S₁: The Native Script Effect

Amalia E. Gnanadesikan

Abstract. This paper claims that the script that a person learns first qualifies as a native script (S_1) in a manner analogous to a native language (L_1). The cognitive pre-eminence of the S_1 results in a *native script effect*, which accounts for various findings in the synchronic study of second-language acquisition and in the diachronic study of script adoption. The native script effect is argued to be an important factor in the historical preference shown for the adoption of pre-existing scripts over the invention of new ones. The claim that S_1 is like L_1 runs counter to the assumptions of linguists of the structuralist and generative traditions, who are agreed in the belief that writing is not language. Language is considered to be cognitively special, the result of a special grammar-learning module. However, writing may be more like primary language than previously believed, and the specialness of language may in fact cause other systems (such as writing) to be analyzed grammatically and entrained into language, with the native script effect being one notable result.

1. Introduction

The fundamental claim of this paper is that literate people have a native script in a way analogous to the way in which they have a native language. That is, the human brain processes a script that is learned early and well in ways that are cognitively similar to how it processes language, with the result that the relationship and interaction between such a first script or scripts (S_1) and a script or scripts learned later (S_2) is similar to the relationship and interaction between a first language or languages (L_1) and language(s) learned later in life (L_2). Furthermore, there are both synchronic and diachronic consequences of the special status of the S1, collectively called the *native script effect*.

If this claim is correct, then the knowledge (implicit and/or explicit) that a literate person acquires of how a script behaves is analogous to the knowledge that speakers have of language. This implies that scripts have grammar, which in turn implies that writing is more like language than

Amalia E. Gnanadesikan   0000-0003-4371-9288
478 Blackshire Rd, Severna Park, MD 21146, USA
E-mail: amalia.gnanadesikan@gmail.com

many linguists have been taught to believe. Yet it is obvious that primary (i.e., spoken or signed) language has a special cognitive and evolutionary status in humankind. Writing does not have that status, but appears to piggyback on primary language to become another modality of language both historically (phylogenetically) and in the acquisition of literacy in the individual (ontogenetically).

To explore this topic, this paper proceeds as follows. Section 2 presents definitions—particularly of the term *script*—which will be essential to the rest of the paper. Section 3 briefly outlines the traditional view of the distinction between writing and language in the structuralist and generative traditions, by which writing is not language and a first script therefore could not have a native status akin to that of a native language. Section 4 considers a number of anecdotal and experiential lines of evidence that suggest that native scripts do in fact exist. Section 5 briefly presents results in the existing literature that argue that the differences and interactions between a first script and later-learned scripts are analogous to those between a first language and later-learned languages. Section 6 applies the concept of the native script effect to the history of writing systems, arguing that the cognitive effect of S_1 accounts for the relative rarity of script invention and radical adaptation when previously unwritten languages come to be written. Section 7 returns to the differences and similarities between primary language and writing, conceding that primary language has a special cognitive status but arguing that the specialness of language in the human brain leads to other complex systems, such as writing, becoming entrained in the linguistic system, with the result that writing becomes language. Section 8 concludes with suggestions for the application of the concept of the native script in policy, pedagogy, and linguistic theory.

2.   Definitions

Before proceeding with the central argument, a few definitions are called for. By *script* I mean a somewhat abstract "set of graphic signs with prototypical forms and prototypical linguistic functions" (Weingarten, 2011, p. 16). A *writing system*, by contrast, is the combination of a specific instantiation of a script with the orthographic rules of a specific language. This use of *script* is in contrast with definitions in which *script* is either synonymous with *writing system*, and thus composed of the combination of a signary and an orthography (e.g., Daniels and Bright, 1996, pp. xliv–xlv), or is merely the collection of signs (the signary) used in a writing system (e.g., Daniels, 2018, p. 155). By the definition used here, the script used in any given written language is *more* than just the signary (since it includes some information about the linguistic function of the signs) but *less* than the writing system (since it does not include all the details of a language-specific orthography). Thus English, Italian, and German all

use the Roman script, but they do not share a single writing system. Similarly, Hindi, Marathi, and Nepali all use the Devanagari script, but again they do not share precisely the same writing system.

It is important for work of the present sort to use a term that allows for the existence of a conceptual entity that is shared across languages. Using the fine-grained level of individual writing system, there is no way to tease apart the process of learning a second language (in a literate context) from that of learning a second writing system: every instance of second language learning in a literate context is an instance of second writing-system learning. However, if we look at the level of script, then it becomes clear that learning some second languages requires learning a new script while learning others doesn't. The two processes can be differentiated.

Looking just at the level of the signary is also the wrong level. For example, the writing systems of English, German, and Italian, or of Hindi, Marathi, and Nepali, share much more than the same basic set of signs. They share important typological features and have many shared or similar values in their linguistic interpretation. Thus, for example, English, German, and Italian use alphabetic letters that write both consonants and vowels, while Hindi, Marathi, and Nepali all use an unwritten "inherent" vowel. In English, German, and Italian, <A>, <E>, <I>, <O> and <U> stand for vowels, and <B> stands for a labial consonant. In Hindi, Marathi, and Nepali, <आ>, <इ>, <उ>, and <ए> stand for vowels, while <ब> stands for a labial consonant. By considering the level of script, we are considering not only a set of largely shared symbols but significant shared ways in how those symbols are used.

In order to study how a learner processes a truly new way of writing, therefore, we must look largely at the level of script. An L$_2$ may or may not share L$_1$'s script. Granted, when people learn to read and write they learn these skills within the context of a particular writing system, not merely at the abstract level of script. In this sense a writing system is analogous to a dialect (or language variety) in that each person learns a specific dialect of a language, while the dialects together comprise a more abstract entity known as a language. Similarly, in becoming literate a person learns a specific writing system, and many writing systems may share the same script.

