

# Alphabetic vs. non-alphabetic writing: Linguistic fit and natural tendencies

Antonio Baroni

This article has two main purposes. The first one is to prove that the alleged superiority of the alphabet to other writing systems (syllabic and logosyllabic ones) is an ethnocentric prejudice and that the optimality of a writing system has to be measured following a series of criteria which cannot be reduced to the faithful mapping of sounds. The second one is to incorporate into the graphemic theory external data and new approaches to develop new methods of investigation and to emancipate graphemics from phonology. The structure of the article is composed of seven parts. First of all, we discuss some definition problems; then, in the introduction, the main points of view about the alphabetic principle are exposed and in chapter 2 the relationships between writing systems and language perception are investigated. In chapter 3 we attempt to define some criteria to judge the degree of optimality of the different writing systems. In chapter 4 we try to find some patterns of predictability of the degree of opacity and transparency of some of the main European writing systems (the opaque English, French and Danish orthographies and the shallow Finnish and Italian orthographies). In chapter 5 we shortly examine the natural evolution of writing in recent times: Internet, SMS and new writing systems. Finally, in chapter 6 we try to draw some temporary conclusions.\*

## *Definitions*

Before starting our investigation about the degrees of optimality of the different writing systems, it would be better to deal with definition problems.

### *\* Symbols and abbreviations*

[a]	phone
/a/	phoneme
<a>	grapheme, graphoneme or allograph
{a}	morpheme
a	morphoneme
sing.	singular
pl.	plural
m.	masculine
f.	feminine

By WRITING, we mean a series of graphic symbols arranged on a surface (be it physical, like a sheet of paper, or virtual, like a screen), in a certain order and in a certain sequence, so that this series is likely to be interpreted (read) by an interpreter (reader) who knows how to decipher the meaning of these signs.

A writing system can then be CENEMIC, if its elements stand for units of phonic expression, or PLEREMIC, if its elements stand for units of content (Coulmas 1989: 38-39). Nevertheless, one must not forget that there is no pure writing system: in every tradition cenemic and pleremic components coexist, to different extents.

Among pleremic writing systems, some display PICTOGRAMS, whose shape recalls iconically the object represented, e.g. <☉> Egyptian pictogram meaning 'sun', whereas IDEOGRAMS, also iconic, represent abstract concepts, e.g. Chinese <三> *sān* means 'three' and is composed of three strokes. In LOGOGRAPHIC SYSTEMS, the graphic elements stand for a word or for a morpheme.<sup>1</sup>

A cenemic writing system can be a:

- SYLLABARY, where every graphic unit stands for one syllable, normally a CV-type of syllable; Japanese kana, Linear B and Cherokee are syllabaries;<sup>2</sup>
- ABJAD (Daniels & Bright 1996), where consonants are represented but vowels are not, even if there is the possibility to add diacritic vocalic signs to disambiguate. This system is typically adopted by Semitic languages, which share the characteristic to possess triconsonantic lexical roots, e.g. from Arabic ذبح *dbḥ* → ذَبَحَ *dabahaḥ* 'he sacrificed', ذَبَحْتِ *dabaḥṭa* 'you-m. sacrificed', ذَبَحَ *dabbaḥa* 'he slaughtered', etc.
- ABUGIDA. It is a syllabary where the graphic elements standing for the consonants and the ones standing for the vowels are recognisable, e.g. Indian devanāgarī, Ethiopian Ge'ez script;
- ALPHABET, where ideally all the phonemes of a language are noted by separate elements;
- FEATURAL WRITING (e.g. Korean hangŭl), an alphabet where the shapes of the graphic signs correlate with phonemic distinctive features.

A SCRIPT selects the modality of graphic expression making it fit to the structure of the language it represents, e.g. Latin and Greek had the same writing system (an alphabet) but employed different scripts.

An ORTHOGRAPHY selects the possibility of a script through uniform and standardised procedures of correspondence between graphic elements and linguistic units. For example, Italian and French share the

same writing system (alphabet) and the same script (Latin) but follow different orthographic rules. Interpunction, spacing and capitalisation are part of the orthography.

An orthography can be TRANSPARENT or SHALLOW if, given a set of basic rules, it is always possible to read and write a word, even an invented one (e.g. Finnish, Italian, Spanish, Turkish, Georgian, etc.). In OPAQUE orthographies, on the contrary, the correspondence between spoken words and written words cannot be reduced to a set of rules and, depending on the degree of opacity, one must learn by heart a certain number of graphic words (e.g. French, English, Danish, Khmer, Chinese, etc.).

We define GRAPHEME the minimal meaningful graphic unit, in every tradition; its meaning can be cenemic (e.g. <a> → /a/), both cenemic and pleremic (e.g. Chinese <下> *bě̀n*, ‘down, under’) or only pleremic (e.g. <h> in Italian *ha* ‘has’ whose function is just to indicate that this word belongs the paradigm of the verb *avere*, ‘to have’, although it does not correspond to any sound). If we consider merely the sound-letter correspondence, then we are not talking about graphemes but GRAPHONEMES (Hořejší 1971: 186), e.g. in French <c> and <h> are graphemes, but <ch> is a graphoneme,<sup>3</sup> inasmuch as it stands for one phoneme (mostly /ʃ/ but sometimes /k/).

## *1. Introduction*

Currently, the most widespread point of view in Western linguistics about graphemics is still biased by several prejudices, all of which can be reduced to the teleological position formally expressed by Ignace Gelb that there would have been a constant improvement in the historical evolution of writing, in which the alphabet would be the pinnacle of perfection, regarded both as the cause and the effect of a high degree of civilisation. It goes without saying that any deviation from this principle is considered as an aberration or an imperfection; non-alphabetic systems are therefore deemed to be inferior to alphabetic ones and opaque orthographies inferior to shallow ones. According to Gelb, writing followed a linear evolution, passing from an early stage of logography and subsequently switching to the syllabic principle and culminating with the alphabet. This path is seen as unavoidable (Gelb 1963: 240).

Other scholars agreed with this view: Diringer (1948) calls the alphabet “a key to the history of mankind”, while Ong (1986) goes so far as to claim that Latin alphabet will replace Chinese characters

as soon as everyone in the People's Republic of China share the same language. The idea of the superiority of the alphabet (and of Western civilisation), as opposed to other writing systems (and cultures), is promoted also by McLuhan (1964), McLuhan & Logan (1977), Innis (1991) and Logan (2004).

Among the opponents to this position, Venezky (1970: 120) claims that “[the fact] that homo sapiens is somehow more at ease with a one-letter one-sound system has often been assumed, but no evidence has ever been produced to substantiate this limitation on man’s mental capacities”. Also Gleason (1961: 419) disagrees with Gelb’s theory, since he doubts that “an alphabet which did accurately record speech would be practical”. Coulmas (2009a: 105) points out that:

“[f]rom a Near- and Far-Eastern point of view [the] validity [of alphabetocentrism] is not so evident. Japanese kana, for example, is much simpler and more elegant than almost all scripts using an alphabetic notation. The system is so simple that children can be expected to have mastered it before they enter elementary school. There is no need to teach it there”.

Daniels (1992: 83) brings also some scientific data, arguing that the phoneme, the unit on which alphabetic writing is based, is not a natural unit,<sup>4</sup> given that

“[i]nvestigations of language use suggest that many speakers do not divide words into phonological segments unless they have received explicit instructions in such segmentation comparable to that involved in teaching an alphabetic writing system”.

## *2. Writing and language perception*

Not surprisingly, perfectly phonemic orthographies do not exist. Moreover, mixed systems such as English and French orthographies, defective writing systems such as the Arabic abjad and logographies such as the Chinese one are the most widespread writing systems in the world. Some scholars argue that it was the invention of the alphabet to affect language perception; modern linguistics could not have been theorised without the constant influence of the alphabet. According to Faber (1992: 127),

“segmentation ability as a human skill may have been a direct result (rather than an impetus to) the Greek development of alphabetic writing. Thus, the existence of alphabetic writing can not be taken *eo ipso* as an evidence for the cognitive naturalness of the segmentation that it reflects. (...) [W]e as linguists feel that, because we *can* describe linguistic system in terms of phonetic segments, we *must* do so [but] (...) every technical linguistic tradition that refers to segments arose in an alphabetic milieu (...). In contrast, the indigenous Chinese linguistic tradition (...) has as phonological primitives syllables *initials* and *finals*,<sup>5</sup> that is, onsets and rhymes. This analytical division is not supported by the logographic Chinese orthography, a lack which strengthens the force of the analysis”.

Several studies (Foss & Swinney 1973, Liberman et al. 1974, Morais et al. 1979, Cossu et al. 1988, Daniels 1992, Goswami 2005) have shown that the perception of the phoneme, though having a psychological reality, always follows syllable-awareness. The phoneme cannot be separated from the syllable, anyone can verify this by recording his/her own voice in a computer and then processing and analysing it using a speech spectrum software (e.g. Praat); one will find out that it is impossible to listen separately to the consonantic and to the vocalic part of the syllable in isolation, the result being something similar to an electronic reverberation. These data would explain the ubiquity, in different cultures, of syllabaries and the rarity of alphabets.

Another experiment (Cho & McBride-Chang 2005) has shown that Korean children apply their syllable-awareness to learn hangŭl, the native Korean script, but when learning English orthography, they use their phoneme-awareness; depending on the script to learn, they use different skills.

