Central Lancashire Online Knowledge (CLoK)

Title	Universality in eye movements and reading: a trilingual investigation
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/22352/
DOI	https://doi.org/10.1016/j.cognition.2015.10.013
Date	2016
Citation	Liversedge, Simon Paul, Drieghe, Denis, Li, Xin, Yan, Guoli, Bai, Xuejun and Hyönä, Jukka (2016) Universality in eye movements and reading: a trilingual investigation. Cognition, 147. pp. 1-20. ISSN 0010-0277
Creators	Liversedge, Simon Paul, Drieghe, Denis, Li, Xin, Yan, Guoli, Bai, Xuejun and Hyönä, Jukka

It is advisable to refer to the publisher's version if you intend to cite from the work.
https://doi.org/10.1016/j.cognition.2015.10.013

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

Universality in eye movements and reading:

A trilingual investigation

Simon P. Liversedge[1], Denis Drieghe[1], Xin Li[2],

Guoli Yan[2], Xuejun Bai[2]

and

Jukka Hyönä[3]

1. University of Southampton, UK

2. Tianjin Normal University, P.R. China

3. University of Turku, Finland

Send correspondence to:

Simon P. Liversedge

School of Psychology,

University of Southampton,

Southampton,

SO17 1BJ, UK

Tel: +44 23 8059 9399

Fax +44 23 8059 4597

Email: s.p.liversedge@soton.ac.uk

Abstract

Universality in language has been a core issue in the fields of linguistics and psycholinguistics for many years (e.g., Chomsky, 1965). Recently, Frost (2012) has argued that establishing universals of process is critical to the development of meaningful, theoretically motivated, cross-linguistic models of reading. In contrast, other researchers argue that there is no such thing as universals of reading (e.g., Coltheart & Crain, 2012). Reading is a complex, visually mediated psychological process, and eye movements are the behavioral means by which we encode the visual information required for linguistic processing. To investigate universality of representation and process across languages we examined eye movement behavior during reading of very comparable stimuli in three languages, Chinese, English and Finnish. These languages differ in numerous respects (character based vs. alphabetic, visual density, informational density, word spacing, orthographic depth, agglutination, etc.). We used Linear mixed modelling techniques to identify variables that captured common variance across languages. Despite fundamental visual and linguistic differences in the orthographies, statistical models of reading behavior were strikingly similar in a number of respects, and thus, we argue that their composition might reflect universality of representation and process in reading.

1. Introduction

The issue of universality has been central to linguistics and psycholinguistics for decades. Chomsky (1965) argued that "…the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages, and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning" (pp. 27-28). In a recent article, Frost (2012) makes a strong argument for a Universality Constraint in relation to reading, suggesting that psychological models of the process of reading should reflect cognitive operations that are common across languages with different writing systems or scripts. In his thesis, Frost goes beyond the traditional Chomskyan notions of universality, making the case for cross-linguistic commonality at the level of cognitive processing. Frost further argues that establishing reading universals is a prerequisite for the development of meaningful, theoretically motivated cross-linguistic models of reading. The responses to Frost's target article are very interesting in that they reveal a broad spectrum of views pertaining to the issue of universality in relation to written language processing, ranging from broad agreement (e.g., Feldman & Moscoso del Prado Martín, 2012) through to the suggestion that there are no such things as universals of reading (Coltheart & Crain, 2012). The views delivered in the article along with the responses to them provide a very relevant context to the experimental project that we report here. We were keen to investigate whether it might be possible to identify factors that could account for common variance across very different written languages in an on-line measure known to reflect moment-to-moment cognitive processing during reading. Our objective in doing this was to first establish

whether such variables did exist, and if so, try to evaluate whether those variables might represent universal aspects of reading. If such universals do exist, they represent common principles by which the written language processing system extracts information from print across different languages. Indeed, if this is the case, then one of the strongest predictions that can be made on the basis of Frost's universality formulation is that whilst different writing systems may visually represent linguistic information in quite different ways, the extraction of meaning from comparable units of language should require a similar amount of time. That is, whilst the moment-to-moment machinations of meaning computation may differ across languages, overall, the time to compute meaning from comparable units of written language should be similar. Arguably, at a fundamental level, universality suggests that an assumption of temporal unity in relation to the attainment of comprehension (regardless of visual format) should hold, and this in turn strongly implies comparability in the time required to attain that state. Finally, our approach in this project also provided an opportunity to pursue a more general objective, namely, to provide comparable cross-linguistic descriptives of reading behavior.

Before developing our claims in detail, it is necessary to be clear about two points. First, unlike the implicit position adopted by Frost (2012), we do not consider theories of written word identification to be the equivalent of theories of reading (see Liversedge, Blythe & Drieghe, 2012). Instead, we consider comprehension of multi-word text to constitute reading, rather than simply the identification of isolated words. Furthermore, it is our view that word identification occurs differently for isolated words than during normal reading (see Rayner & Liversedge, 2011). Thus, whilst word identification is clearly a central and critical aspect of reading, numerous other cognitive processes are also required for successful text comprehension (e.g.,

syntactic, semantic, discourse processes, anaphor resolution, inferential processing, etc.). For these reasons, when we discuss reading in the present article we include consideration of processing beyond word identification. Our second qualification concerns exactly what we mean when we refer to universality. As should become clear, we do not restrict our use of the term to the notion of Formal and Substantive Universals as originally stipulated by Chomsky (1965). Instead, perhaps unsurprisingly, we will consider universality in relation to representations and cognitive processes that are common to reading across languages (with the exception of Braille). It is in this sense that our claims will be about aspects of written language processing that are universal.