It is also worth noting that there is no claim being made here that monolingualism and monoliteracy are the only options for L$_1$ and S$_1$, or are even normative. In this paper any set of scripts learned well at roughly the same time in childhood are considered collectively as S$_1$, just as any set of languages learned well in early childhood are considered L$_1$.[1]

---

1. I leave aside for now the question of how first-script literacy that is gained in adulthood might differ from that acquired in childhood. If the analogy with primary

3.   Traditional Assumptions About Language and Writing

Linguists of the American structuralist and generative schools have traditionally held dogmatically to the belief that, as Leonard Bloomfield famously put it, "Writing is not language, but merely a way of recording language by means of visible marks" (Bloomfield, 1933, p. 21). The same attitude was recorded rather colorfully by Fred Householder when he listed first among "the propositions intuitively felt to be basic by friend and enemy alike" among Bloomfieldian linguists the proposition that "Language is basically speech, and writing is of no theoretical interest" (Householder, 1969, p. 886). This attitude was inherited by the generative school of linguistics and has continued into the twenty-first century, resulting in a dampening effect on efforts to apply linguistic analysis to writing systems. James Myers, for example, describes having abstracts rejected at linguistics conferences with dismissive comments such as, "This paper does not deal with linguistic matters" (Myers, 2019, p. x). I have myself been told after giving a talk on writing systems to a linguistics department that "this is not interesting," on the grounds that writing, not being language, is not about the fundamental character of the human brain.

Indeed, there are important differences between spoken or signed language—which I will collectively call *primary language*—and writing. These differences can be found laid out in any typical introductory linguistics textbook in the generative tradition. For example:

> Speaking and writing are different in both origin and practice. Our ability to use language is as old as humankind, and reflects biological and cognitive modification that has occurred in the evolutionary history of our species. *Writing*… is a comparatively recent cultural development, having occurred within the past five thousand years and only in certain parts of the world. The contrast between speech and writing comes into sharper focus when we consider that spoken language is acquired without specific formal instruction, whereas writing must be taught and learned through deliberate effort. There are entire groups of people in the world today, as well as individuals in every literate society, who are unable to write. While spoken language comes naturally to human beings, writing does not. (Dobrovolsky and O'Grady, 1997, p. 553, emphasis in original)

One reason that writing and primary language are considered to be fundamentally different is that primary language is considered to be

---

language holds in this respect, there will be significant differences between the two, since failure to learn a primary language in childhood leaves a person with a permanent language deficit (Pinker, 1994). While the initial acquisition of literacy in adulthood is possible, the acquisition of fluent reading is difficult for adults and relapse into illiteracy is common (Abadzi, 1994). Thus the analogy with primary language may indeed hold. However, the effects of age on first literacy acquisition are not yet well understood (ibid.).

special, cognitively speaking, and thus necessarily distinct from other human behaviors, including writing. As mentioned in the quote above, primary language is found everywhere that humans are found. It is therefore claimed (with good reason) that language has a special cognitive place in the human mind—and only in the human mind. Noam Chomsky has long championed

> the Cartesian view that man alone is more than mere automatism, and that it is the possession of true language that is the primary indicator of this... (Chomsky, 1964, p. 8)

Or, as Chomsky has more recently put it,

> There is no serious reason today to challenge the Cartesian view that the ability to use linguistic signs to express freely-formed thoughts marks 'the true distinction between man and animal' or machine... (Chomsky, 2000, p. 3)

In the generative framework, the human faculty for language is considered to arise from a "language acquisition device" (Chomsky, 1965, pp. 32–33). This faculty for language has been termed a "language instinct," which is active during the critical period (mostly strongly from birth to the age of six or so, and phasing out by puberty), during which L_1 learning takes place automatically and implicitly, without explicit instruction (Pinker, 1994).

By this view, language is cognitively special, but it is specifically the native language that expresses the full range of this cognitive distinctiveness. Thus L_1 and L_2 learning are fundamentally different. L_1 learning, assuming it occurs during the critical period (as might fail to happen to a deaf child of hearing parents, or a child raised under circumstances of unusual social deprivation) is fast, automatic, implicit, and more or less perfect. L_2 learning, by contrast, is slow, difficult, and error-ridden, and it leaves the learner with a permanent foreign accent. Furthermore, properties of the L_1 will influence a person's ability to perceive and/or learn features of L_2, resulting in both positive and negative transfer from L_1 to L_2 (Ringbom, 1987). In other words, features of the L_2 that are similar to L_1 will be learned easily (positive transfer), while features of L_1 may persist in a learner's use of L_2 even when they are not appropriate to that language (negative transfer).

If language is indeed cognitively special and writing is indeed not language, then the relationship between a first-learned script and a later-learned script should not resemble the relationship between L_1 and L_2. The following sections set out to examine to what extent the difference between S_1 and S_2 does in fact resemble the difference between L_1 and L_2. If the differences between the two pairs are similar, that suggests that S_1 and L_1 may be more similar than the structuralist/generative view would allow. And while that does not undermine

the claim that language has a special cognitive status, it suggests that the special cognitive status of language spills over onto writing in the development of literacy. This point will be returned to in Section 7.

4.   Impressionistic Evidence for a Native Script

Evidence for the phenomenon of native scripts comes from a number of sources, some admittedly impressionistic and others more rigorous. I begin with the impressionistic. While the interpretations of these impressions may be subjective, I suspect that many adult learners of second scripts will be able to relate to them.