However, one must not forget that, even if writing affects the metalinguistic view of language users, any writing system invented through history could not have seen the light without a prior linguistic consideration. After all, if Greeks had not thought about a concept similar to that of phoneme, they could not have invented a segmental writing system.

### *3. The optimality of writing systems: some possible criteria*

Shall we therefore conclude that the prestige acquired by the alphabet since its first appearance is solely due to ethnocentric prejudices? Well, we think that the alphabet has its own merit beyond its prestige.

The syllable is a more salient perceptive unit so an ideal script, if it exists, should be based on it. But syllabaries have always caused several problems, since they should represent all the syllables of a language. When a language has a syllabic structure that is mainly CV, the syllabogram inventory is still manageable, but when it comes to Indo-European languages like German or Russian, the list of syllabograms would be endless. The devanāgarī script uses graphemes for single vowels and graphemes for consonants with an inherent vowel /e/ (transliterated <a>). If the syllable to represent graphically comprehends another vowel, diacritics are used. Conjuncts are employed for complex syllables like CCV or CCCV but it is often hard to reconstruct the way in which they have been composed. Moreover, the number of symbols to memorise, if one considers the conjuncts as well, rises considerably (Masica 1993: 162). The devanāgarī script is an abugida which, albeit being a sophisticated system, is less flexible than the alphabet in transcribing foreign words and consonant clusters.

Logographic scripts are very cumbersome systems to learn; it takes several years for Chinese children to master the basic graphic system and the knowledge of a larger number of characters is a life-long learning process. Japanese children need considerably less time even if they have to learn not only one, but three systems: the kanji series (Chinese characters with an ideo/logographic function), hiragana (syllabograms) and katakana (syllabograms used mainly for foreign words), not to mention the Latin alphabet. Nevertheless, Chinese logographic system has one great advantage: it is interdialectal; the People's Republic of China is linguistically fragmented and regional languages and dialects are not always mutually intelligible, but everybody manages to communicate thanks to writing and that is one of the reasons why Chinese script is still in use today (Coulmas 1983: 246).

The script that seems closest to the ideal is the Korean hangŭl, inasmuch as it is iconic in relation to both speech articulation and syllable organisation and is able to graphically render very complex syllables such as *sang* <쌍> and *balp* <뵵>, not being excessively tied to its own linguistic fit. Besides these particularities, hangŭl is an alphabet since every phoneme is represented; so does that mean that the alphabet is indeed the best writing system? It is difficult to judge the optimality of a given writing system because different criteria have to be taken into account and they are often in opposition with each other. We will consider five main criteria:

### *3.1. Maximum distinctiveness*

According to the first criterion, the Latin alphabet would be one of the least functional, since the differences between the various graphemes are not really striking and it is easy to confuse them with each other. On the contrary, Chinese characters, Egyptian hieroglyphs or Maya glyphs differ greatly one from the other. Compare sequences as <aeiou> vs. <嘸軋芗事>, <𠄎𠄎𠄎𠄎𠄎𠄎> and <𠄎𠄎𠄎𠄎𠄎𠄎>.

Sampson reports an anecdote (Chiang 1973: 3-4), according to which literate Chinese have claimed that European script gives an impression of monotony and lack of distinctiveness, something similar to what we might experience when encountering a page printed in Morse code (Sampson 1985:164). Apparently, Hebrew abjad is even less visually distinct than Latin alphabet; the latter (in its lower case version) presents at least ascending and descending lines such as in <b, f, k, t> and <g, j, p, y> which make for a greater recognisability (Sampson 1985: 94). Thanks to these lines, pre-literate children would already be able to distinguish some written words basing their decision on the “bouma” (from the name of Dutch psychologist Herman Bouma, meaning the contour of written words), normally when they are around 4 years old (Pontecorvo 1994: 278). On the contrary, most Hebrew graphemes consist of a horizontal line on the top and a vertical line on the right: <סםתתופם> and some experiments (Gray 1956:59) suggest that readers of Hebrew and Arabic make longer eye-fixations than European readers.

### *3.2. Size of the graphemic inventory*

According to the second criterion, the Latin alphabet, with a number of graphemes comprised between 20 and 40, depending on the orthography considered, would be relatively fast to memorise, whereas it takes several years to learn the thousands of Chinese characters.

The different mnemonic weight has an evident neurological counterpart: it appears that Chinese children who suffer from dyslexia, characteristically present an under-activation of the Exner area (in the left medial frontal region of the brain), whereas the problem for European dyslexic children seems to be caused by an anomaly in the left temporal lobe; this is because the main difficulty of learning written Chinese is not at the phonological level but at the mnemonic level, since one needs to remember at least 3000 characters. To cope with that, Chinese kids use the visuo-motory memory necessary to draw the characters (Dehaene 2009: 283-285).

### 3.3. Cognitive salience

By the criterion of cognitive salience we mean the cognitive effort required to transfer words from speech to writing and vice versa. The higher the level of abstraction, the greater the required cognitive effort. Gleitman & Rotzin (1977 cited in Martlew 1983: 261), two psycholinguists, affirm that “[y]oung children (...) are aware of language as meaning units, only later aware of the phonological and syntactic substrata of language”. From this point of view, the alphabet would be the most complicated system and pictography would be the simplest.

### 3.4. Maximum naturalness

According to the criterion of maximum naturalness, the simplicity of the graphemes is regarded as one of the most important factors. Also in this field, alphabets are among the easiest to learn, but so are Japanese kana and Korean hangŭl.

Perhaps it is not obvious what we mean by simplicity: our intuition consists in the fact that recurring features which are found in most existing writing systems are somewhat natural for human beings. An interesting study (Changizi & Shimojo 2005: 267) compared more than 100 writing systems, attempting to find similarities between them and it turned out that

“[b]ecause writing systems are under selective pressure to have characters that are easy for the visual system to recognise and for the motor system to write, these fundamental commonalities may be a fingerprint of mechanisms underlying the visuo-motor system”.

The results indicate that a character is preferably composed of three strokes; this is because it is still possible, for the brain, to process, in a parallel way, three elements but beyond this number, more time is required.<sup>6</sup> It goes without saying that writing systems containing several characters composed of more than 3 strokes are considered harder to reproduce manually.

Redundancy seems to be another important factor, since most graphemes are redundant, namely, not all the strokes composing them are necessary for their recognisability (Changizi & Shimojo 2005: 273).

According to the neurophysiologist Stanislas Dehaene, all writing systems share traces of an underlying structure. Although different writing systems were historically invented independently one from the other, his NEURONAL RECYCLING THEORY (NRT) predicts that

human inventiveness must be limited by the organisation of the circuits of our brain (Dehaene 2009: 201-204). The NRT tries to explain how human beings are so at ease with reading even if writing was invented only 5400 years ago and the alphabet is just 3800 years old, since our genome would not have had the time to develop brain circuits specific for reading. The brain would not have developed new circuits, but neurons which once were employed for one task, switched their function and they specialised in discriminating graphemes from other visual stimuli, such as faces, objects, numbers, etc.

The Japanese neurophysiologist Keiji Tanaka has discovered that chimps possess neural sensors which react to elementary shapes and their function is to recognise objects (Tanaka 2003: 90-99). These simple shapes are a sort of alphabet because, combining them, every object can be described and, moreover, they look surprisingly like some elements of our writing systems (Dehaene 2009: 153). That is why Dehaene calls them “protoletters”, among which the most widespread in all writing systems are similar to <T, F, Y, L, 8>. What do these protoletters have in common? When they impress our retina, they appear as structured objects, and the cerebral cortex judges them as non-accidental (Dehaene 2009: 169).

All writing systems, be they alphabetic, syllabic or logographic, draw on a small set of stroke configurations whose spread follows a universal tendency; the most frequent configurations are more likely to be found also in nature and are therefore encoded by neurons in the inferior temporal cortex even before learning how to read (Changizi et al. 2006: 117-139).

Since neurological studies (Dehaene 2009: 66) also showed that, regardless of the writing system considered, human beings use the same part of the brain to read (the left occipitotemporal region, with minimal differences), some configurations underlying our graphemes must be universally easier (more natural) than others.

### *3.5. Inner consistency*

When we talk about the criterion of inner consistency, we refer to the degree of iconicity<sup>7</sup> in relation to language mapping. To put it down more clearly, we will make some examples. Among cenic writing systems, hangŭl displays a high degree of consistency, since similar characters stand for similar sounds, whereas the Latin alphabet is only partially consistent: some graphemes which stand for similar sounds look similar too, e.g. <m> and <n> (both nasal), <s> and <z> (both sibilant, dental or alveolar, depending on the language), <b> and

<p> (both labial), <u> and <v> (both labial), but they are the minority of the cases, since the similarities in shape of other characters are completely unrelated to the sounds they stand for: <N> and <Z>, <b> and <d>, <p> and <q>, etc. The háček employed in some Slavic languages and the German umlaut are good examples of consistency of a graphic system, since: /s/ : /ʃ/ = <s> : <š>, for example, in Czech, so that the háček stands for [+palatalised] and /u/ : /y/ = <u> : <ü> in German, so that the umlaut stands for [+front].