As mentioned, reading is a visually mediated psychological process. Humans process visual information via the eyes. Visual information, and more specifically in the case of reading, text, is encoded and then represented in an abstract form after which it is linguistically processed by later cognitive systems. Written language comprehension results in the formation of a representation of the meaning of text, often referred to as a discourse representation. In this sense, the human visual processing system (including "the brain's letter-box", Dehaene, 2009) sub-serves the linguistic processing system, delivering the information that the language processor needs in order to carry out its computations. As already indicated, the eyes are the means by which visual information is encoded for subsequent processing, and the human eye has a particular physiological make up that has important implications for the eyes' functional role in the uptake of visual linguistic information. At (approximately) the middle of the retina there is the fovea, a small circular area (roughly 2 degrees), that provides high acuity visual information, and beyond which, in the parafovea and the periphery, vision is of much reduced visual acuity.

5

Consequently, this causes humans to visually sample their environment by making a series of fixations, which are short periods where the eye is comparatively still (usually lasting about quarter of a second during reading), and saccades, which are fast, ballistic rotations of the eyeball. During fixations humans cognitively process the visual information that they have encoded, whereas during saccades, there is no useful visual input. All humans across all cultures who have an undisrupted visual system visually sample their environment in this way (Findlay & Gilchrist, 2003), and it has been argued that saccadic sampling has evolved due to its efficiency in relation to visual information processing (Gilchrist, Brown & Findlay, 1997). Furthermore, eye movements are very largely under cognitive control, and measurement of temporal and spatial properties of saccades and fixations during reading provides an excellent on-line index of cognitive processing (Liversedge & Findlay, 2000; Rayner, 1998; 2009). Thus, despite our perceptual experience during reading being one of a smooth, continuous flow of information, in fact, it takes place via a staccato succession of discrete snapshots, each providing detailed information from a small portion of the sentence (usually a word or two). In other words, detailed visual information necessary for linguistic processing beyond the centrally fixated (foveal) region is not available. It is important to understand, however, that readers do not exclusively process text directly at fixation. If this were the case, then linguistic processing would be extremely tightly yoked to specific fixations (c.f., the Eye-Mind Assumption, Just & Carpenter, 1980). Instead, there has been substantial work (see Rayner, 1998; 2009) showing that readers partially pre-process upcoming text in the parafovea in the direction of reading (McConkie & Rayner, 1975; Rayner,1975). In sum, saccadic eye movements during reading are common throughout the human species, regardless of culture or language, and saccadic sampling and the retinal

make-up constrain the rate at which visual information is encoded and delivered by the visual system to the language processing system.

Not only does commonality exist in relation to human eye movements during reading, but also certain linguistic effects on eye movement behaviour occur across languages. For example, lexical frequency effects are known to occur robustly across most languages such that words that are more frequent are read more quickly than words that are less frequent (Ellis, 2002). Also, word length effects have been demonstrated across languages, whereby longer words take longer to read than shorter words (Bertram & Hyönä, 2003; Just & Carpenter, 1980; Rayner, Sereno & Raney, 1996). Finally, words that are more predictable on the basis of preceding sentential context are read more quickly than words that are less predictable (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Inhoff, 1984). The fact that word frequency, word length and word predictability effects (the "big three" in reading, Clifton, Ferreira, Henderson, Inhoff, Reichle & Schotter, 2015), are found across languages provides evidence for the more general suggestion of the importance of word based processing during reading across languages (see Li, Bicknell, Liu, Wei & Rayner, 2014).

The next point that we will consider concerns the script, or the physical form of a written language. As Perfetti and Harris (2013) make clear, reading depends on the writing system that encodes the language. We strongly concur with this view. Here, we will consider the writing system of the language in relation to two issues relevant to reading: (1) the visual and informational density of the written language, and (2) the intricacies of the orthography (notational system) that capture and represent linguistic characteristics.

Scripts vary across languages to a very significant degree. Some written languages are extremely, visually dense (e.g., Chinese), whilst others are less dense and (usually) horizontally spatially extended (e.g. English, or even more so, Finnish). To be clear, by visual density, we mean the amount of visual information that is available per unit of text. This definition, in itself, raises complexities in relation to what actually constitutes a unit of text. For the moment, however, let us sidestep this question and consider visual density in relation to one of the three languages we have chosen to examine. In written Chinese, visual density can be indexed in terms of the stroke complexity of characters and words (e.g., Liversedge, Zang, Bai, Yan & Drieghe, 2014) since all characters occupy the same unit of space, and some characters are comprised of many strokes, whilst others are comprised of comparatively few; the more strokes that comprise a character or word, the greater the visual density. Note, though, that not all strokes carry equivalent weight within a character (Wang, Schotter, Angele, Wang, Pomplun & Rayner, 2013), and for this reason, stroke count in relation to the relative importance of those strokes may offer a more veridical index of informational density. In English, Finnish and other alphabetic languages, the letters that comprise words are directly comparable, and to this extent, words of equal length can be considered to be comparably visually dense. Furthermore, it is generally accepted that in alphabetic languages words comprised of more characters are more visually complex than words comprised of fewer characters, and in this sense, alphabetic word length is a proxy for visual complexity (Liversedge et al. 2014). However, on average, words are longer in Finnish than in English (and most often, there will be fewer words in the Finnish than the English version of a directly translated sentence), and therefore, in relation to the amount of information a character in a word conveys, then the informational density of words in English is

greater than in Finnish. That is to say, on average, in English more information is

packed into a smaller word unit than is the case in Finnish. This is partly due to

prepositions being expressed as inflectional suffixes in Finnish. From that perspective,

Finnish words may be argued to contain more information than English words. Yet,

information density per character is still greater in English due to written Finnish

marking all phonemes with separate graphemes which lengthens the written words

(see below, for more details concerning grapheme-phoneme correspondence in the

studied languages).