First, fluency in a new script comes frustratingly slowly. An example from personal experience is shown in Figure 1. The non-Roman script on the left is Thaana, the script in which the Maldivian language, Dhivehi, is written (Gnanadesikan, 2012). The Roman-script text on the right is the same text in the official Romanization of Thaana. My personal experience shows that an adult can learn the Thaana script with a day's concentrated effort but that fluency (in the script *as distinct from* the language) takes years. This means, for example, that as an S_2 reader I must *choose* to read a text in Thaana rather than having the reading happen automatically just because my eye landed on it. It makes the text on the right substantially more appealing to me, drawing my eye even against my will. It means, further, that skimming Thaana is difficult to impossible for me, and that I can't read Thaana text upside down, although I have observed S_1 readers do so easily. It means that I cannot automatically (and even involuntarily) pick my own name out of a text as I can with a Roman-script text (as in the transliteration at right), even with some variation in the spelling. Automaticity and fluency do develop over time, but very slowly.
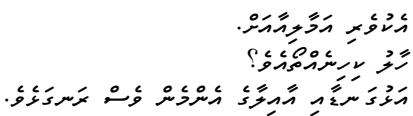
އެކުވެރި އަމާލިއާއަށް.   Ekuveri Amaaliaa-ah.
ހާލު ކިހިނެއްތޯއެވެ؟   Haalu kihinehthoa-eve?
އަޅުގަނޑާއި އާއިލާގެ އެންމެން ވެސް ރަނގަޅެވެ.   Alhugan'daai aailaage emmen ves ran'galheve.

FIGURE 1. A short Dhivehi text in the author's S_2 (left) and S_1 (right). (The text reads, 'Dear Amalia. How are you? I and everyone in the family are fine.')

Another example in shown in Figure 2. In this example, in which a short Chinese text is presented for the learning reader in Hànzì (characters) and Pīnyīn (Romanization), the eye of an S_1 Roman-script reader will be drawn to the Pīnyīn, just as it is to the Romanization in Figure 1, despite the fact that the Hànzì characters are larger. The additional point in this example is that the characters, being morphographic,

contain more information than the Pīnyīn does. Each Hànzì character uniquely identifies a morpheme, while each Pīnyīn syllable could in principle refer to any of several homophonous morphemes (although in context the ambiguities are largely resolved, at least for fluent speakers). Thus both 有 'have' and 友 'friend' are <yǒu> in Pinyin, for example. If one does not know all of the characters, resorting to the Pīnyīn is expected. What can be frustrating to S₂ readers, however, is that even when they *do* know each character, the Pīnyīn will still ineluctably draw the eye, depriving them of authentic S₂ reading practice.

Pīnyīn	Wǒ	yǒu	yī	zhǐ	xiǎo	hēi	māo.
Hànzì	我	有	一	只	小	黑	猫。
Pīnyīn	Tā	de	míng	zì	jiào	wū	lóng.
Hànzì	她	的	名	字	叫	乌	龙。
Pīnyīn	Tā	kàn	qǐ	lái	jiù	xiàng	yī
Hànzì	她	看	起	来	就	像	一
Pīnyīn	zhǐ	xiǎo	hēi	bào.			
Hànzì	只	小	黑	豹。			

FIGURE 2. A short Chinese text in Hànzì and Pīnyīn. The eye of a Roman S₁ reader will be drawn to the Pīnyīn, despite the Hànzì being larger and more informative. Example courtesy of Gitanjali Gnanadesikan. (The text reads, 'I have a black kitten. Her name is Oolong [Black Dragon]. She looks just like a small black panther.')

Another line of evidence comes from the reactions of S₁ readers to instances of script mimicry. *Script mimicry* is the use of graphs from one script (or graphs that look like they come from a particular script) as graphs in another script.[2] A simple example is Devanagari ॐ, spied on a yoga T-shirt. The message intended for Roman-script readers who are not readers of Devanagari is <om>, the sacred syllable of South Asian religions. A reader of Devanagari, however, will read this as <ṭʰal>, its actual value in Devanagari. A more extensive case is shown in Figure 3. While this text is written in English in Roman script, it mimics Japanese

---

2. Alessandrini (1979) uses the term *exotype* to refer to a typeface that, while writing Roman script, is clearly influenced in its letter forms by another script. The font in Figure 3 is an exotype. The term *script mimicry* is related but encompasses a wider range of cases, including ones that use only actual graphs from another script (as in the Devanagari ॐ above), ones that operate between two non-Roman scripts, and ones that occur in handwriting.

katakana and kanji, with the result that while an S_1 reader of Roman script can read it after a moment or two's adjustment, an S_1 reader of Japanese—according to anecdotal evidence—will often fail to be able to read it (Raymond Larabie, personal communication). The S_1 exerts too strong a pull to allow for easy decoding as Roman script.

AS WE PASSED DOWN QUEEN STREET TOWARD CITY HALL, WE WERE STOPPED SO MANY TIMES BY TOTAL STRANGERS, ASKING ANXIOUSLY ABOUT OUR CUTE LITTLE ROBOT PAL.

FIGURE 3.   Script mimicry by Roman script of Japanese katakana and kanji in Electroharmonix font.   Example from `https://typodermicfonts.com/electroharmonix/`, used with permission of Raymond Larabie, the font's designer.

Yet another line of evidence comes from the length of time it takes to learn a language that is written in S_2. Programs of study and measures of success in learning vary greatly, making comparisons difficult in second-language learning. However, a certain degree of standardization can be assumed by considering the courses offered by the US Foreign Service Institute (FSI), since the types of use to which the languages are put and the level of proficiency desired for those uses will be comparable across languages. According to FSI's website[3], languages offered there are divided into four levels of difficulty for (English-speaking) American learners. The languages are tabulated in Table 1.

As shown in Table 1, there are no languages at difficulty level 1 or 2 which have a non-Roman script. In other words, in this sample, rapid language learning (in 36 weeks or less) for S_1 readers of English never involves learning a new script. At level 3, where adequate language proficiency may be achieved after 44 weeks, 28 of the 48 languages use a non-Roman script. In calculating this figure, it was noted that five of the level 3 languages are written in more than one script. In the absence of access to the FSI curricula for these languages, the script that is associated with the language's use as an official national language or its

_____

3. Department of State, "Foreign Language Training: Foreign Service Institute," `https://www.state.gov/foreign-language-training/`.

TABLE 1.   Difficulty level and length of time allotted for English speakers to learn non-English languages at the US Foreign Service Institute. (Language counts and difficulty level tabulated from `https://www.state.gov/foreign-language-training/`.)