Pictographies should be, ideally, highly consistent, since they should represent in a univocal way concepts but here one needs a common background or the risk is to incur serious misunderstandings. Ideographies and logographies are not consistent at all in relation with phonology and may display different degrees of iconicity with the meanings they convey, depending on the given system.

Of these five criteria that we tentatively outlined, one can not choose the most important one: the degree of greater or lesser simplicity will also depend on the cognitive strategies of individuals.

### 3.6. *Other criteria*

One should also consider:

- the point of view of the reader vs. the point of view of the writer;
- the point of view of the native speaker vs. the point of view of the non-native speaker.

In the act of reading, words function as units of meaning. The design of the word may be composed of strokes (as in ideograms) or of letters of the alphabet (also composed of strokes), it does not matter to the reader as long as his/her mental orthographic lexicon is activated by the word recognition process, in which written words are perceived as visual Gestalts by expert readers. The same does not apply to the writing process, in which the production of the graphic sequence is analytic rather than synthetic.

For the reader, the criteria of maximum distinctiveness and maximum naturalness are extremely useful, whereas the writer is probably more comfortable with a reduced number of symbols easy to reproduce. Similarly, for native speakers a phonologically detailed information is not necessary, they just need a phonological cue and are then able to identify the word thanks to the context. This is not true for non-native speakers. So, to whom should we pay more attention? Readers or writers? Native or non-native speakers? According to Sampson (1985: 212),

“[a]ny literate adult, even a professional author, reads far more than he writes; so if (...) the ideal script for a reader is a somewhat unphonemic script, (...) the balance of advantage has been tending to move towards the reader and away from the writer: extra trouble in writing a single text can now be massively repaid by increased efficiency of very many acts of reading that text. (...) [I]t is worth spending more time nowadays to learn an orthography, if the extra time is the cost of acquiring a system that is relatively efficient once mastered, because the period during which the average individual will enjoy mastery of an orthography is now longer than it used to be”.

So far, it seems unavoidable to consider other elements to define the optimality of a writing system (cf. Coulmas 2009b). The most likely conclusion we can reach is that there is not such thing as a system that would be optimal for every language in every society, since every language has got its own linguistic fit.

#### *4. The linguistic fit of the Latin alphabet: Opaque and shallow orthographies*

##### *4.1. English*

English orthography is undoubtedly one of the most debated topics in graphemic studies and this is due to the prestige of English as the global lingua franca. One striking aspect of English orthography is its sobriety, since it does not display any diacritics or special letters, it is therefore optimal according to the criterion of the ease of reproduction because one never needs to use special keyboards or to worry about accents, diaeresis, and so on.

Nevertheless, English orthography is considered to be very hard to master, both for native and non-native speakers. It used to be closer to spoken English in the past, but around 1400, when a common orthographic standard was established, a peculiar linguistic phenomenon, the Great Vowel Shift, started to change the pronunciation of almost all vowels and it came to an end only around 1600. That is why English vocalic graphemes are pronounced so differently from all the other European languages; they normally have two possible phonic meanings, corresponding to the “free pronunciation” or to the “checked pronunciation”. The former is usually indicated by the diacritic grapheme <e> on the following graphic syllable, compare *mat* /mæt/ vs. *mate* /meɪt/, *bit* /bɪt/ vs. *bite* /baɪt/, *con* /kɒn/ vs. *tone* /təʊn/, etc. The diacritic <e> has provoked a lack of isomorphism between the

graphic syllable and the phonic syllable, inasmuch as words composed of two graphic syllables often correspond to one phonic syllable.

Moreover, during the Renaissance until the 18<sup>th</sup> century, many etymological letters were introduced, so that *cors*, *langage*, *doute*, *samon* became *corpse*, *language*, *doubt*, *salmon*. Aesthetic and typographic questions modified the length of some words: <e> was added or removed to adjust the margins of the page: *doe*, *goe*, *heere* vs. *do*, *go*, *here* (Mitton 1996: 19).

Throughout the last century, many have tried to understand if English orthography possessed a sort of underlying principle (see Venezky 1970, Albrow 1972) and some have pointed out that the so-called chaos of English orthography is, unexpectedly, often useful. The most widespread example is the distinction of homophones, e.g. *rite* vs. *write* vs. *right* vs. *Wright*; in this way, like visual morphemes (Bolinger 1946), these words speak directly to the eyes avoiding the possibility of confusion in the process of understanding the text.

According to Chomsky and Halle, English orthography is a good representation of the underlying form of lexemes (Chomsky & Halle 1968: 47-49). An optimal orthography, they argue, should be a deep one, having a single representation for each lexical entry. Spanish spelling would thus be somewhat inferior to that of English: the conjugation of the verb *pedir* 'to ask' undergoes a morphonemic alternation: the sound /e/ of the root {ped-} becomes /i/ in a stressed position, so that the paradigm is *pido*, *pides*, *pide*, *pedimos*, *pedís*, *piden*. Since this transformation is quite regular and predictable for a native speaker, a deep orthography would always write <ped-> (Sampson 1985: 200). One might agree with Chomsky and Halle, if English orthography were truly a deep one, but this assertion can not be accepted uncritically analysing a number of examples chosen ad hoc. Firstly, it seems implausible that native English speakers have such a fine and detailed awareness of the deep phonological processes of their own language, and secondly, there are many cases where the alleged depth of English orthography does not show: one writes *speak* but *speech*, *collide* but *collision*, *sight* but *see*, etc.

A solution is to consider written English as a system that initially was strongly based on phonology and then went slowly acquiring a somewhat logographic component. This trend could be seen as an adaptation of writing to the change of its functions: a phonological notation was absolutely necessary in a period when the only *raison d'être* of a written text was to be read aloud, but with the diffusion of silent reading, this is no longer needed. As Berry (1977: 10) suggests, "[t]he reader reads fluent English or French or German

efficiently only insofar as he treats the written language as if it were ideographic”.

In the particular case of English, the evolution of its spelling shows also a striking parallelism with the evolution of spoken English: the spelling was more shallow at the time when English was a flexive language, closer to other European ones, and evolved into a graphic system with a high degree of logography and morphography until today, while English is considered typologically closer to an isolating language like Chinese rather than to its Indo-European cousins.<sup>8</sup>

#### *4.2. French*

The development of written French is somehow tied to that of English orthography, since French is responsible for many changes that occurred in written English, either directly through the Norman invasion, or indirectly, thanks to the prestige of French culture during most of the Modern Age. The two graphic traditions share many similarities: French also, at the beginning, was written in a way which reflected the phonetic reality simply and precisely. It was the spelling employed by jesters, singers and poets who, because of their job, needed a rapid transcription of the texts they used, characterised by a typically oral structure. When French became important in public life replacing Latin in official documents, the scribes began to add many unnecessary letters. Flaunting the presence of etymological letters that were no longer pronounced was a way to confer French the same prestige of its ancestor (Fournier 1940: 261).

Among the main difficulties of current French orthography, we can name the following:

- the alternation between diacritical accents and silent letters to indicate the timbre of a vowel, e.g. *fidèle* /fidɛl/ vs. *nette* /nɛt/;
- /E/ in an open syllable can be realised as [e], [ɛ] or [ə], whereas in a closed syllable it is always [ɛ]; however, there is a medium [E] that is halfway between [e] and [ɛ] and usually appears before a syllable containing [ə]; the instability of the schwa makes the syllable itself unstable and it can be realised as an open syllable or as a closed syllable, so that the timbre can be itself open or closed. The French Academy normally indicates this medium [E] by <é>, that usually stands for [e], although the pronunciation tends towards [ɛ], e.g. *médecin* [medɛsɛ̃] can switch to [mɛdsɛ̃];
- among the sixteen vocalic phonemes of French, four are nasal phonemes, whose notation is still problematic;

- many graphemes are employed to distinguish between homophones (*cinq*,<sup>9</sup> *sein*, *sain*, *seing*, *saint* stand all for [sɛ̃]; cf. English *buy*, *bye*, *by* for [bai], Danish *vær*, *hver*, *værd*, *vejr* for [vɛ̃ʁ]); moreover, many letters are conserved for matters of prestige, sometimes etymologically motivated, (*philosophie*, *xénophobie*) sometimes not (*lys* from Latin *lilium*, *nénuphar* from Arabic *ninufar*).

One of the most striking aspects of French orthography is the discrepancy between oral and written morphology. Written French requires much more grammatical knowledge than spoken French. In speech, there is often no difference between the singular and the plural of nouns and only the article, context or agreement can disambiguate the grammatical number of the noun. Some nouns are not even differentiated even between the masculine and the feminine form. In verbal conjugation, the first three persons and the sixth have often the same desinence. In writing, on the contrary, masculine and feminine are normally differentiated, the plural is always indicated by <-s>, the verbal conjugation preserves different desinences for almost every grammatical person, and so on, e.g. <l'ami> 'the friend (m.)' [lami] vs. <l'amie> 'the friend (f.)' [lami]; in the series *je parle*, *tu parles*, *il parle*, *nous parlons*, *vous parlez*, *ils parlent* ('I speak, you-sing. speak, he speaks, we speak, you-pl. speak, they speak') *parle*, *parles*, *parlent* stand all for [parl(ə)].<sup>10</sup> These data could lead us to the conclusion that French native speakers' morphological competence is often expressed only graphically.