A key point is that the visual density of Chinese, and the informational density

of English and Finnish directly impact on how readily the written form of the

language can be encoded during a fixation, or across multiple fixations (Liversedge,

et al., 2013). Recall that detailed visual information is only encoded from a limited

portion of the retina. When a script is visually or informationally dense, a greater

amount of information is foveally available to be processed on any particular fixation,

and therefore fewer fixations are required to encode that information. However, at

least to some extent, due to the increased visual or informational density of that

information, the duration of the fixations required to encode it will be increased. In

contrast, for less dense scripts, readers need to make a greater number of fixations in

order to encode a comparable amount of information (since less of the script is

foveally available during any single fixation). Consequently, for less dense scripts,

additional fixations are required for successful encoding. Note, however, fixations

during reading of such languages will be shorter in duration because the amount of

information encoded is reduced. In this basic respect we can start out by formulating

a strong experimental prediction in relation to eye movement behavior during reading

across languages. Specifically, we can anticipate a trade-off between fixation

durations and numbers of fixations that is related to the visual or informational density of the written language. For more dense scripts, fewer fixations of longer duration will occur. In contrast, for less dense scripts, there will be more fixations but these will be of shorter duration.

The visual and informational density of scripts is not the only variable that is important in relation to differences in eye movements during reading in different languages. Different scripts also vary in the way linguistic information is conveyed in their annotated form, that is, there are differences in their orthography. It has been argued that there are three broad categories of written language: logographic (or morpho-syllabic), syllabic and alphabetic languages, with each category being characterised according to the predominant correspondence between the unit of spoken form and the unit of the written language (e.g., Gelb, 1952). A good example of a logographic language is Chinese; a syllabic language is Japanese, and English and Finnish are both alphabetic languages. There are, however, further relevant and important distinctions between different orthographies of written languages that relate to the systematicity of the mapping relationship between written form and sound. In logographic languages like Chinese, in which syllables rather than individual phonemes comprise the basic units of sound, many visually different characters can represent the same syllable. This means that the relationship between the visual and spoken form of a character is arbitrary to a significant degree. Orthographic depth refers to the degree of consistency between the orthographic and phonological forms: the less consistent the relationship, the deeper the orthography. In relation to our examples, Chinese has a very deep orthography, with little by way of consistency in the mapping between characters and syllables. In English, a less deep orthography, whilst there is significant consistency, there is also a degree of inconsistency (e.g.,

mint, *pint*). Finally, in Finnish, a language with a very shallow orthography, the relationship is so consistent and transparent that the manner in which words in the language are pronounced is fully specified by their written form. Thus, phonological ambiguity in relation to orthographic form is very prevalent in Chinese, present but less abundant in English and virtually absent in Finnish. A final noteworthy point is that it has been consistently demonstrated (Durgunoglu, 2006; Ellis & Hooper, 2001; Hanley, Masterson, Spencer & Evan, 2004; Spencer & Hanley, 2003; Seymour, Aro & Erskin, 2005; Ziegler, Bertrand, Tóth, Csépe, Reis, Faísca, Saine, Lyytinen, Vaessen, & Blomert, 2010) that children's reading, as indexed by their ability to read out loud, appears to develop more rapidly in languages with shallower orthographies (though whether there are corresponding differences in comprehension remains an open question, Seidenberg, 2011). The point to take here is that scripts of languages not only differ in their visual and informational densities, but also in how the orthography conveys linguistic information. To this extent, written languages also vary in their linguistic specificity. We need to now consider this fact in relation to the arguments we have developed concerning commonality in visual processing and eye movement behavior during reading across cultures, and further, how this constrains the delivery of information to the language processing system. It should be very clear that because visual encoding (both in relation to retinal acuity and the nature of visual sampling) is constant across readers of different languages, but that the visual and linguistic characteristics of different written languages are themselves very different, then the rate and nature of cognitive processing during reading will be necessarily constrained to differing degrees for those different languages.

Given this argument, the reader may be confused by our intention to investigate universality of processing during reading. Indeed, thus far, the issues we

have raised seem to point very directly to the suggestion that, if anything, there should be a significant degree of language specificity in relation to the rate and nature of linguistic processing during reading. However, if there were universal representations associated with reading, and those representations mediated the nature of such processing across different languages, then a strong hypothesis would be that even in the light of marked linguistic differences between the written forms of languages, it should still be possible to observe commonality in relation to representation and process. In order to show this, however, it would be necessary to construct maximally comparable translations of text stimuli across three quite different writing systems (e.g., Chinese, English & Finnish). This would ensure that any differences that did occur would not be due to language specific content differences. Furthermore, it would be necessary to record eye movements using identical methods from expert native readers of comparable linguistic proficiency, thereby avoiding the possibility that cross-linguistic effects could be attributable to measurement or participant differences. If these requirements could be met then we might argue that universality of process (and representation in relation to such process) should be associated with variables that capture common cross-linguistic statistical variance in the eye movement data. To reiterate our claim, given the combination of visual encoding that is operationally the same across languages, and cross-linguistic specificity in relation to visual and linguistic characteristics of text, one might assume that linguistic processing during reading, and more precisely, the construction of a phonological representation on the basis of orthography, proceeds in a language-specific, idiosyncratic fashion. Alternatively, despite cross-language differences in relation to visual and linguistic form, and the fact that these are encoded via a visual system that is operationally the same regardless of the particular characteristics of a script, at a

12

very fundamental level, the linguistic representations and processes that exist across languages might share commonality.