Difficulty Level	Total Languages	Non-Roman Script
1 (24−30 weeks)	9	0
2 (36 weeks)	5	0
3 (44 weeks)	48	28
4 (88 weeks)	5	5

likely use for diplomatic purposes was counted.[4] At level 4, that of the "super-hard languages" requiring 88 weeks of training, none of the five languages uses the Roman script.

Granted, there is a clear confound here with the degree of relationship between the language itself and English. It is no surprise to find Dutch in level 1, for example. And in fact, all of the level 1 languages (Dutch, Danish, French, Italian, Norwegian, Portuguese, Romanian, Spanish, and Swedish) are Germanic or Romance languages. However, level 2 includes Swahili (a Bantu language) and Malay and Indonesian (Austronesian languages) beside German and Haitian Creole (a French-based creole). While Swahili, Malay, and Indonesian bear little resemblance to English, they are at least written in the Roman alphabet, sparing the learner the effort of acquiring an S_2.

A final line of suggestive evidence comes from the history of the Cherokee syllabary, famously invented in the early nineteenth century by Sequoyah. When the Cherokee syllabary was first disseminated in the 1820s, "Cherokee children who took up to four years to read and write English reportedly learned the syllabary in a few days and put it to use"; yet by the early 2000s the syllabary was "considered by many native speakers to be an extremely difficult writing system to learn and use" (Bender, 2002, p. 28). Evidently, a significant change in perceived difficulty took place between the early years of the syllabary's use and the present century. The most plausible cause of this difference was the introduction of universal English-language education. Nowadays Roman script is S_1 for Cherokee children. Not only does this mean that the Cherokee script is, by contrast, S_2, but some of the same sorts of confusion as those caused by deliberate script mimicry are at play, since many

---

4. For example, Azerbaijani (or Azeri) is written in the Roman script in Azerbaijan and in Perso-Arabic script in Iran. Since Azerbaijani is the national language of Azerbaijan, an independent nation to which a US diplomatic mission is posted, but is not the official language of Iran, Azerbaijani is considered for the purposes of Table 1 to be written in Roman script.

Cherokee signs resemble Roman letters. For example, Cherokee <W> is /la/ and <D> is /a/.

5.   Synchronic Consequences of a Native Script

Once the possibility of a native script is allowed, a number of results in the existing literature can be interpreted as consequences for the special status of S_1 and its primacy over S_2, analogous to the primacy of L_1 over L_2. L_1 learning is characterized as fast, implicit, and complete, while L_2 learning is slow, often mediated by explicit instruction, and incomplete (leaving an accent, and affected by both positive and negative transfer from L_1). The relationship between S_1 and S_2 is surprisingly similar. This section lists briefly a few works that make this point.

First, acquisition of adequate fluency in S_2 is painfully slow, as already mentioned to in Section 4. As Elliott (2012) puts it, "Inefficient decoding can quickly lead to frustration and diminishing motivation, in turn resulting in less reading practice/time on task" (ibid., p. 66). Elliott suggests that learners may need practice with simplified texts, as authentic texts may well be too difficult.

Secondly, there is evidence for an analog to a foreign accent in handwriting. Certain hand motions are more or less characteristic of one script as compared to another, particularly if the two scripts run in opposite direction. Machine learning experiments have succeeded at distinguishing between S_1 and S_2 writers of Arabic script with 100% accuracy at the document level (Farooq, Lorigo, and Govindaraju, 2006), and between S_1 and S_2 writers of Roman-script English (where the S_2 writers have various scripts native to India as S_1) with up to 97.67% accuracy (Ramaiah, Utkarsh, and Venu, 2012).[5] Furthermore, efforts to identify the specific accent (i.e., the specific S_1, Chinese Hànzì or Devanagari) of S_2 writers of English with machine learning have achieved up to 89.19% accuracy (Ramaiah, Arti, and Venu, 2013).

Thirdly, scripts are sensitive to transfer from S_1 to S_2. An extensive body of research reviewed by Bassetti (2013) shows that literacy skills transfer to a new writing system, but that such a new writing system is more easily learned if the new writing system is typologically similar

---

5. A potential confound that the authors do not discuss is that Roman script as written in different parts of the world (in this case India and the United States) may have different regional "accents," separately from any effect of whether they are a person's first or second script. Thus even an S_1 writer of Roman script schooled India may write detectably differently than an S_1 writer of Roman script schooled in the United States. However, this possibility does not negate the existence of accent in handwriting; it merely adds to the kinds of accents that one should expect. As such it strengthens the analogy with spoken accents, which may be either regional or foreign.

to the previously learned one, allowing positive transfer.[6] As in primary language learning, where bilingualism is an advantage in learning a third (or subsequent) language, biliteracy is an advantage in learning a third writing system. Negative transfer occurs when readers read a word incorrectly, assigning values that would be correct in their native writing system. While the examples Bassetti cites occur within a script (e.g., English speakers reading Spanish <v> as /v/ rather than /b/), my own experience with learning Thaana (shown above in Figure 1) included frustratingly many misreadings of 𐑝 as /v/ rather than the correct /k/.

An S$_2$ may be read with different neural processing patterns depending on the S$_1$, showing that the transfer from S$_1$ to S$_2$ happens at a neurological level. For example, Kim, Liu, and Cao (2017) found that Chinese S$_1$ and Korean S$_1$ readers showed different brain activation when reading English, the Korean S$_1$ readers showing more activation in the right inferior frontal gyrus than the Chinese S$_1$ readers. This was attributed to the fact that the Korean writing system encodes phonemes but the Chinese writing system does not. Chinese S$_1$ readers showed more activation in the left middle frontal gyrus, an area which is particularly active in S$_1$ Chinese reading.