It is interesting here to point out that written French morphology is much more natural according to the Natural Morphology framework (see Dressler 1987a, 1987b, 1990, Dressler et al. 1987, Mayerthaler 1981, Kilani-Schoch 1988, Wurzel 1994, Dressler, Mayerthaler et al. 1987) than its spoken counterpart. According to the principle of DIAGRAMMATICITY or CONSTRUCTIONAL ICONICITY, forming the plural adding a suffix, such as the morpheme {-s}, corresponds to the first degree of the scale of diagrammaticity, namely, the most natural one, and so it happens in written French: <chien> (sing.), <chiens> (pl.). But in spoken French, the plural is normally formed without any alteration of the base, through the morphological technique of METAPHORICITY, eg. /ʃjɛ̃/ (sing.), /ʃjɛ̃/ (pl.). Metaphoricity corresponds to the third degree in the scale of diagrammaticity. If we consider the formation of the feminine, written French still employs the AGGLUTINATIVE AFFIXATION (1<sup>st</sup> degree) but spoken French employs SUBTRACTION, which is considered anti-iconic (since feminine is more marked than masculine).

Concerning morphological naturalness, writing has the primacy also when it comes to MORPHOTACTIC and MORPHOSEMANTIC TRANSPARENCY (a form is morphotactically and morphosemantically transparent if it possible to find, in the structure of the signifier, morphs which correspond to components of meaning) and TRANSPARENCY OF ENCODING (according to which, synonymy, homonymy, suppletion and allomorphy are not natural) (Thornton 2007: 164-165, Crocco Galèas 1998: 25-41). Phonological naturalness is often in contrast with morphological naturalness; the natural phonological processes tend to diminish the articulatory effort, whereas morphological processes tend to increase it to improve the perception for the listener.

Part of the opacity of French orthography may be due to the need to preserve a certain degree of distinctiveness in a language where many important grammatical distinctions were lost because of phonological processes.

#### *4.3. Danish*

It is interesting to also consider Danish orthography here because it is somewhat similar to English orthography, but unlike the English one its use is limited to a relatively small community.

According to a small scale study comparing 13 European orthographies (Seymour, Aro & Erskine 2003), Danish and English children lag far behind other children in reading and writing performances by the end of the first year of school. The origin of Danish's orthographic depth may be found in the choice of the first scribes, during 13<sup>th</sup> and 14<sup>th</sup> centuries, to select diasystemic spelling forms that did not reflect any specific dialect (a sort of interdialectal writing system, like Chinese script). As a result, the orthography they created reflected the archaic pronunciation, so for example, *lov* [lav] was spelled <logh>, although the last phoneme was already [v]; similarly, final [ð], originally [θ], was still spelled <th>.

Since the 1200s, Danish has undergone several phonological changes that did not affect related languages such as Swedish, which has a much more shallow orthography (Elbro 2003: 33). Danish, Swedish and Norwegian maintain a certain degree of mutual comprehensibility, but spoken Danish is the most difficult to understand (Basbøll 2005: 7). What makes the Danish language particularly unsuitable for a phonologically transparent notation through the Latin alphabet are the following features:

- an extremely rich vowel system, with far more distinctions in vowel quality, both in rounded and unrounded vowels, than other related languages;

- radical reduction processes, which make the language very hard to understand for foreigners; these phenomena affect vowels, obstruents and approximants/glides as well;
- schwa-assimilation, which means that the schwa, as a final neutral vowel, can either manifest itself as such or be assimilated by the preceding consonant, turning it into a syllabic segment, e.g. *hedde* ‘to be called’, [‘heðə] vs. [‘heðð];
- a unique feature of Danish is the *stød* (glottal stop), whose presence is unpredictable in synchrony and affects word formation and other morphological processes (Basbøll 2005:8). Even though *stød* is not generally indicated consistently in the orthography, there is a certain correlation between *stød* and unpronounced <d>, e.g. *spild* ‘waste’ [sɕilʔ], *hund* ‘dog’ [hunʔ], *mand* ‘man’ [manʔ] vs. *spil* ‘play’ [sɕel], *hun* ‘she’ [hun], *man* ‘one, people’ [man];
- many morphonemes can have different phonological and phonetic counterparts, e.g. the morphonemes |v| and |g| can be both realised phonologically as /v/ and phonetically as [ʊ]; in these cases the orthography tends to operate at a morphonemic level, rather than at a phonological or phonetic one, cf. <filolog> ‘philologist’, |filolo:g|, /filo‘lo:ʔv/, [filo‘lo:ʔʊ] and <sav> ‘saw’ (noun), |sa:v|, /sa:ʔv/, [‘sæ:ʔʊ] (Basbøll 2005: 74-77);
- given a certain number of cases of homophony between words which differ greatly with regard to their grammatical function, Danish employs the orthography as a means of disambiguation, e.g. *at bore* (‘to drill’, infinitive form) and *borer* (‘drills’) sound the same but are spelled differently; similarly, the homophonous endings *-ene* (plural definite nouns) and *-ende* (present participle of verb) are distinguished in writing for morphological reasons (Elbro 2005: 40).
- another striking feature of Danish compared to other European languages is the distinction between prosodic commas and grammatical commas; during the 20<sup>th</sup> century, people tried to follow the grammatical use of commas, namely, commas around sentences regardless of utterance prosody, but later, in the 90s, the existence of two competing systems began to be felt as a problem by many. Today the placement of a comma before a subordinated sentence is optional (Basbøll 2005: 89-90), cf. English *He says (that) it tastes good*, Italian *Dice che è buono*, Danish *Han siger, det smager godt*.

#### *4.4. Finnish*

In the scientific literature, the most common example of a shallow orthography is undoubtedly Finnish; this Finno-Ugric language appears, in its written form, to be very close to the phonemic ideal, inasmuch as each grapheme corresponds to one phoneme and there are no ambiguous or contextual graphemes (the only exception is the digraph <ng> that stands for /ŋ/). Furthermore, Finnish orthography represents both vowel and consonant length, which are distinctive in speech, by doubling the grapheme, e.g. <mutta> /mut:a/ 'but' vs. <muuttaa> /mu:t:a:/ 'to change' vs. <muta> /muta/ 'mud'.

How is such a nearly perfect system possible? Firstly, the Finnish language is spoken by more or less 6 million people, the dialectal fragmentation is minimal and almost the entirety of its speakers live between Finland and Sweden; secondly, before the 16<sup>th</sup> century there was no Finnish literature, as the Bible and the academic publications were written in Latin or in Swedish and a true orthographical standard was reached only in 1880, after Finnish had already undergone several important phonological transformations (/ð/ > /d/, /θ/ > /ts/, /ɣ/ > /v~Ø/); thirdly, the Latin alphabet fits Finnish phonology. Finnish possesses 13 consonants and 8 vowels, with a ratio of 96 consonants each 100 vowels in speech. Vowels conserve their full value in unstressed syllables and the dominant principle in word formation is to avoid any phoneme that requires a difficult articulation; a syllable never begins with a consonantic cluster and this greatly limits the phonological resources and the number of monosyllabic roots.

The voiced consonants /b, d, g/ do not belong, originally, to the Finnish language, but lately people pronounce them under the influence of foreign loans and of orthography, which employs the graphemes <b, d, g>, triggering a fortition process (but, according to Brown & Koskinen 2011, only the phonological status of /d/ is undisputed, whereas /b/ and /g/ are still pronounced [p] and [k] in everyday life). The trend of the most educated speakers to pronounce <b, d, g> as voiced consonants has provoked a more tense pronunciation of voiceless stops /p, t, k/ to keep the phonemes distinct. Given that many morphological oppositions are indicated by both vocalic and consonantic length, Finnish pronunciation requires a greater accuracy than other European languages (Hakulinen 1961: 5-17).

If we compare the Finnish structure to that of French and English it is clear that:

- Finnish has 21 phonemes, vs. 36 of English and 37 of French;

- Finnish is spoken by 6 million people in Finland and Sweden, whereas 205 million people speak French, in France, Belgium, Switzerland, Canada, Haiti, in many African countries, etc. Even more people speak English (around one billion and 351 million, including those who speak it as second language) and it is the official language in several countries (UK, Ireland, USA, Canada, Australia, New Zealand, etc.);
- the first attempts to write in English and French date back to Middle Age, Finnish literature began in the 19<sup>th</sup> century;
- since they have an orthographic standard, French and English have undergone several phonological processes (especially with regard to vowels in English and consonants in French), whereas Finnish phonology was relatively stable when a graphic standard was established;
- Finnish has few monosyllabic words, whereas French and English have many.

#### 4.5. Italian

Italian, just like Finnish and German, owes its standardisation to the fixation of a written language, created by mixing features of different dialects. Written Italian is a sort of ‘amended Florentine’, namely an orthography based upon Florentine speech without the local features that were too marked (like Tuscan “gorgia”, the fricativisation of voiceless stops in post-vocalic position) and with some Lombard influences.

The stability of Italian orthography is a relatively recent phenomenon and until the 19<sup>th</sup> century there were many possible spellings, especially to indicate the palatal phonemes that did not exist in Latin (such as /ʎ/ and /ɲ/), to distinguish between /ɛ/ and /e/ and between /ɔ/ and /o/, but few attempts were made to indicate /i/ and /u/ differently from /j/ and /w/ (Migliorini 1994: 146-147, 206-207).