To investigate this issue, we conducted an eye movement experiment across three languages. The languages that we selected were Chinese, English and Finnish. We selected these languages very purposefully due to their particular linguistic characteristics. As described earlier, Chinese is logographic, whereas English and Finnish are alphabetic. Chinese has the deepest orthography of the three languages. English and Finnish also differ in terms of their orthographic depth, even though they are both alphabetic. English is a comparatively deep orthography, whereas Finnish is one of the most shallow of all orthographies. Another important difference between the languages concerns the fact that it is not the case that there is perfect word-to-word correspondence across the three languages - some words that occur in one language do not occur in the other languages. For example, most prepositions in Finnish are coded morphologically as part of a noun. Thus, there are a significant number of characters in Chinese, and short function words that appear in English that do not appear in Finnish. Similarly, articles appear in English, but not in Chinese or Finnish, and some characters that feature in Chinese do not have corresponding words in English or Finnish (e.g., operators of quantity; characters marking possessives, characters marking words as an adjective or a noun, etc.).

In addition to cross-linguistic differences, visual differences between the three selected scripts are also plentiful. Chinese is very dense with over 90% of the words in the language comprising one or two characters. Furthermore, the characters themselves have constituent radicals that are comprised of individual strokes, and importantly, there is structure to the arrangement of these strokes. In fact, the stroke structure of a character has implications for processing during Chinese character

identification and reading (Wang et al., 2013; Yan et al., 2012). In stark contrast to Chinese, because Finnish and English are alphabetic, the written form is horizontally spatially extended to a greater degree, with the constituent letters of words being presented horizontally adjacent to each other. Alphabetic letters are less visually dense than Chinese characters. Note, also, that since Finnish is an agglutinative language, being comprised of a high proportion of multi-morphemic words, many of the words in Finnish are very long (words of 12 or more letters are quite common). Thus, the horizontal spatial extent of words and sentences is greater in Finnish than English, and far greater than in Chinese. A final characteristic of Chinese that makes it visually distinct from Finnish and English, is that Chinese is an unspaced language, that is, there are no spaces between the words in Chinese sentences. The lack of word spacing in character-based languages has been shown to be very important in relation to eye movements, saccadic targeting and reading (see Bai, Yan, Liversedge, Zang, & Rayner, 2008; Blythe, Liang, Zang, Wang, Yan, Bai & Liversedge, 2012; Li, Liu & Rayner, 2011; Sainio, Hyönä, Bingushi & Bertram, 2007; Shen, Liversedge, Tian, Zang, Cui, Bai, Yan, & Rayner, 2012; Yan, Kliegl, Richter, Nuthmann & Shu, 2010; Zang, Liang, Bai, Yan & Liversedge, 2012). The lack of spaces between words in Chinese contributes further to its reduced horizontal extent, and this also means that a process of word segmentation is required for word identification to occur that is unnecessary in English and Finnish (with the exception of long, multimorphemic compound words, Bertram, Pollatsek, & Hyönä, 2004).

Eye movement recordings have been used to investigate aspects of alphabetic reading and aspects of non-alphabetic reading separately for many decades (Huey, 1900; Javal, 1878, 1879; Shen, 1927; Tinker, 1936a,b). However, to date, there have only been two experiments that have directly compared eye movements during

reading across alphabetic and character-based languages (Feng, Miller, Shu & Zhang, 2009; Sun, Morita & Stark, 1985), and both of these examined Chinese and English readers. To date, no study has investigated Chinese reading in relation to an alphabetic language with a shallow orthography such as Finnish (and indeed, no study has investigated differences in relation to reading of alphabetic languages with shallow and deep orthographies such as Finnish and English respectively).

In the study by Sun et al. (1985) participants whose native language was Chinese were required to read paragraphs of text (taken from *Scientific American*) that had been translated into Chinese, as their eye movements were recorded. Their eye movement recordings were compared with eye movement data obtained from native English speaking participants with a similar level of scientific training reading the original English versions of the texts. Sun et al.'s analyses of Chinese and English reading behavior were largely descriptive with few formal statistical comparisons, and overall, they focused on the similarities in eye movement behavior that existed between the two languages. They showed that the general pattern of saccades and fixations was broadly similar in the two languages, with readers making a succession of left to right saccades and fixations as the text was read. They also argued that fixation durations were comparable in the two languages (though note that the average fixation duration was longer for English [270ms] than for Chinese [260ms]), and that there were similar numbers of regressive eye movements. Furthermore, when they applied a scaling factor of 1.5 (number of Chinese words to English), then further eye movement metrics also became quite comparable. For example, saccades were 1.7 Chinese units relative to 1.8 words in English, and under this conversion, reading rates were 390 units per minute in Chinese compared to 380 words per minute in English. Indeed, as Sun et al. argued, these values are quite comparable, however,

15

as we will see later, the method of scaling employed by Sun et al. is quite arbitrary (see also Yang & McConkie, 1994). Quite what it means to say that 1.5 words in Chinese is the equivalent of 1 word in English is not at all clear. Furthermore, such scaling can mask real, important differences that exist in eye movement behavior during reading of Chinese and English. We will return to these issues below.