Despite the commonalities between S$_1$ and L$_1$ described in the preceding few paragraphs, the obvious failure of the parallel between S$_1$ and L$_1$ is that S$_1$ is explicitly taught, as mentioned in Section 3. Children are taught to read and write but learn to speak and understand their L$_1$ automatically, without explicit instruction. Nevertheless, there is evidence that some learning of a writing system is implicit. For example, Pacton, Perruchet, Fayol, and Cleeremans (2001) report on an experiment in which children learning to read and write in French showed sensitivity to aspects of French orthography that they are never taught. Specifically, they learned implicitly that French vowel letters are never doubled and that only certain consonant letters are.

Additional evidence for implicit learning comes from Tsai and Nunes (2003), who present evidence that children learning Chinese Hànzì (characters) in Taiwan, where character structure is not explicitly taught, nevertheless internalize the schemas of character composition and become increasingly adept at judging whether a novel character conforms to the schemas between five and nine years of age.

To summarize this section, not only does S$_2$ involve greater difficulty, a foreign "accent," and other types of transfer from S$_1$, but the S$_1$ is to

---

6. Bassetti (2013) discusses biliteracy at the level of the writing system (more specific than that of script), so I have used that wording here. Any difference of script implies a difference of writing system. Not all differences of writing system involve a difference of script, but just as one speaks only a specific variety of one's native language as L$_1$, the S$_1$ will be instantiated in a specific writing system, so that similar but weaker S$_1$ effects should be expected across writing systems that share a script. See Section 6 for more on within-script S$_1$ effects.

some extent learned implicitly, strengthening the analogy with L_1. Admittedly, it could be argued that the difficulties associated with switching from S_1 to S_2 are merely the same sorts of difficulties associated with overcoming any ingrained habit, such as driving on the right-hand or the left-hand side of the road. However, the same argument could then be made for primary language being simply a habit, since the difficulties of switching from L_1 to L_2 are analogous to those of switching from S_1 to S_2. The similarities in the relationship between S_1 and S_2 to the relationship between L_1 and L_2 suggests that script and primary language are in the same boat, whether because of the operation of a special language instinct or merely habit.

6.   Diachronic Consequences of a Native Script

This section turns to the diachronic consequences of literate people having a native script, an application not made elsewhere in the literature, to my knowledge. I claim here that the native script effect is the answer to a question that is not often asked but deserves to be, namely, why are there so few scripts in the world? This is not a question about the number of languages that are written as compared to the number of languages that are not written. Rather, it is a question about why so many languages share a script, despite large differences in their phonological and morphological characteristics that would suggest that different scripts would be more appropriate for them. While some scripts (such as Thaana) are indeed confined to a single language, other scripts have come to be used for many languages. In fact, a few blockbuster scripts, such as Roman, Cyrillic, and Arabic, dominate the world. Why is this the case? Why is innovation so rare in the history of script design?[7]

When a language first comes to be written, there are in theory three ways in which the pairing of a language and script could come about. The first, the independent invention of writing, characterized the first scripts of their respective cultural spheres (such as Sumerian cuneiform or the oldest Chinese writing). In such a case people who have no prior knowledge of writing invent a way to write. The second way, script invention by stimulus diffusion, starts from the background knowledge that writing exists but is not beholden to a prior script for its design features. A famous example of this type is the Cherokee syllabary mentioned in Section 4, since Sequoyah was aware of the existence of writing but was not literate before he invented the Cherokee syllabary. The

---

7. I have elsewhere commented on such lack of innovation in the history of writing by calling the alphabet "a monument to … hidebound conservatism" (Gnanadesikan, 2009, p. 143).

third way, script adoption, is the use of a pre-existing script for the newly written language.

In practice, the first type (creation *de novo*) no longer occurs, since some knowledge of the existence of writing has spread to every, or virtually every, part of the globe. Also in practice, there is something of a spectrum between the second and third types. In other words, there is a spectrum between the invention of a completely new script and the wholesale adoption of a pre-existing script, with some scripts becoming substantially adapted in the transfer to a new language. For example, the Roman alphabet arose from the Greek alphabet and is very similar to it but different enough to qualify as a different script. Nevertheless, the question remains: Why is the end of the spectrum nearer to outright adoption as common as it is? Why don't large typological differences between languages more often lead to large differences in script?

Examples of scripts being borrowed more or less wholesale abound. A few examples (taken from Gnanadesikan, 2009) will suffice here. Chinese characters (Hànzì) were historically adopted to write Vietnamese, Korean, and Japanese, none of which are Sino-Tibetan languages, and two of which (Korean and Japanese) are morphologically synthetic as opposed to Chinese, which is morphologically analytic. The Aramaic script spread from Syria to Manchuria over the course of about two and a half millennia. In the process it spread from Semitic languages to Indo-European languages to Turkic, Mongolic, and finally Tungus languages. The letter forms were quite different by the time they came to be used for Manchu—and the direction of writing had rotated by ninety degrees—but at each step along the way the changes were relatively minor. More recently, the Cyrillic alphabet has come to be used for many minority languages of Russia and the former Soviet Union. Cyrillic as used for the Slavic language Russian has 33 letters (of which 21 are consonants), yet it has been adapted to write the Northwest Caucasian languages Abkhaz and Karbardian, each with about 50 consonants. Similarly, the Roman alphabet, with 21 consonants and 5 vowels, has come to be used for languages as diverse as Vietnamese, an Austroasiatic language with 11 vowels and 6 tones, and Xhosa, a Bantu language with 12 clicks and 43 other consonants (Baker, 1997). Simple metrics of the fit of the script to the phonology of the languages would surely suggest that these sorts of script adoptions would be dispreferred.

While in some cases extra letters or diacritics may be added, as in Vietnamese, in many cases digraphs (and even trigraphs) are called on to stretch the script to fit the language, as in Xhosa. Going so far as to alter the inventory of letters is rare, however. Baker notes that "Strong objections to the very idea of using special characters in orthography design are sometimes held by otherwise rational people, and seem to stem from a deep-rooted conviction that the Roman alphabet is somehow inviolable" (ibid., p. 137).