Even if Italian is thought to have a very transparent orthography, it is less shallow than others. It maintains some etymological letters, such as <q><sup>11</sup> in words of Latin origin, silent <i> in words like *scienza*, *efficiente*, *deficiente*, *cielo* and <h> in *ho*, *hai*, *ha*, *hanno* ‘I have, you-sing. have, he/she/it has, they have’.<sup>12</sup> The main stress, albeit variable, is indicated by a graphic accent only in polysyllabic oxytones and in some monosyllabic words with a morpholexical distinctive value. The timbre of <e> and <o> is not generally indicated and there are some other inconsistencies. However, Italian spelling is based upon a straightforward relation between writing and speech

and a native speaker generally knows how to write or how to read a word that he or she has never heard before.

Among the Romance languages, Italian is the one which remained closer to Latin phonology (if we do not consider Sardinian), so the alphabet still fits the language. From a strictly formal point of view, Italian orthography is defective, because it does not distinguish between /ɛ/ and /e/, /ɔ/ and /o/, /i/ and /j/, /u/ and /w/, /s/ and /z/, /tʃ/ and /dʒ/, but these distinctions are not relevant for most native speakers. A more phonetic orthography would risk to impose an artificial pronunciation and to discriminate those who do not follow the Tuscan model. Eventually, the choice not to indicate these differences turned out to be functional.

Nevertheless, if in Finland a highly transparent orthography correlates with a high degree of literacy of the population (around 100%), in Italy teachers have pointed out that the average level of orthographic competence of Italian students is decreasing more and more dramatically and there are many cases of adult illiteracy and functional illiteracy (Giscler 2007).

#### *4.6. Opacity vs. shallowness*

In summary, we may assume that there is a high probability that an alphabetic orthography is opaque if:

- there is a very old literary tradition, so that the need to conserve the graphic image of words has come to light and is felt by language users;
- the language is widespread and there are many local varieties (one exception is Danish, which has an opaque orthography but is spoken only by 6 million people<sup>13</sup>); in these cases, opaque orthography permits people to communicate even if their spoken varieties differ greatly, just as a logographic system would do;
- the available graphemes are not sufficient to transcribe all the phonemes of the given language;
- the language has undergone many phonological changes during or after the standardisation of the spelling;
- the language undergoes significant phenomena of vowel reduction. It is often the case of stress-timed languages (such as English, Swedish, Russian),<sup>14</sup> which are less likely to have an univoque written representation of vowels than syllable-timed languages (such as Finnish, Italian, Spanish). This is not always true, though; for example, French is a syllable-timed language but has an opaque orthography, whereas German is

a stress-timed language but it is quite regular in the graphic representation of vowels;

- the given language has many homophonous words.

By contrast, alphabetic orthographies tend to be transparent if:

- the standard orthography was introduced (or revised) in relatively recent times;
- the language has a relatively stable phonological structure in which most vowels are pure and vowel reduction phenomena only happen at a phonetic level but not at a phonemic one;
- the number of phonemes and the number of graphemes do not differ too much;
- the language has not many varieties or it is not very widespread outside its country (some important exceptions are Turkish, Spanish, Italian);
- the language has few cases of homophony;
- the language has accepted an underspecified notation, in which some phonemes are neutralised in writing (e.g. Italian orthography).

### *5. The naturalness of artificiality*

According to Walter Ong (1986: 124), writing is a kind of technology, since it has been invented by human beings and it is not natural and spontaneous like speech, but even if writing is artificial, yet “artificiality is natural for human beings”.<sup>15</sup>

Our assumption here is that, since writing conveys language through graphemes, as much as speech conveys language through sounds and signing conveys language through gestures, then writing, which is itself artificial, starts to be subjected to the same constraints of language and to work as language. Its conservative nature led many scholars to think that orthography can be reformed at will and that it should perfectly reflect speech. But this is not how language works. We as linguists know very well that the different planes of language are very seldom in a 1:1 relationship and that isomorphism is a concept artificially constructed in grammar books, but that does not show in the language used by people every day. So, most of the so-called discrepancies of written language might be merely a result of the adaptation of the written medium to the specific logic of language. Primus (2005: 240) points out that “[o]rthographies have been criticised for mapping spoken language imperfectly. But functional imperfection is a natural trait of language”.

The spontaneous deviations from the norm operated by language users are not really disturbing for the system, it has been the over-zealousness of pedantic purists to affect orthographies in unnatural ways throughout history.

As we have seen, the history of English and French orthographies is normally depicted as a corruption from a shallow to an opaque orthography. Similarly, Gelb and other scholars thought that all the writing systems had to eventually evolve into alphabets, because of the alleged inherent perfection of the alphabetic principle. But linguists should not think like that; when considering writing, they should be able to find what is natural in artificiality, that is, how natural linguistic processes find their way and show themselves through the artificial written medium. If we do that, it is clear that there is no such thing as a linear evolution from picto/logographic to phonemic writing. Each writing system evolved and keeps on evolving but they hardly become purely phonemic.

One of the most ingenious writing system today is the Japanese one, which employs all possible solutions in writing: ideography, logography, syllabography and phonography. Besides or thanks to that, very high rates of literacy are achieved in Japanese society. English orthography works in a very similar way to Japanese one, it is just less apparent since English uses only the Latin alphabet and Japanese employs kanji, hiragana, katakana, furigana (small kana printed next to a kanji to indicate its pronunciation), Arabic numerals and Latin letters.

### *5.1. Natural processes in private written communication*

If the hypothesis that human beings are more at ease with a perfectly phonemic system were true, we would expect spontaneous, privately written communication to work in this way, but if we analyse the very recent evolution of writing habits developed with Short Message Texting and the Internet, we still find the same old “tricks” employed by Ancient Egyptians (e.g. rebus, such as *m8* for ‘mate’) or by the scribes of the Middle Age, e.g. abbreviations such as <q> for *qué* in Spanish or <w<sup>f</sup>, þ<sup>t</sup>, þ<sup>u</sup>> for *with, that, thou* in Middle English (McLaughlin 1963: 44-45).

Some other examples:

- <I love U> or <I ♥ U>; the name of the grapheme <u> /ju/ is homophonous with the pronoun ‘you’ (rebus); in the second case, the verb ‘love’ is replaced by the symbol of a heart (ideography); the use of <♥> has become so common lately that, to

cope with its absence from the keyboard, internet users write <3. Given that all the words in the sequence *I love you* are monomorphemic and monosyllabic, writing <I ♥ U> would, at the same time, satisfy the economy principle and create an isomorphic relation between graphic signs and syllables;

- In Italian, <x> is both a letter and the mathematical symbol for the multiplication called *per*; because of that, <x> can replace both the preposition *per* ‘for’ (logography) and the phonic sequence /per/ (rebus), e.g. *Lo faccio x te* ‘I do it for you’, *xfetto*, *xdo*, *xso* instead of *perfetto*, *perdono*, *perso* ‘perfect, forgiveness, lost’. The homophony between the names of numbers and other words allows to write *for* as <4> in English and to write *sei* ‘you-sing. are’ as <6> in Italian (cf. Mioni 2009: 36-38);
- Very common abbreviations reduce words to a three-grapheme root, a process which reminds closely of Semitic writings such as Hebrew and Arabic scripts: e.g. English *mmt* for *moment*, Spanish *tmb* for *también*; Italian *cmq* for *comunque*, etc. It is interesting to note that very often these triconsonantic abbreviations ignore the postvocalic nasal (*moment*, *también*, *comunque*), a phenomenon very well documented in many Mediterranean writing systems (Miller 1994: 19, Justeson 1976: 76), as well as in children’s early spelling in English, Dutch and Spanish (Read 1986: 80-86);
- Another bit of evidence that expert readers and writers do not produce, spontaneously, phonologically detailed utterances, is the general phenomenon of the abandonment of diacritics. Young French and Czech speakers, whose orthographies make a great use of diacritics, when writing texts with their mobile phone or chatting on the internet, do not use them. A study about Portuguese native speakers communicating on the internet with a non-native keyboard underlined that they tried to mark words (e.g. using an apostrophe instead of an unavailable accent) only when there was the possibility of semantic ambiguity but they had no interest to do so when the use of diacritics was related only to phonology (Jensen 1995). Context itself helps decoding meaning. About this, Nina Catach (1992: 24) defines French diacritics “gênants, lourds à gérer, peu utiles dans bien des cas” and points out that in private their use is already decreasing. This seems a general phenomenon, applying to very distant languages: Coulmas (1989: 237) talks about Vietnamese, where diacritics “[i]n handwritten script (...) are easily and often omitted; and in print they make for a clutter-

tered appearance". This tendency does not appear surprising if one thinks of Hebrew: the only written materials where vowels are normally noted by *matres lectionis* are, besides the Bible, books for children and poetry, but in everyday life, adult readers do not need them (Sampson 1985: 89).

The situation is even more complex: alongside with a return to logography, consonantic writing, etc. we also encounter phonetic spellings, e.g. French *koi* and *jamé* instead of regular *quoi* /kwa/ and *jamais* /zame/ or American English *don't cha* or *whatcha say* for *don't you* and *what (do) you say*.