Feng et al. (2009) undertook the second study that attempted to quantify differences in reading behavior between alphabetic and non-alphabetic languages. In their study, as with the study by Sun et al. (1985), they compared reading behavior for Chinese and English. However, Feng et al. were primarily concerned with cross-linguistic changes in eye movement behavior during reading across development, focusing both on the unique characteristics of the different orthographies, along with psychological changes associated with maturation and reading development. Feng et al. focused on three aspects of the languages for which there were pronounced differences; the fundamental linguistic units of each language, its orthographic depth, and the presence of boundary demarcations between linguistic units (word spacing). They tested Chinese and English readers in three age groups; 9 years, 11 years and adults. For their stimuli they used a combination of texts that were direct translations from English to Chinese (16% of stimuli), and texts that differed in content between the age groups and languages (84% of stimuli), and in their analyses, they reported results from data collapsed across each type of stimulus. This is a critical aspect of this study, in that this design, at least in principle, allows for the possibility that differences in content between stimuli could have contributed to any effects (or lack of them) between languages (or ages). Nevertheless, the main findings from the adult participants reported by Feng et al. are relevant to some of the issues that we investigated in the current study, and we will, therefore, consider these briefly. To

summarise, for reading times per word, number of fixations per word, mean fixation durations per word, refixations on words and number of progressive fixations, Feng et al. found no reliable differences between Chinese and English readers. Saccades were longer on average in Chinese than in English readers, though note that the physical size of the text was made approximately equal, and given that word units rather than text size is a primary influence on saccade targeting (Morrison & Rayner, 1981) this could have artificially inflated the length of saccades in Chinese relative to English. Finally, Feng et al. found that, on average, Chinese readers made more regressions than English readers, although they tentatively suggested this effect might be an artefact associated with the software they used to establish word boundaries in their Chinese stimuli.

As should be clear, both the study by Sun et al. (1985) and the study by Feng et al. (2009) suggest commonality in the characteristics of the eye movements of Chinese and English readers. However, in both studies it is not unambiguously clear that the failure to find differences in eye movements between Chinese and English readers, or that those differences that were found, occurred due to extraneous factors. In the present study, therefore, we took several steps to avoid potential confounds in our experimental procedures in order that we might more unequivocally assess the extent of commonality in eye movement behavior during Chinese, Finnish and English reading. To do this, we constructed a series of short expository texts that we translated across Chinese, English and Finnish. In constructing the texts across languages we were very careful to ensure that the content and constructions were as similar as possible, thereby permitting maximal comparability of eye movement behavior across the stimuli in each language (see Figure 1).

Football **can** bridge *social*, cultural and religious divides.

□□□□□□ *//////////□*

Jalkapallo **voi ylittää** *sosiaaliset*, kulttuuriset ja uskonnolliset rajat.

Figure 1. An example of one of the sentences from the texts that were used as stimuli in the experiment. Note that whilst there is one to one correspondence between words across languages in this example sentence, this was not always the case. Also, Chinese is unspaced and comparatively dense, whereas English and Finnish are less dense and word spaced. The average word length is longer for Finnish than for English.

Our rationale in this respect was to use texts that caused readers to engage in linguistic processing that was as similar as possible across the languages. Recall that this approach is quite different, and arguably more powerful, in comparison to the study of Feng et al. (2009), in that our use of minimally different texts across languages minimised variability due to content differences.

To ensure comparability of eye movement measures during reading, it was necessary to decide upon regions of text that adequately corresponded across languages (and within limits, regardless of text size). This was a non-trivial decision since the nature of the written form of the three languages differs to such a degree. To allow for comparison of equivalents, it was necessary to compute measures of eye movement behavior over portions of text that unambiguously convey the same information in each language. For this reason, the region of analysis that we adopted in this investigation was the sentence. For sentences, in all of the languages it is the case that all the constituents from the first word in the sentence until the following period comprise the text that conveys comparable information. To adopt a more granular level of analysis would bring into question the level of homogeneity of linguistic information contained within corresponding regions over which fixations

were made and eye movement measures computed. Under such circumstances, comparison of eye movement measures between languages may not represent a comparison of like with like (in the course of this project we have come to refer to this issue as the "apples and pears" issue). Our general hypothesis was that, in the absence of cross-linguistic differences in comprehension, any commonality across languages in relation to linguistic representation and process should be revealed as commonality in statistical models accounting for variance in the same measures of eye movement behavior during reading of comparable regions of text in the different languages. If such commonality did exist, then given similar constraints on human visual encoding, but marked visual and orthographic differences between the written forms of Chinese, English and Finnish, commonality may be taken to represent universality in aspects of representation and process during reading of these languages.

2. Method

2.1 Participants. Twenty-five Chinese undergraduate students from Tianjin Normal University, twenty-one English undergraduate students from the University of Southampton and twenty Finnish students from the University of Turku took part in the experiment. The Chinese participants read texts in Chinese, the English participants texts in English and the Finnish participants read texts in Finnish. All participants were undergraduate psychology students, and all had normal or corrected to normal vision. All of the participants were naïve regarding the purpose of the experiment.

2.2 Materials and Design. We constructed 8 short expository texts on a variety of topics (sheep, a car race, football, oil, sugar, restaurant tipping behavior, walking as exercise and wind energy) that were used as stimuli in the experiment. One of these

texts was used as a practice text, and the remaining 7 were used as experimental

stimuli. The texts were initially constructed in English and then translated into

Chinese. Any sentence structures that did not translate directly into Chinese were

changed or removed. After the initial translation into Chinese, the sentences were

then translated back into English to check for consistency. This process was repeated

(and wherever necessary changes made) until the translations were stable. For the

Chinese stimuli, the characters corresponding to words were identified in the

Contemporary Chinese Dictionary (2005). Any multi-character unit that had an entry

in the dictionary was taken to be a word. The English sentences were also translated

into Finnish. Once again, any sentences that did not translate directly were changed,

this time in both the English and Chinese stimuli. Again, after the initial translation

into Finnish, the sentences were translated back to English to ensure consistency, and

the process repeated until stability was achieved. We were careful to ensure that we

did not oversimplify the texts, nor make their constituent sentences syntactically

uniform. In this way, we developed three counterpart stimulus sets, one for Chinese,

one for English and one for Finnish that were maximally comparable in terms of their

linguistic structure and content (see Figure 1)[1]. Each of the texts was split down into

between 2 and 6 (mean = 4.1) pages presented one at a time successively on the

presentation screen to allow for comfortable reading whilst eye movements were

recorded. Each page contained between 1 and 8 sentences (mean = 4.1), and in the

English version, between 28 and 98 words (mean = 59.9). Our experimental design

was between participants. The Chinese stimuli were presented in Song font and the

size of each Chinese character was 21 square pixels. One Chinese character

subtended 0.67° visual angle. The Finnish and English stimuli were presented in

Courier New font and the size of each character was 14 pixels. One character

subtended .46° visual angle.

