of a word but the way English employs this feature is historical not educational, and it clearly shows inconsistency between orthography (or morphology) and phonology. However, since a new language will not have this historical precedence of phonology over morphology, its writing system will be able to combine a morphological feature as well as orthographic regularity.

In summary, an efficient writing system should include a regular orthography, conventional and visually distinguished letters, and a possible morphological feature. Firstly, a regular orthography can be developed by the consistent relationship between letters and phonemes. The one-to-one correspondence between a phoneme and a letter or a letter-string is essential. Although irregularity in an artificial language is uncommon, the hidden overlaps of phonemes may exist for letters due to the context or the combination with other let-

ters. In Esperanto, the different pronunciations apply for the letters ‘*ŭ*’ and ‘*j*’ according to the position of the letters.

(16) a. ending –*aŭ*

<*ŭ*> /w/ as in *will* → –*aŭ* /au/
e.g., *antaŭ* /antau/ ‘before’

b. ending –*oj* (Nom.pl.)

<*j*> /j/ as in *yell* → –*oj* /oɪ/
e.g., *libroj* /libroɪ/ ‘books’

Secondly, it is difficult for learners to familiarise unknown letters or symbols. Subtle differences between two letters would also result in inefficient processing. Therefore, visual features of orthography have to be controlled as simple as possible. The use of the basic Roman alphabet is advantageous for many learners unless no additional symbols are included. Finally, a morphological feature within a word is effective on lexical processing as it stimulates visual processes of a word to the lexicon, not only phonological encoding. Thus, the morphological feature should be included as long as orthographic regularity is maintained. On the other hand, the etymology is not very important if the morphology is incorporated because morphemes become active information to access the meaning as well as to produce new words.

In Esperanto, there are many consistent spellings of morphemes, e.g., the plural –*j* and the present tense –*as*, or consistent spellings to describe word status, e.g., the noun –*o* and the infinitive –*i*. These endings are useful to connect the visual form and the meaning or the grammatical status directly without deep analysis such as consulting the word order. However, the overuse of the visual consistency at the morphemic level may result in redundant information or production of unnecessary rules. For example, the present tense form and the infinitive could be represented with the same suffix or one with

and the other without a suffix, however Esperanto adopted the two different suffixes, *-as* and *-i*, respectively; or it would be sufficient if nouns represent the plural form, but Esperanto requires the plural form for both adjectives and nouns. These elements can make the language not only visually noisy but also phonologically tedious. The efficient writing system should have the consistent grapheme-phoneme correspondence as well as the minimum number of regular spelling rules in the orthographic structure; in addition, importantly those two aspects should be well balanced.

7. Conclusion

The English mental lexicon is supposed to be different from the French mental lexicon in terms of the phonological, orthographic and syntactical information it contains. However, as we have seen in recent second language acquisition research, the mechanisms of processing words, e.g., how the lexicon is accessed and how information is retrieved, can have common features across languages. Processing skills of written language develop in a particular language aspiring to the maximum efficiency in reading. Nevertheless, these skills can be applied to a different language in which mechanisms are relatively similar to the first language: L1 processing transfer can be found when common mechanisms exist. If a mechanism is completely different, learners may need to develop some new skill or strategy.

When it comes to the creation of an artificial language, the most important point for a language is 'easy to learn and use'. The ideal language should be compatible with the learner's first language so that they do not need to develop a different mechanism or processing strategy and rather they can make use of the existing linguistic system. This paper has been emphasizing the importance of efficiency in lower level language (i.e., orthographic and phonological)

processing and has pointed out that Esperanto needs the modification in order to provide an efficient writing system. The importance of orthographic regularity and simplicity should not be underestimated in terms of the acquisition of a new language.

Writing is a tool that humans have devised in order to communicate effectively, and presently it constitutes a very important part of the language function. Especially when people learn an artificial language, most of them would start from written language. Therefore, it is important to provide simple orthography and the simple writing system because even just for reading a word there are underlying processes and mechanisms which are developed and operated in the human brain.