As we have already pointed out, these strategies are not necessarily related to the new media but are as old as writing. Let us think about cursive handwriting: in private use one does not feel the need to draw the shape of each grapheme in a clear, univocal way; on the contrary, the more expert one is, the faster he/she writes and probably the less clear his/her writing will appear. We can then assist to phenomena of graphemic distinctiveness reduction, in which many graphemes look very similar, if not identical, e.g. <u>, in the handwriting of many people, could stand for <n> or for <u>; <l> normally stands for <l> but in fast handwriting could merge with <t> or <f>; the same applies to <e> and <c> and so on. Something similar happens in East Asia, where

"to write neatly to an educated man could actually be seen as insulting, since it suggested that he was thought incapable of reading cursive forms. Normal handwriting (...) is some way removed from the neatness of print, fusing what are printed as separated dots and strokes into continuous, smooth motions of the brush". (Sampson 1985: 192).

After all, cursive handwriting stand to print (in writing) as allegro forms stand to lento forms (in speech; cf. Dressler 1975); when one speaks fast, in a familiar context, with no or little social pressure, he/she applies a series of phonological reductions but yet, his/her speech is still comprehensible to a native speaker; similarly, graphic reductions, if they are not excessive and are put in the right context, do not affect the intelligibility for the reader.

## *5.2. New writing systems*

In the last centuries, many languages that were only spoken acquired a written form. Normally when a society needs its own script, it adapts or adopts an existing one but from to time to time

there have been some individuals who, once aware of the existence of literacy, invented a writing system from scratch. The creation of a script is called GRAMMATOGENY, an operation that can be SOPHISTICATED if it implies a certain degree of phonetic knowledge, or UNSOPHISTICATED, if the maker of the script cannot read any language and does not know anything about phonetics (Daniels 1996: 579).

Among sophisticated systems it is worth to consider the Pollard script, invented by Samuel Pollard in Southern China around 1887 to transcribe the Western Hmong language. The script is composed of 32 letters corresponding to single consonants or consonantic groups and 37 diacritical letters corresponding to vowels and nasal finals, placed differently (at the top, upper right, middle or bottom) depending on the tone. The shape of the symbols resembles the Latin letters, besides being more geometrical, and the system works almost like hangül, i.e. an alphabet where the graphemes are grouped in syllabic glyphs (Daniels 1996: 580).

It is said that the Pollard script was influenced by the Cree syllabary, invented in 1840 for Cree and Ojibwe languages in Canada and then adapted to Athabaskan and Inuit. Like Pollard script, Cree can be defined as a “featural-cum-abugida” system (Daniels 2001); single vowels are indicated by a triangle shape and the rotation changes according to vowel quality; consonant-initial syllables indicate the consonant by the shape and the vowel by the orientation (e.g. Inuit < Δ ▶ ◀ > stand for /i, u, a/ and < Λ > < > stand for /pi, pu, pa/) (Nichols 1996: 608).

Pollard and Cree scripts seem to work quite well for the languages they convey, but the study of unsophisticated writing systems is more effective in identifying natural tendencies. It appears that writing systems devised independently from each other end up more often being syllabaries rather than alphabets and some of them pass through a logosyllabic stage. Edgerton, in his criticism of Gelb’s theory (Edgerton 1952: 287), points out that, among new scripts, Cherokee (created by Sequoyah around 1810), Vai (designed by Dualu Bukele of Jondu in the 1820s in Liberia), Alaska (devised by Uyaqoq between 1901 and 1905) and Bamum (invented by King Njoya of the Bamum tribe after he had a dream about it) started out as mainly logographic systems but quickly became syllabaries. The first part of Gelb’s theory is then confirmed, but the inevitability of the evolution of these systems into alphabets does not show, since they are still employed as syllabaries, although Alaska and Bamum scripts show “some tendencies toward alphabetisation” (Gelb 1963: 209).

Autochthonous African alphabets, such as Bassa and N’ko, were introduced respectively by Flo Darvin Lewis, a Bassa native speaker

who went to America to study medicine (he was therefore deeply affected by English alphabetic orthography) and by Soulemayne Kante in 1949, as an identitary action against those who considered Africans inferior for not having a script of their own (Dalby 1969: 162, cf. Pasch 2008).

The Neo-Tifinagh script (ⵜⴰⴳⴷⵉⴷⴰⵏⵜ <tifinay>) is an alphabet and is currently used in Morocco and Algeria to transcribe various Berber languages; in its original form it was an abjad but vowels are now noted as well. The passage from consonantic to alphabetic writing might depend on three main factors: the refusal of the Arabic script as an identitary action,<sup>16</sup> the influence of the Latin alphabet and the linguistic structure of Berber languages for which an abjad is not well suited (O'Connor 1996, Pasch 2008).

As suggested by Daniels (1996: 579), “observable script inventions have much to teach about the possible scenarios of the three ancient grammatogenies (Sumerian, Chinese, Maya)”. All these three languages have a similar typology, i.e. morphemes are mostly monosyllabic, a feature which allows a good fit with a logosyllabic system (Daniels 1992: 83).

As we have pointed out at → 3.3, the most intuitive unit for a human being is the word or the concept, so it comes as no surprise that the first attempts of both Sequoyah and Uyaqoq consisted in devising a symbol for every word, and it is even less surprising that eventually they both chose to denote syllables and not phonemes.

Put in other words, logographic writing is very natural but not economic at all (since it contrasts with the criterion of the size of the inventory → 3.2), so the most natural unit right after the word/morpheme is the syllable (cf. Dressler & Dziubalska-Kořaczyk 1994). The existence of brand new alphabets cannot be taken as an evidence for the naturalness of a segmental notation, since the prestige of the Roman script is today too important to ignore and it sure has a weight in modern grammatogenies.

## *6. Conclusion*

The main aim of our article was to give a quick overview about the linguistic fit of the different writing systems, in the attempt to refute the Western axiom according to which the alphabet is the best system whatsoever and shallow orthographies are superior to opaque ones. While doing this, we were also hoping to contribute to the still scarce amount of graphemic studies, namely, a linguistic approach to

the problems of writing. To do so, we have tried to abandon the usual path (a forced parallelism between graphemics and phonology) and we ventured in two domains which, to our knowledge, had not met yet graphemics in a consistent way: Natural Linguistics (see Dressler et al. 1987) and neurolinguistics.

If linguists wish to treat graphemics as a somewhat independent plane of language,<sup>17</sup> they must find natural tendencies in the approach of human beings to reading and writing. The current studies in neurology have demonstrated that specific areas of our brain respond to the sounds of language and that other areas respond to graphemes. If it seems plausible that spoken language has a biological basis and may be largely innate, the same theory seems unlikely when it comes to writing, due to its relatively recent invention. According to Dehaene, it was writing that modelled itself to be adequate to brain circuits, not vice versa (Dehaene 2009: 1-10). Once we realise that the shape of our writing systems is not wholly accidental but that shared similarities between geographically and chronologically distant traditions are not coincidences, but are due to neuronal constraints, we may dare to identify natural tendencies in reading and writing.

If we consider the question without any relationship to speech, the most natural writing system would display characters composed of no more than three strokes (for the writer) but whose configuration is perceptively redundant (for the reader).

When it comes to the naturalness of the encoding process of a specific language, pictography, ideography and logography seem the most natural methods, but they require a huge mnemonic effort (and a great neuronal recycling). Moreover, it appears that, even if in the reading process both the graphemic-phonemic and the graphemic-lexical ways are activated, the most important one remains the former, since “the *universal phonological principle* (UPP) (...) predicts automatic activation of phonological information in words in all languages, that is, prelexical phonological coding, even in deep orthographies such as Chinese” (Perfetti & Zhang 1991, 1995 cited in Chikamatsu 1996: 64). For the human being, the syllable would be a more natural unit than the phoneme, so syllabaries and abugidas would be somewhat superior to alphabets. Nevertheless, alphabets require less mnemonic effort and are more suitable for languages with a complex syllabic structure. After all, each language has its own fit in relation to writing, so any claim about the primacy of the alphabet has no linguistic value, but only an ideological one.

A linguistic analysis of opaque alphabetic orthographies reveals that some alleged inconsistencies may play a positive role (e.g. main-

taining a certain degree of morphological naturalness in languages where phonological processes have heavily affected morphotactic and morphosemantic transparency) and may be a natural evolution caused by the loss of the original linguistic fit of the script (e.g. English, French, Danish, etc.).

Finally, empirical observations of spontaneous written utterances in private communications show that, once a sufficient level of isomorphism with speech is established, all kinds of relation with all planes of language (phonetics, phonology, morphology, etc.) can be fruitfully exploited and that the level of accuracy can decrease, to some extents, without interfering with the communication process.

### *Address of the Author*

Università degli Studi di Padova, Facoltà di Lettere e Filosofia,  
Dipartimento di Studi Linguistici e Letterari, via Beato Pellegrino 1,  
35137, Padova, Italy <antonio.baroni@studenti.unipd.it>

### *Notes*

<sup>1</sup> Actually, it has been proven that logographic systems are indeed LOGOSYLLABARIES or more precisely MORPHOSYLLABARIES, since “each character stands for a morpheme, and the characters can be used for the sound of the morpheme as well as for its meaning”. (Daniels 2001: 43).