2.3 Apparatus. All of the participants were tested using the same type of eye tracker,

a SR Research EyeLink 1000 machine (sampling rate = 1000 Hz, spatial resolution

<0.5°). The Chinese experimental stimuli were presented on a 19-inch DELL monitor

with a 1024 × 768 pixel resolution; the English experimental stimuli were presented

on a ViewSonic P227F 20" with similar resolution; the Finnish experimental stimuli

were presented on a ViewSonic G225F 20" monitor, again, with the same resolution.

All stimuli were presented at a viewing distance of 70cm. In all three testing

situations, head movements were minimised by the use of a chin rest and head

restraint.

2.4 Procedure. Each participant was tested individually. Participants were informed

that they would be required to read and understand sentences that would be presented

passage by passage on the display screen. When they finished reading the final

sentence in each page, they pressed a button to terminate the display. After each text,

two questions were asked concerning some aspect of the content explicitly mentioned

in the text. The questions were answered by pressing either a Yes or a No key (e.g.,

Can sugar cause skin wrinkles?) on the computer keyboard. Half of the questions

required a Yes response and half a No response. Prior to the start of the experiment a

nine-point calibration procedure was completed (the same across the three languages)

and computer software calculated the position of the point of fixation on the basis of

the calibration. After a successful calibration (average calibration error less that .5°)

the passages were presented in turn. Participants were recalibrated after each passage.

In total the experiment took approximately 40 minutes.

3. Results and Discussion

Trials where there was tracker loss were removed from the data prior to the analyses. Fixations shorter than 80ms that were within one character of the previous or following fixation were merged, all other fixations shorter than 60ms that were within one character of the previous or following fixation were merged and all other fixations shorter than 60ms or longer than 800ms were removed. Also, when calculating the eye-movement measures, for each participant, data more than 2.5 standard deviations from the mean for that participant were removed. Overall, data loss was below 7.5% across measurements for each language. We assessed comprehension based on responses to the comprehension questions and found that there was no significant difference in comprehension between the three languages $F1$, $F2 < 1$ (Chinese mean = 88%, s.d. = .10, English mean = 89%, s.d. = .07, Finnish mean = 86%, s.d. = .05). These results indicate that readers understood the sentences well, and to a similar degree in each of the languages.

As described earlier, we selected the sentence as our unit of analysis and we computed four measures that reflect global properties of the eye movement behavior when reading (see Liversedge et al., 2004): These were (1) Total sentence reading times, (2) Average Number of Fixations, (3) Average Forward Saccade Size, and (4) Average Fixation Duration.

3.1 Analysis model. For the analyses we ran linear mixed-effects models using the *lmer* programme of the *lme4* package (version lme4_1.1_6) for R (2012) specifying 3 crossed random factors: the participants, the sentence that was being read, and the story to which the sentence belonged. Significance values reflect both participant, item and story variability. Based on previous research (e.g. Rayner, 2009), we

identified the following 3 predictors which were entered as fixed factors into the models: The number of words in the sentence, the average word length of the words in the sentence and the average log frequency of the words featured in the sentence. For the analysis of the Chinese data, we also considered entering the following predictors as fixed factors: The average number of strokes per character as an index of visual density, and total number of characters in the sentence. However, these factors correlated very highly in Chinese with either average word length, number of words, or average frequency (for instance in Chinese, the highest frequency words tend to be single character words with relatively few strokes) leading to issues of multicollinearity. In contrast, for the average number of radicals, multicollinearity was not so much of a problem (all correlations with other factors were below .09), and therefore we did enter this additional factor into the models for the Chinese data.

Table 1 lists the average values of the predictors as a function of language. These data clearly demonstrate some of the differential properties of the languages discussed in the Introduction. Finnish as an agglutinative and morphologically rich language contains fewer words than Chinese or English since a single morphologically complex Finnish compound word often corresponds to several Chinese or English words. Thus, there is a lower overall number of words, a lower average number of words calculated across languages and a higher average word length in Finnish compared to the other two languages in the corpus. Table 1 also clearly indicates that the average frequency calculated across the words in a sentence is considerably higher in English compared to Chinese and Finnish. This is due to the fact that English is the only one of the three languages that has (extremely frequent) articles and prepositions featuring as separate words. In Finnish this information is incorporated morphologically in the noun, and in Chinese articles and prepositions do

not occur. To make sure our measurement of frequency was comparable across languages we standardized the log frequency before using it as a predictor. This transformation was especially important given that the frequencies of both the English and the Finnish data had to be entered simultaneously when we constructed the bilingual[2] model (see below). Finally, note that Chinese has a considerably smaller average word length than the other two languages. Consideration of this difference should be taken with care in that whilst the metric of word length in English and Finnish is based on the number of letters in a word (and is therefore directly comparable), word length in Chinese is based on the number of characters in a word. Since most words in Chinese are one or two characters long, with only a small proportion of words with three, four or more characters, the average word length is shorter and the degree of variability in word length is significantly reduced relative to the alphabetic languages. It is also important to note that one sentence was removed from the Chinese corpus in these analyses after it became apparent that the whole sentence in Chinese consisted of only a single three character word, and as such represented an extreme outlier in relation to both the average word length and the number of words. We also removed this sentence from the English and Finnish corpus in these analyses.