References

- Allport, D. & E. Funnell. 1981. Components of the Mental Lexicon. *Philosophical Transactions of the Royal Society of London B* 295, 397-410.
- Carney, E. 1994. *A Survey of English Spelling*. London: Routledge.
- Crystal, D. 1997. *The Cambridge Encyclopedia of Language* (2nd edition). Cambridge: Cambridge University Press.
- Eco, U. 1995. *The Search for the Perfect Language*. Oxford: Blackwell.
- Ellis, A. 1848. *A Plea for Phonetic Spelling* (2nd edition). London: Phonetic Depôt.
- Gholamain, M. & E. Geva. 1999. Orthographic and Cognitive Factors in the Concurrent Development of Basic Reading Skills in English and Persian. *Language Learning* 49.2, 183-217.
- Graddol, D., J. Cheshire, & J. Swann. 1994. *Describing Language* (2nd edition). Buckingham: Open University Press.
- Halliday, M. 1987. *Spoken and Written Modes of Meaning, Comprehending Oral and Written Language*. San Diego, CA: Academic Press.
- Harley, T. 2001. *The Psychology of Language: From Data to Theory*. Hove: Psychology Press.
- Harris, M. & M. Coltheart. 1986. *Language Processing in Children and Adults*. London: Routledge.

- Jacob, H. 1943. Otto Jespersen: His Work for an International Auxiliary Language. International Language (Ido) Society of Great Britain. Available at URL <<http://www.geocities.com/Athens/Forum/5037/OJ.html>>.
- Jespersen, O. 1930. *Novial Lexike*. Heidelberg: Karl Winters.
- Katz, L. & R. Frost. 1992. The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In R. Frost & L. Katz (eds.), *Orthography, Phonology, Morphology and Meaning* 67-84 Amsterdam: Elsevier.
- McClelland, J. & D. Rumelhart. 1981. An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of the Basic Findings. *Psychological Review* 88, 375-407.
- Perfetti, C. & L. Tan. 1998. The Time Course of Graphic, Phonological, and Semantic Activation in Chinese Character Identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.1, 101-118.
- Rubenstein, H., S. Lewis, & M. Rubenstein. 1971. Evidence for Phonemic Recoding in Visual Word Recognition. *Journal of Verbal Learning and Verbal Behavior* 10, 645-657.
- Sasaki, M. 2003 (to appear). Bilingual Processing Strategies in Word Reading. Proceedings of the Second Symposium on Bilingualism, October 2003, Vigo.
- Shallice, T., P. McLeod, & K. Lewis. 1985. Isolating Cognitive Modules with the Dual Task Paradigm: Are Speech Perception and Production separate Processes? *Quarterly Journal of Experimental Psychology* 37A, 507-532.
- Snowling, M. 2000. *Dyslexia* (2nd edition). Oxford: Blackwell.
- Van Orden, G. 1987. A ROWS is a ROSE: Spelling, Sound, and Reading. *Memory and Cognition* 15, 181-198.
- Venezky, R. 1970. *The Structure of English Orthography*. The Hague: Mouton.
- _____. 1995. How English is Read: Grapheme-phoneme Regularity and Orthographic Structure in Word Recognition. In Taylor, I. & D. Olson (eds.), *Scripts and Literacy* 111-129. Dordrecht: Kluwer.
- Wydell, T., K. Patterson, & G. Humphreys. 1993. Phonologically Mediated Access to Meaning for Kanji: Is a Rows Still a Rose in Japanese Kanji? *Journal of Experimental Psychology: Learning, Memory and Cognition* 19.3, 491-514.
- Wydell, T. & B. Butterworth. 1999. A Case Study of an English-Japanese Bilingual with Monolingual Dyslexia. *Cognition* 70, 273-305.