<sup>2</sup> Syllables other than CV are rarely represented. In the Linear B, CCV would be written as <CV>+<CV>, e.g. <ko-to-i-na> for [ktoyna], <de-ka-sa-to> for [deksato], etc. (Miller 1994: 18-19). In Old Accadian, CVC sequences are often rendered as <CV> + <VC>. Moreover, the few existing graphemes standing for CVC sequences do not reflect the actual syllable boundaries, e.g. <is-pur-am> for [ispuram], rare variant of <is-pu-ra-am>. Signs for complex rhymes, such as <VCC>, are not attested in any writing system (Dressler & Dziubalska-Kořaczyk 1994: 68-69).

<sup>3</sup> Graphonemes are the object of study of what Mioni (2009: 19) calls *systematic graphemics* (as opposed to *autonomous graphemics*, which deals with the internal structure of writing systems, regardless of their phonetic value).

<sup>4</sup> Many authors other than Daniels made similar proposals, see Gleitman & Rozin (1977), Bellamy (1989), Coulmas (1989), Aronoff (1992), Faber (1992). According to Miller (1990, 1994) phonemes are part of the implicit linguistic knowledge (but not of the explicit linguistic knowledge, see Chomsky 1986).

<sup>5</sup> Probably inspired by Indian grammars made known in China by the Buddhist teaching.

<sup>6</sup> This is evident if we consider Roman numerals: the numbers from one to three are represented by, respectively, one stroke, two strokes and three strokes but to represent ‘four’ and the following numbers, other methods are used: <I, II, III, IV, V, etc.>.

<sup>7</sup> From a diachronic point of view, the graphemes of Latin and Greek alphabets derive, through Phoenician, from Proto-Sinaitic glyphs based on pictograms. If

iconicity is still there in modern writing systems, it is likely to be residual or accidental, but even so, that does not imply that users cannot find a certain degree of motivation in the shape of graphemes, be it related to sound, meaning or the relationship with other graphic elements.

<sup>8</sup> In both languages, the majority of words are monosyllabic, even though Chinese is slowly acquiring more and more polysyllabic words through the creation of polymorphemic compounds.

<sup>9</sup> *Cinq* is pronounced [sɛ̃k] before a vowel and can either be pronounced [sɛ̃] or [sɛ̃k] in the other possible contexts.

<sup>10</sup> Although in liaison an epenthetic [t] appears.

<sup>11</sup> <q> in Italian stands for /k/ but can appear only before <u>. In this position it rivals with <c> and their distribution is based on etymological criteria, cf. *cuore* vs. *quale*; both come from Latin, the former from *cor*, *cordis*, the latter from *qualis*, *qualis*.

<sup>12</sup> Spelling distinguishes <ho, hai, ha, hanno> from their homophones <o, ai, a, anno>, which mean, respectively, 'or', 'to the-pl.', 'to', 'year'.

<sup>13</sup> The opacity of Danish orthography can partially be due to the necessity to conserve the mutual intelligibility between Nordic languages, considering also the central role of Denmark in the Kalmar Union (1397-1523), during which Denmark, Sweden, Norway, Iceland, Greenland and the Faroe Islands were united. Norway and Denmark remained a single political entity until 1814 and even today Bokmål, one of the two official Norwegian written languages (the other one being Nynorsk), differs minimally from written Danish. Denmark lost its rule over Iceland only in 1943, whereas Greenland is still part of the Danish Kingdom.

<sup>14</sup> As Linell (1979: 56) suggests, "careful pronunciations need not be, and are very often not, the most frequent or normal pronunciations of the word forms involved. On the contrary, they will be somewhat artificial and pedantic, particularly perhaps in languages that have heavy stresses and thus normally a great deal of reduction (e.g. English, Danish, Russian). In such cases, it may be that speakers may even construct full-vowel plans which are virtually never realized as such (...). Possibly such abstractness may be due to conventional orthography".

<sup>15</sup> Artificiality pertains to the Peircean sign type SYMBOL, which is more complex than the ICON. As a matter of fact, orthographies tend to be more symbolic and less iconic. Orthographic rules are LEGISIGNS, namely, laws that are signs (cf. Peirce 1980).

<sup>16</sup> The choice of a script is a very strong identitary act for a society. As Sebba (2006: 100) points out, "debates on orthography become symbolic battles over aspects of national, regional or ethnic identity". Let us take the Tatar language as an example. As a Turkic language, Tatar used to be written using the Arabic script, but switched to the Latin alphabet in the 1920s and later to the Cyrillic. In the late 1990s, the Tatarstan government decided to gradually restore the Latin alphabet, but soon after, the Russian government accused Tatarstan to threaten Russian unity by doing so, and in 2002 a bill was approved imposing all national languages of the Russian Federation to use alphabets based on Cyrillic. Consequently, in Tatarstan, the use of Latin alphabet to write Tatar has diminished in public situations, but individual users still employ it in their personal websites or in private communication (Suleymanova 2010: 55-56).

<sup>17</sup> Written language had already been analysed as partially independent from speech by the members of the Prague School, with a functionalist approach (cf. Artymovyč 1932, Vachek 1939, 1973).

*Bibliographical References*

- ALBROW Kenneth Harold 1972. *The English writing system: Notes toward a description*. London: Schools Council.
- ARONOFF Mark 1992. Segmentalism in linguistics: The alphabetic basis of phonological theory. In DOWNING, LIMA & NOONAN 1992. 71-82.
- ARTYMOVYČ Aenor 1932. Fremdwort und Schrift. *Charisteria Guilelmo Mathesio quinquagenario a discipulis et Circuli Linguistici Pragensis sodalibus oblata*. Prague: Pražský Linguistický Kroužek. 114-118.
- BASBØLL Hans 2005. *The Phonology of Danish*. Oxford & New York: Oxford University Press.
- BELLAMY Rufus 1989. Bellerophon's Tablet. *Classical Journal* 84. 289-307.
- BERRY Jack 1977. The 'Making of Alphabets' Revisited. In FISHMAN Joshua Aaron (ed.) 1977. *Advances in the Creation and Revision of Writing Systems*. The Hague & Paris: Mouton. 3-16.
- BOLINGER Dwight Le Merton 1946. Visual Morphemes. *Language* 22.4. 333-340.
- BROWN Jason & Päivi KOSKINEN 2011. On Voiced Stops in Finnish. *Linguistica Uralica* 47.2. 94-102.
- CATACH Nina 1992. À propos de l'accent plat. *Liaisons-HESO* 19-20. 20-25.
- CHANGIZI Mark A. & Shinsuke SHIMOJO 2005. Character Complexity and Redundancy in Writing Systems over Human History. *Proceedings: Biological Sciences* 272. 267-275.
- CHANGIZI Mark A., Qiong ZHANG, Hao YE & Shinsuke SHIMOJO 2006. The structure of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *American Naturalist* 167.5. 117-139.
- CHIANG Yee 1973. *Chinese Calligraphy*. Cambridge & London: Harvard University Press.
- CHIKAMATSU Nobuko 1996. The effects of L1 orthography on L2 word recognition: A study of American and Chinese learners of Japanese. *Studies in Second Language Acquisition* 18. 403-432.
- CHO Jeung-Ryeul & Catherine McBRIDE-CHANG 2005. Levels of Phonological Awareness in Korean and English: A 1-Year Longitudinal Study. *Journal of Educational Psychology* 97.4. 564-571.
- CHOMSKY Noam 1986. *Knowledge of Language: Its nature, origin, and use*. New York: Praeger.
- CHOMSKY Noam & Morris HALLE 1968. *The sound pattern of English*. New York, Evaston & London: Harper & Row Publishers.
- CONSANI Carlo, Cristiano FURASSI, Francesca GUAZZELLI & Carmela PERTA (eds.). *Atti del 9° Congresso dell'Associazione Italiana di Linguistica Applicata – Oralità e Scrittura. In memoria di Giorgio Raimondo Cardona* (February 19-20, 2009). Pescara: Guerra.
- COSSU Giuseppe, Donald SHANKWEILER, Isabelle Y. LIBERMAN, Leonhard E. KATZ & Giuseppe TOLA 1988. Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics* 9. 1-16.
- COULMAS Florian & Konrad EHLICH (eds.) 1983. *Trends in Linguistics, Studies and Monographs – Writing in Focus* 24. The Hague: Mouton.