Table 1. Stimulus Descriptives: Total Number of Words, Average Number of Words Per Sentence (Standard Deviations in Parentheses), Average Log Word Frequency Per Sentence, and the Average Word Length Per Sentence (in Characters) in the Text Corpus for the Three Languages.

	Chinese	English	Finnish
Total Number of Words in Corpus	1774	1762	1301
Average Number of Words per Sentence	14.78 (7.63)	14.68 (7.32)	10.84 (5.40)
Average Word Frequency per Sentence	5.68 (0.87)	7.76 (0.73)	5.97 (1.04)
Average Word Length per Sentence (in characters)	1.55 (0.20)	5.63 (0.80)	8.32 (1.44)

Because we did not have strong a priori predictions concerning two-way interactions between our fixed factors we adhered to the following algorithm when constructing the models: (1) Our initial model contained the fixed factors described above without any interactions, (2) model comparisons examined whether the addition of a two-way interaction resulted in a significantly better fit compared to the simpler model without the interaction. This was done for every potential two-way interaction (and also for three-way interactions, though they never contributed), and (3) when after the previous steps one of the fixed factors was not significant either by itself or within a two-way interaction, we ran an additional model comparison to see whether a model without the factor provided a similar fit to the data. If this was the case, the most parsimonious model (i.e. without the fixed factor) was selected for further model comparisons. As a result, the models presented in this paper are sometimes different in terms of listed predictors for different measures.

When the models that resulted individually for English and Finnish were similar, we also ran a model on the data combined for both languages with Language as an additional fixed factor. For these models English was selected as the baseline. We did not construct any trilingual models because average word length was indexed

by number of characters in Chinese and by number of letters in the alphabetic languages, and so might reasonably be considered to amount to comparing the proverbial "apples and pears".

Visual inspection of the dependent variables indicated that most of them were not normally distributed, necessitating transformation. A Box-Cox power transformation was carried out indicating that a square root transformation most effectively normalised the total sentence reading times. The same analysis was carried out for the average fixation duration and the average rightward saccade length. For both these dependent variables a log transformation was most effective and was therefore carried out. It is important to note that for total sentence reading times, the log transformation which is most often used in the field of eye movement reading research, simply resulted in replacing a rightward skew in the data with a leftward skew (hence our adoption of a square root transformation for this measure). Finally, for the number of fixations measure no transformation was carried out because this measure is a count. Table 2 lists the means and standard deviations for our 4 dependent variables across each language. The total sentence time is strikingly similar across the three languages, indicating that sentences matched for content are read at a comparable speed independent of language (indeed, an ANOVA showed no overall difference for this measure, $F1(2,56) = 1.26$, $p > .20$, $F2(2,354) = 2.03$, $p > .10$). The importance of this result should not be overlooked. The similarity in total reading times is itself suggestive of commonality in relation to fundamental aspects of processing associated with the construction of a semantic representation of a sentence. Despite very different orthographic (surface) forms across the languages, the primary propositions in the sentences remain the same, and overall comprehension times for

the sentences reflect this similarity regardless of language specificities. We will return to this point in the General Discussion.

Table 2. Global Eye Movement Measures: Total Sentence Reading Times, Number of Fixations, Saccade Size Rightwards (in Characters) and Average Fixation Duration (ms). Standard Deviations Appear in Parentheses.

	Chinese	English	Finnish
Total Sentence Reading Times (ms)	3428 (716)	3093 (777)	3063 (943)
Number of Fixations	13.88 (7.51)	14.81 (2.93)	15.22 (4.44)
Saccade Size Rightwards (characters)	3.19 (0.67)	8.53 (1.55)	9.35 (1.74)
Average Fixation Duration (ms)	245 (27)	207 (32)	199 (25)

Table 3. Likelihood of a Word Being Skipped, Re-fixated or Have a Regression Originate From It. Standard Deviations Appear in parentheses.

	Chinese	English	Finnish
Skipping Rate	.47 (.23)	.36 (.28)	.22 (.28)
Re-fixation Rate	.20 (.18)	.23 (.19)	.38 (.25)
Regression Rate	.15 (.13)	.13 (.12)	.17 (.15)

Table 4. Fixed Effect Estimates for Total Sentence Reading Times. Models for Chinese, English and Finnish.

		Chinese			English			Finnish		
		M	SE	t	M	SE	t	M	SE	t
Intercept		55.89	1.66	33.75***	53.11	1.62	32.69***	52.58	1.86	28.34***
Average Frequency		-1.41	.60	-2.34*	-1.73	0.50	-3.44***	-0.99	0.56	-1.77[+]
Number of Words		1.79	.08	22.32***	1.61	0.06	27.78***	2.15	0.09	23.76***
Average Word Length		6.99	3.71	1.88[+]	3.03	0.65	4.70***	2.68	0.40	6.77***
Average Frequency x Number of Words		-0.17	.08	-2.10*						

[+] p <.10, * p < .05, ** p < .01, *** p <.001

Table 5. Fixed Effect Estimates for Total Sentence Reading Times for the Bilingual Model (English – Finnish).

	Bilingual Model English – Finnish		
	M	SE	t
Intercept	54.07	1.28	42.31***
Language	-4.05	2.46	-1.64
Number of Words	1.88	0.05	35.66***
Average Frequency	-2.25	0.53	-4.24***
Average Word Length	2.74	0.32	8.52***
Language x Number of Words	0.53	0.10	5.36***
Language x Average Frequency	2.17	0.90	2.42*

+ p <.10, * p < .05, ** p < .01, *** p <.001

Turning to the other measures, a very interesting picture emerges that illustrates differences in processing as a function of the density of the language whereby the least dense language, Finnish, produces more and shorter fixations, along with slightly longer saccade sizes compared to a more dense language such as English, and even more so Chinese, which is the most visually dense of the three. These findings are exactly in line with our broad overall predictions that fixations and saccades during reading would be influenced by the density of the language with fewer, longer fixations positioned more closely together for Chinese (the most dense language), and more, shorter fixations spaced more sparsely for Finnish (the least dense language). Also, as predicted, the data for English patterned intermediately relative to the other two languages. We will defer consideration of these results in relation to the existing literature until the General Discussion.