A similar pattern of preference for an existing writing system can be seen not just in the choice of script as a whole but also within a script in the choice of specific orthographies. Grenoble and Whaley (2006) discuss several cases where orthography designs for minority languages have failed or succeeded depending on how similar they were to the orthographies of the majority languages with which the speakers of the minority languages were familiar. Thus two orthographies designed for Coreguaje (a Tucanoan language) failed because they were not enough like Spanish. An orthography modeled after French designed for Athapaskan languages failed because most of the speakers were familiar with English orthography. On the other hand, the orthography for Zapotec (an Oto-Manguean language), based on Spanish orthography, has succeeded despite a poor match with the Zapotec phonology, since Spanish is the language of education in the Zapotec area.

There are many reasons for the spread of a script. The Arabic script, for example, spread along with Islam as the script of the Holy Qurʾān (Kaye, 1996). However, the existence of a native script effect suggests that at least some of the reason for the frequency of script spread as compared to the rarity of script invention lies in cognitive factors.

The situation is shown schematically in Figure 4. If a native speaker of an unwritten language (that person's L_1) receives an education, it will be in the regional written language of education (that person's L_2). The speaker therefore learns to read in the script of the L_2. The upshot of this situation is that *the speaker's S_1 is the script associated with the speaker's L_2*. This kind of situation is extremely common historically, from the days of Akkadian students learning Sumerian cuneiform to minority children learning majority languages across the world today.
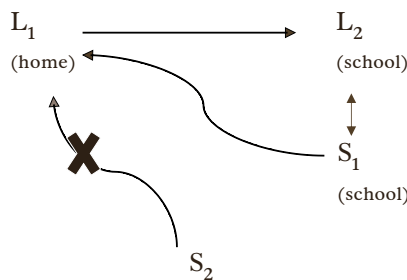

FIGURE 4. A schematic showing how the script of a language learned for educational purposes becomes S_1 and will therefore tend to be adopted for a previously unwritten L_1.

Once S_1 is established as the native script, if the speakers of L_1 want to write their language, there will be a strong predilection for using S_1

(L_2's script). While a different script might be invented or adopted, it will be at a cognitive disadvantage, since it will be competing with S_1.

The natural consequence, therefore, of literate people having a native script is that existing scripts spread. The trend is as old as the adaptation of cuneiform to write Akkadian in the third millennium BCE. And the more powerful a script is, the more it will continue to spread. Educators, missionaries, and policy makers who are not native speakers of the L_1 of Figure 4 but are often native readers of S_1 also play a role, since they too are cognitively biased in favor of their S_1. Their role in the history of script adoption is perhaps more expected, however, as yet another case of domination and/or imperialism by cultural elites. My claim here, however, is that cognitive factors influence all players—including the speakers of the previously unwritten language themselves—toward the adoption of a previously existing, commonly known script, and against script invention. While the invention of new scripts by previously literate individuals for their native languages does happen, as in the case of King Sejong's invention of Han'gŭl for Korean (Kim, 2005) or the invention of Thaana for Dhivehi (Gnanadesikan, 2012), it is relatively rare.[8]

7.   Is Writing Language?

If there really is a native script effect similar to the native language effect, then writing and primary language have significant properties in common, which implies writing cannot be merely dismissed as irrelevant to language, as Bloomfield so famously did. But then what actually is the relationship between the two? Is writing language or not?

While many of the special properties of a native language are also found in a native script (including even some implicit learning), it is also clear that writing and primary language are different in important neurocognitive respects. As mentioned earlier in Section 3, primary language is a universal of human societies, while writing is a later and spottily adopted invention. Additionally, different types of writing systems are processed differently in the brain, a fact that allows for the detection

---

8. I suspect that these cases are examples of the *biliterate advantage* (Bassetti, 2013), by which readers who already know two writing systems are advantaged in learning (or in this case, designing) a third. King Sejong knew Hànzì script and is believed by some to have been inspired by 'Phags pa (Ledyard, 1966). Whether or not he knew 'Phags pa specifically, he would have been well positioned to learn other scripts, as his school for diplomats offered classes in several foreign languages (Ledyard, 1997). The inventor of Thaana clearly knew both Arabic script and an older indigenous writing system, as features of both are incorporated in the design of Thaana (Gnanadesikan, 2012).

of neurological transfer from S_1 to S_2 of the type documented by Kim, Liu, and Cao (2017) and discussed in Section 5.

What we have, then, is a system that acts very much like primary language in some ways but is clearly distinct from it in certain other important ways. How essential are these differences?

Primary language, whether spoken or signed, is believed to be unique because it uses an inborn, implicitly acting grammar-learning module. It is a rule-based (grammatical) system. Yet writing, though not inborn, can also be described as a grammatical system. Myers (2019), for example, analyzes the Chinese script as having a grammar—that is, as following rules of wellformedness—both in its formal properties and in users' processing of it. On another tack altogether, the stroke order of writing letters in both English and Hebrew has been successfully modelled using Optimality Theory, a theoretical paradigm developed for and primarily used to model phonological grammar (Ellenblum, 2019).

One way to resolve the tension between the similarities and dissimilarities between primary language and writing—between the innateness of only primary language on the one hand and the grammar-based properties of both primary language and writing on the other—is suggested by James Myers when he states that "Once this flexible neural system [of language/grammar] evolved, it may have become as trigger-happy as our face-processing system (which detects 'faces' anywhere, even in clouds), automatically switching on whenever it encounters any sufficiently complex communication challenge" (Myers, 2019, p. 22). In other words, the grammar-building language instinct is so strong that it entrains other communicative systems into its orbit. If this is so, then it is no surprise that scripts show grammatical properties and other similarities to primary language.

The view that emerges here is that language is indeed cognitively special but that this specialness lies not so much in being unique but in being overpowering. That is, the language module(s) of the brain will process as language—as grammatically constituted—as many systems as it can. For a literate individual, that includes writing.