- COULMAS Florian 1983. Writing and literacy in China. In COULMAS & EHLICH 1983. 239-254.
- COULMAS Florian 1989. *The writing systems of the world*. Oxford: Blackwell.
- COULMAS Florian 2009a. Back to the Future: Literacy and the Art of Writing in the Age of Cyberspace. In CONSANI, FURASSI, GUAZZELLI & PERTA 2009. 93-114.
- COULMAS Florian 2009b. Evaluating Merit – the evolution of writing reconsidered. *Writing Systems Research* 1.1. 5-17.
- CROCCO GALÈAS Grazia 1998. *The parameters of natural morphology*. Padova: Unipress.
- DANIELS Peter T. & William BRIGHT (eds.) 1996. *The world's writing systems*. New York: Oxford University Press.
- DANIELS Peter T. 1992. The syllabic origin of writing and the segmental origin of the alphabet. In DOWNING, LIMA & NOONAN 1992. 83-110.
- DANIELS Peter T. 1996. The Invention of Writing. In DANIELS & BRIGHT 1996.
- DANIELS Peter T. 2001. Writing Systems. In ARONOFF Mark & Janie REES-MILLER (eds.) 2001. *The Handbook of Linguistics*. Malden & Oxford: Blackwell. 43-80.
- DEHAENE Stanislas 2009. *I neuroni della lettura*. Milano: Cortina. (Orig. vers. Paris: Odile Jacob 2007).
- DIRINGER David 1948. *The Alphabet: A Key to the History of Mankind*. New York: Philosophical Library.
- DOWNING Pamela, Susan D. LIMA & Michael NOONAN (eds.) 1992. *The Linguistics of Literacy*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- DRESSLER Wolfgang U. 1975. Approaches to Fast Speech Rules. In DRESSLER Wolfgang U. & MAREŠ František V. 1975 *Phonologica 1972*. Vienna: Wilhelm Fink. 219-234.
- DRESSLER Wolfgang U. 1987. Word Formation as Part of Natural Morphology. In DRESSLER, MAYERHALER, PANAGL & WURZEL 1987. 99-126.
- DRESSLER Wolfgang U. 1990. Sketching Submorphemes within Natural Morphology. In MÉNDEZ DOSUNA Julián & Carmen PENSADO (eds.) 1990. *Naturalists at Kreams*. Salamanca: Ediciones Universidad de Salamanca. 33-41.
- DRESSLER Wolfgang U., Willi MAYERHALER, Oswald PANAGL & Wolfgang U. WURZEL 1987. *Leitmotifs in Natural Morphology*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- DRESSLER Wolfgang U. & Katarzyna DZIUBALSKA-KOLACZYK 1994. Evidence of syllabic writing against the phonological syllable. In DRESSLER Wolfgang U., Martin PRINZHORN & John R. RENNISON (eds.) 1994. *Phonologica 1992*. Torino: Rosenberg & Sellier. 65-76.
- EDGERTON William F. 1952. On the Theory of Writing. *Journal of Near Eastern Studies* 11.4. 287-290.
- ELBRO Carsten 2005. Literacy Acquisition in Danish: A Deep Orthography in Cross-Linguistic Light. In MALATESHA & AARON 2006. 31-45.
- FABER Alice 1992. Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In DOWNING, LIMA & NOONAN 1992. 111-134.

- FOSS Donald J., David A. SWINNEY 1973. On the Psychological Reality of the Phoneme: Perception, Identification and Consciousness. *Journal of Verbal Learning and Verbal Behavior* 2. 246-257.
- FOURNIER P. F. 1940. Sur l'origine des complications de l'orthographe française. *Le français moderne* 8. 257-266.
- GELB Ignace Jay 1963. *A study of writing*. Chicago: University of Chicago Press (1<sup>st</sup> edition 1952).
- GISCCEL 2007. *Educazione linguistica democratica a trent'anni dalle 10 Tesi*. Franco Angeli: Milano.
- GLEASON Henry Allan 1961. *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston (1<sup>st</sup> edition 1955).
- GLEITMAN Lila & Paul ROZIN 1977. The structure of acquisition in reading. In REBER S. Arthur & Don L. SCARBOROUGH (eds.) 1977. *Towards a psychology of reading*, Hillsdale, New Jersey: Lawrence Erlbaum Associates. 1-141.
- GOSWAMI Usha 2005. Orthography, Phonology and Reading Development: A Cross-Linguistic Perspective. In MALATESHA & AARON 2006. 463-480.
- GRAY William S. 1956. *The Teaching of Reading and Writing (Monographs on Fundamental Education, X)*. Paris: UNESCO.
- HAKULINEN Lauri 1961. *The structure and development of the Finnish language*. The Hague: Mouton.
- HOREJŠÍ Vladimír 1971. Formes parlées, formes écrites et systèmes orthographiques des langues. *Folia linguistica* 5. 185-193.
- INNIS Harold 1991. *The Bias of Communication*. Toronto: Toronto University Press.
- JENSEN John B. 1995. Writing Portuguese Electronically: Spontaneous Spelling Reform. *Hispania* 78.4. 837-845.
- JUSTESON John S. 1976. Universals of Language and Universals of Writing. In JUILLAND Alphonse, A. M. DEVINE & Laurence D. STEPHENS (eds.) 1976. *Linguistic Studies offered to Joseph Greenberg* 1. Saratoga: Anna Libri. 57-94.
- KILANI-SCHOCH Marianne 1988. *Introduction à la morphologie naturelle*. Berne, Francfort-s. Main, New York & Paris: Lang.
- LIBERMAN Isabelle Y., Donald SHANKWEILER, William F. FISCHER & Bonnie CARTER 1974. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18. 201-212.
- LINELL Per 1979. *Psychological Reality in Phonology. A Theoretical Study*. Cambridge, London, New York & Melbourne: Cambridge University Press.
- LOGAN Robert K. 2004. *The Alphabet Effect: A Media Ecology Understanding of the Making of Western Civilization*. New York: Hampton.
- MALATESHA JOSHI R. & P. G. AARON 2006. *Handbook of Orthography and Literacy*. Oxford & New York: Routledge.
- MARTLEW Margaret 1983. The development of writing: Communication and Cognition. In COULMAS & EHLICH 1983. 257-276.
- MASICA Colin P. 1993. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- MAYERHALER Willi 1981. *Morphological Naturalness*. Ann Arbor: Karoma.
- MCLAUGHLIN John 1963. *A Graphemic Phonemic Study of a Middle-English Manuscript*. The Hague: Mouton.

- McLUHAN Marshall & Robert K. LOGAN 1977. Alphabet, Mother of Invention. *Etcetera* 34. 373-383.
- McLUHAN Marshall 1964. *Understanding Media: The Extensions of Man*. London: Latimer.
- MIGLIORINI Bruno 1994. *Storia della lingua italiana*. Milano: Bompiani.
- MILLER Gary D. 1990. Homer and Writing: Use and misuse of epigraphic and linguistic evidence. *Classical Journal* 85. 171-179.
- MILLER Gary D. 1994. Ancient Scripts and Phonological Knowledge. *Current Issues in Linguistic Theory* 116. Amsterdam: John Benjamins Publishing Company.
- MIONI Alberto 2009. Conoscenze, memorie e riti della scrittura e della parola – Continuando il viaggio di Giorgio R. Cardona. In CONSANI, FURASSI, GUAZZELLI & PERTA 2009. 13-46.
- MITTON Roger 1996. *English Spelling and the Computer*. London & New York: Longman.
- MORAIS José, LUZ CARY, JÉSUS ALEGRIA & Paul BERTELSON 1979. Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition* 7.4. 323-331.
- NICHOLS John D. 1996. The Cree Syllabary. In DANIELS & BRIGHT 1996. 599-611.
- O'CONNOR Michael 1996. The Berber Scripts. In DANIELS & BRIGHT 1996. 112-119.
- ONG Walter J. 1986. *Oralità e scrittura. Le tecnologie della parola*. Bologna: Il Mulino (Orig. vers. London & New York: Methuen 1982).
- PASCH Helma 2008. Competing scripts: The introduction of the Roman alphabet in Africa. *International Journal of the Sociology of Language* 191. 65-109.
- PEIRCE Charles Sanders 1980. *Semiotica*. Torino: Einaudi (Orig. vers. Cambridge: Harvard University Press 1931-1935).
- PERFETTI Charles A. & Sulan ZHANG 1991. Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory and Cognition* 17.4. 633-643.
- PERFETTI Charles A. & Sulan ZHANG 1995. Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21.1. 24-33.
- PONTECORVO Clotilde 1994. Iconicity in Children's First Written Texts. In SIMONE Raffaele (ed.) 1994. *Current Issues in Linguistic Theory* 110. Amsterdam & Philadelphia: John Benjamins Publishing Company. 277-307.
- PRIMUS Beatrice 2004. A Featural Analysis of the Modern Roman Alphabet. *Written Language & Literacy* 7.2. 235-274.
- READ Charles 1986. *Children's Creative Spelling*. London & Boston: Routledge & Kegan Paul.
- SAMPSON Geoffrey 1985. *Writing systems*. London: Hutchinson.
- SEBBA Mark 2006. Ideology and alphabets in the former USSR. *Language Problems and Language Planning* 30.2. 99-125.
- SEYMOUR Philip H. K., Mikko ARO & Jane M. ERSKINE 2003. Foundation literacy acquisition in European languages. *British Journal of Psychology* 94. 143-174.

- SULEYMANOVA Dilyara 2010. International Language Rights Norms in the Dispute over Latinization Reform in the Republic of Tatarstan. *Caucasian Review of International Affairs* 4.1. 43-56.
- TANAKA Kenji 2003. Columns for complex visual object features in the infero-temporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex* 13.1. 90-99.
- THORNTON Anna M. 2007. *Morfologia*. Roma: Carocci.
- VACHEK Josef 1939. Zum Problem der geschriebenen Sprache. *Travaux du Circle Linguistique de Prague* 8. 94-104.
- VACHEK Josef 1973. *Written language. General problems and problems of English*. The Hague: Mouton.
- VENEZKY Richard L. 1970. *The Structure of English Orthography*. The Hague & Paris: Mouton.
- WURZEL Wolfgang U. 1994. Natural Morphology. In E. ASHER Ronald & John M. Y. SIMPSON (eds.) 1994. *The Encyclopaedia of Language and Linguistics*. Oxford: Pergamon Press. 2590-98.