The upshot is that while writing does not start out as language, it *becomes* language. This is true both phylogenetically (in the origins of writing) and ontogenetically (in the acquisition of writing by an individual). Historically, writing was not invented to be language. It was not even invented to *record* language but rather to record certain types of information. "[E]arly writing did not reflect spoken language, nor was it invented to do so." (Woods, 2010, p. 20). The world's earliest writing systems, in Egypt and Mesopotamia, took half a millennium or so before they "achieved a relatively full notation of language, including its grammar" (Baines, 2004, p. 150). Yet today the recording of language is considered by many scholars of writing systems to be essential to the

definition of writing (e.g., Gelb, 1963, p. 13; Rogers, 2005, p. 2; Daniels, 2018, p. 157).

Similarly, in the life of the individual it is the primary (spoken or signed) language that is learned with the full drive of the language instinct. The written language requires some explicit instruction. But, as noted above, children learning to read also learn aspects of their writing system implicitly, suggesting that they are applying their grammatical system to entrain language.

The question of whether writing is language may not be answerable with a simple yes or no. My claim here is that writing *becomes* language.

## 8.	Conclusions and Applications

This paper has argued that a literate person's first script has a special cognitive status—including pre-eminence over later-learned scripts— that is analogous to the special status of a native language. In other words, a literate person has a native script. Other scripts learned later in life suffer the same sorts of disadvantages as second languages: learners find them hard to process, use them with an accent, and experience transfer from their S$_1$.

This paper applies the concept of a native script to the historical preference for adopting existing scripts and the comparative rarity of newly invented scripts. Native speakers of an unwritten language who are educated in a written language will have the script of that non-native written language as their S$_1$. Thus educated speakers of the language, educators, and policy makers will all tend to agree in the identity of, and their preference for, their S$_1$. This preference for S$_1$ means that an established script spreads, even more so than the language with which it is originally associated. The result is a world with many written language but remarkably few different scripts.[9]

The various synchronic consequences of the native script effect are worth considering. These effects occur in the areas of pedagogy and policy. In pedagogy, the question arises of when Romanization should be used in second-language instruction (Elliott, 2012). In Figures 1 and 2 above, the Romanization was a distraction, reducing S$_2$ input for the learner. On the other hand, if all L$_2$ input must be filtered through a slowly and painfully read S$_2$, language learning as a whole will be slowed

---

9. A partial exception is the linguistic area of South Asia, where many different scripts are used and "there is… a widespread feeling that a self-respecting language should have its own unique script to confirm its status as a language" (Masica, 1996, p. 774). Even though this feeling has led to the invention of a number of scripts for previously unwritten languages, even in South Asia many minority languages are written in the script of the official state language.

and perhaps even abandoned. The best way to use Romanization and when to withdraw it is a question that merits further research.

In orthographic policy, the choice of a script—and of a specific orthography within a script—for a newly written language is likely to involve the native script effect on the part of literate speakers in the community, educators from outside the community, and sometimes even professional linguists. The examples cited earlier from Grenoble and Whaley (2006) and Baker (1997) show this. While the prior existence of a native script for some members of a community does not necessarily imply that that script should be chosen (or not chosen), it may be useful to be aware of the native script effect as one factor influencing speakers' preferences.

There are also consequences to linguistic theory in the native script effect. On the one hand, if scripts have grammar and behave like language, becoming entrained by the brain into the grammatical system, then the study of writing is a more legitimate undertaking for linguists than previously believed by members of the structuralist and generative schools. And the tools and models of linguistics (such as Optimality Theory, as in Ellenblum, 2019) can be appropriately used to study writing.

On the other hand, one implication of the native script effect is that linguists themselves are influenced by their native scripts. This has consequences for linguistic theory, especially phonological theory. For example, the concept of the phoneme developed in the context of alphabetic writing, while the phonological existence of the syllable was slower to gain acceptance in modern linguistics (though well established in other contexts). Famously, Chomsky and Halle's *Sound Pattern of English* (1968) does not contain the word *syllable* (the term *syllabic* is used, but is a feature of vowels). The insight that one's script influences one's view of phonology is not new. The influence of alphabetic writing on phonological theory is noted by Aronoff (1992), who describes "segmentalism" in linguistics, and even more strongly by Faber (1992), who argues that phonemes are no more than epiphenomena of alphabetic literacy. More recently, Port and Leary (2005) have argued that phonological theory has made a fundamental error in positing that the phonological system acts on symbolic, graph-like entities.

In evaluating such claims in light of the native script effect, it is on the one hand possible that Western phonologists have been fooled by their native script into creating a phonological theory that resembles their script. On the other hand, it is also reasonable to suppose that typological features, such as phoneme-sized units, that survive in writing systems (having been successfully grammaticalized by the language system) can be expected to have analogs in primary language, even if such units are not the only valid levels of analysis. A more thoughtful aware-

ness of how writing and primary language interact will be to the benefit of the study of both.

References

Abadzi, Helen (1994). *What We Know About Acquisition of Adult Literacy: Is There Hope?* Washington DC: World Bank.

Alessandrini, Jean (1979). "Une nouvelle classification typographique: Codex 1980." In: *Communication et langage* 43, pp. 35–56.

Aronoff, Mark (1992). "Segmentalism in Linguistics: The Alphabetic Basis of Phonological Theory." In: *The Linguistics of Literacy*. Ed. by P. Downing, S.D. Lima, and M. Noonan. Amsterdam: Benjamins, pp. 71–82.

Baines, John (2004). "The Earliest Egyptian Writing: Development, Context, Purpose." In: *The First Writing: Script Invention as History and Process*. Ed. by Stephen D. Houston. Cambridge: Cambridge University Press, pp. 150–189.

Baker, Philip (1997). "Developing Ways of Writing Vernaculars: Problems and Solutions in a Historical Perspective." In: *Vernacular Literacy: A Re-Evaluation*. Ed. by Andrée Tabouret-Keller et al. Oxford: Clarendon Press, pp. 93–141.

Bassetti, Benedetta (2013). "Bilingualism and Writing Systems." In: *The Handbook of Bilingualism and Multilingualism*. Ed. by Tej K. Bhatia and William C. Ritchie. Malden, MA: Wiley-Blackwell, pp. 649–670.

Bender, Margaret (2002). *Signs of Cherokee Culture: Sequoyah's Syllabary in Eastern Cherokee Life*. Chapel Hill, NC: University of North Carolina Press.

Bloomfield, Leonard (1933). *Language*. New York: Holt, Rinehart.

Chomsky, Noam (1964). *Current Issues in Linguistic Theory*. The Hague: Mouton.

——— (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——— (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper & Row.

Daniels, Peter T. (2018). *An Exploration of Writing*. Sheffield: Equinox.

Daniels, Peter T. and William Bright (1996). "Abbreviations, Conventions, and Definitions." In: *The World's Writing Systems*. Ed. by Peter T. Daniels and William Bright. New York: Oxford University Press, pp. xxxix–xlv.

Dobrovolsky, Michael and William O'Grady (1997). "Writing and Language." In: *Contemporary Linguistics: An Introduction*. Ed. by William O'Grady, Michael Dobrovolsky, and Mark Aronoff. New York: St. Martin's, pp. 553–586.

Ellenblum, Gali (2019). "Cognitive and Computational Modeling of Handwriting." PhD thesis. John Hopkins University.

Elliott, Christine (2012). "Second Language Writing System Word Recognition (with a Focus on Lao)." In: *Journal of the National Council of Less Commonly Taught Languages* 11, pp. 53–73.

Faber, Alice (1992). "Phonemic Segmentation as Epiphenomenon." In: *The Linguistics of Literacy*. Ed. by Pamela Downing, Susan D. Lima, and Michael Noonan. Amsterdam: Benjamins, pp. 111–134.

Farooq, Faisal, Liana Lorigo, and Venu Govindaraju (2006). "On the Accent in Handwriting of Individuals." In: *Tenth International Workshop on Frontiers in Handwriting Recognition*. Ed. by Guy Lorette. La Baule, France: Suvisoft.

Gelb, Ignace J. (1963). *A Study of Writing*. 2nd ed. Chicago: University of Chicago Press.

Gnanadesikan, Amalia E. (2009). *The Writing Revolution: Cuneiform to the Internet*. Malden, MA: Wiley-Blackwell.

———— (2012). "Maldivian Thaana, Japanese Kana, and the Representation of Moras in Writing." In: *Writing Systems Research* 4, pp. 91–102.

Grenoble, Lenore A. and Lindsay J. Whaley (2006). *Saving Languages: An introduction to Language Revitalization*. Cambridge: Cambridge University Press.

Householder, Fred W. (1969). "Review of Language and its Structure: Some Fundamental Linguistic Concepts by Ronald W. Langacker." In: *Language* 45.4, pp. 886–897.

Kaye, Alan S. (1996). "Adaptations of Arabic script." In: *The World's Writing Systems*. Ed. by Peter T. Daniels and William Bright. New York: Oxford University Press, pp. 743–762.

Kim, Jeongsu (2005). "The History and Future of Hangeul: Korea's Indigenous Script." In: trans. by Ross King. Folkestone (Kent): Global Oriental.

Kim, Say Young, Li Liu, and Fan Cao (2017). "How Does First Language (L1) Influence Second Language (L2) Reading in the Brain? Evidence from Korean-English and Chinese-English Bilinguals." In: *Brain & Language* 171, pp. 1–13.

Ledyard, Gari K. (1966). "The Korean Language Reform of 1446: The Origin, Background, and Early History of the Korean Alphabet." PhD thesis. University of California, Berkeley.

———— (1997). "The International Linguistic Background of the Correct Sounds for the Instruction of the People." In: *The Korean Alphabet: Its History and Structure*. Ed. by Young-Key Kim-Renaud. Honolulu: University of Hawai'i Press, pp. 31–87.

Masica, Colin (1996). "South Asia: Coexistence of Scripts." In: *The World's Writing Systems*. Ed. by Peter T. Daniels and William Bright. New York: Oxford University Press, pp. 773–776.

Myers, James (2019). *The Grammar of Chinese Characters: Productive Knowledge of Formal Patterns in an Orthographic System*. New York: Routledge.

Pacton, Sébastien et al. (2001). "Implicit Learning Out of the Lab: The Case of Orthographic Regularities." In: *Journal of Experimental Psychology: General* 130.3, pp. 401–426.

Pinker, Steven (1994). *The Language Instinct*. New York: Harper Collins.

Port, Robert F. and Adam P. Leary (2005). "Against Formal Phonology." In: *Language* 81.4, pp. 927–964.

Ramaiah, Chetan, Shivram Arti, and Govindaraju Venu (2013). "A Bayesian Framework for Modeling Accents in Handwriting." In: *12th International Conference on Document Analysis and Recognition*, pp. 917–921.

Ramaiah, Chetan, Porwal Utkarsh, and Govindaraju Venu (2012). "Accent Detection in Handwriting based on Writing Styles." In: *10th IAPR International Workshop on Document Analysis Studies*, pp. 312–316.

Ringbom, Håkan (1987). *The Role of the First Language in Foreign Language Learning*. Clevedon: Multilingual Matters.

Rogers, Henry (2005). *Writing Systems: A Linguistic Approach*. Malden, MA: Blackwell.

Tsai, Kuan-Chun and Terezinha Nunes (2003). "The Role of Character Schema in Learning Novel Chinese Characters." In: *Reading Development in Chinese Children*. Ed. by Catherine-McBride-Chang and Hsuan-Chih Chen. Westport, CT: Praeger, pp. 109–125.

Weingarten, Rüdiger (2011). "Comparative graphematics." In: *Written Language & Literacy* 14.1, pp. 12–38.

Woods, Christopher (2010). "Visible Language: The Earliest Writing Systems." In: *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond*. Ed. by Christopher Woods, Emily Teeter, and Geoff Emberling. Chicago: Oriental Institute, pp. 15–25.