Although detailed analyses at the word level are outside of the scope of the current article, we also present the likelihood of a specific word being skipped, re-fixated or have a regression originate from it for the three languages in Table 3. These data show that whereas there are only limited differences in regression rates between the three languages, noticeable differences do occur for the likelihood of a word being skipped or re-fixated with relatively lower skipping rates for Finnish compared to Chinese (and to a lesser extent English) and more re-fixations in Finnish compared to the other two languages. These latter two findings clearly indicate the impact of the relatively increased word length in Finnish compared to the other two languages.

3.2 Total Sentence Reading Times. The fixed factor effects are presented in Table 4 and the total sentence reading times as a function of the fixed factors are presented in

Figure 2. For the three languages an effect of Average Frequency was observed such that sentences with a high average word frequency were read faster than those with a low average word frequency. Not surprisingly a higher number of words led to a longer sentence reading time in all three languages. An effect of average word length was also observed with sentences containing longer words leading to longer reading times, though, note that for Chinese this effect was only marginally significant. No interactions were observed for the three languages with the exception of that between average frequency and number of words in Chinese. The fixed effect associated with this (numerically small) interaction was negative indicating that the processing cost of adding an extra word to a sentence was less pronounced in a sentence containing, on average, high frequency words. This finding makes sense if we assume that reading is a process of incremental interpretation. The interpretation of a word with respect to its preceding sentential context is presumably harder when that context is more difficult to read (i.e., comprised of lower frequency words) than when it is easy to read (i.e. comprised of higher frequency words). The important point to note here is that there is a strong degree of similarity between the separate models for each language, indicating that despite their quite pronounced visual and linguistic differences, to a very significant degree, the same variables captured variance in the eye movement data.
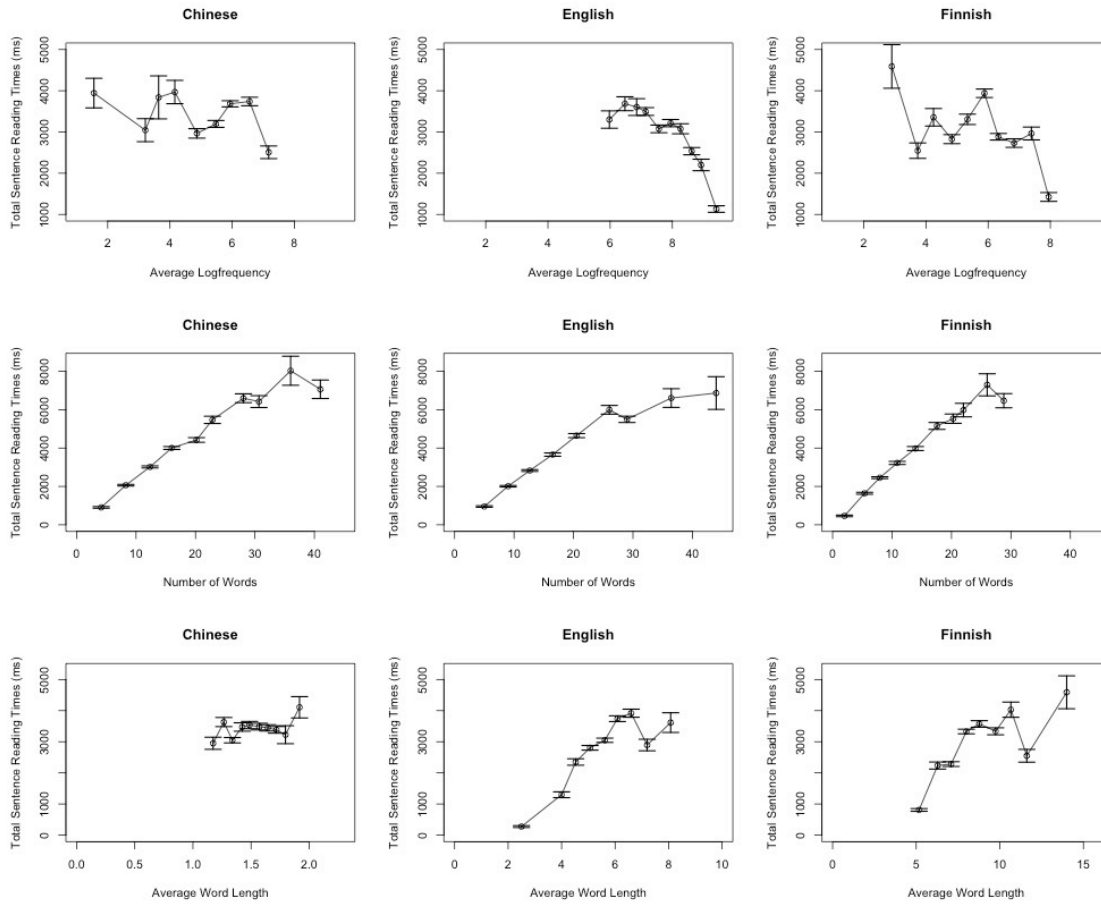
Fig. 2. Total Sentence Reading times (ms) as a function of Average Logfrequency (Top row), Number of Words (Middle Row) and Average Word Length (Bottom Row) for Chinese (Left Column), English (Middle Column) and Finnish (Right Column). For each predictor, fixations were binned into categories containing minimally 10% of the data. Error bars are standard errors of the mean within-subject variance.
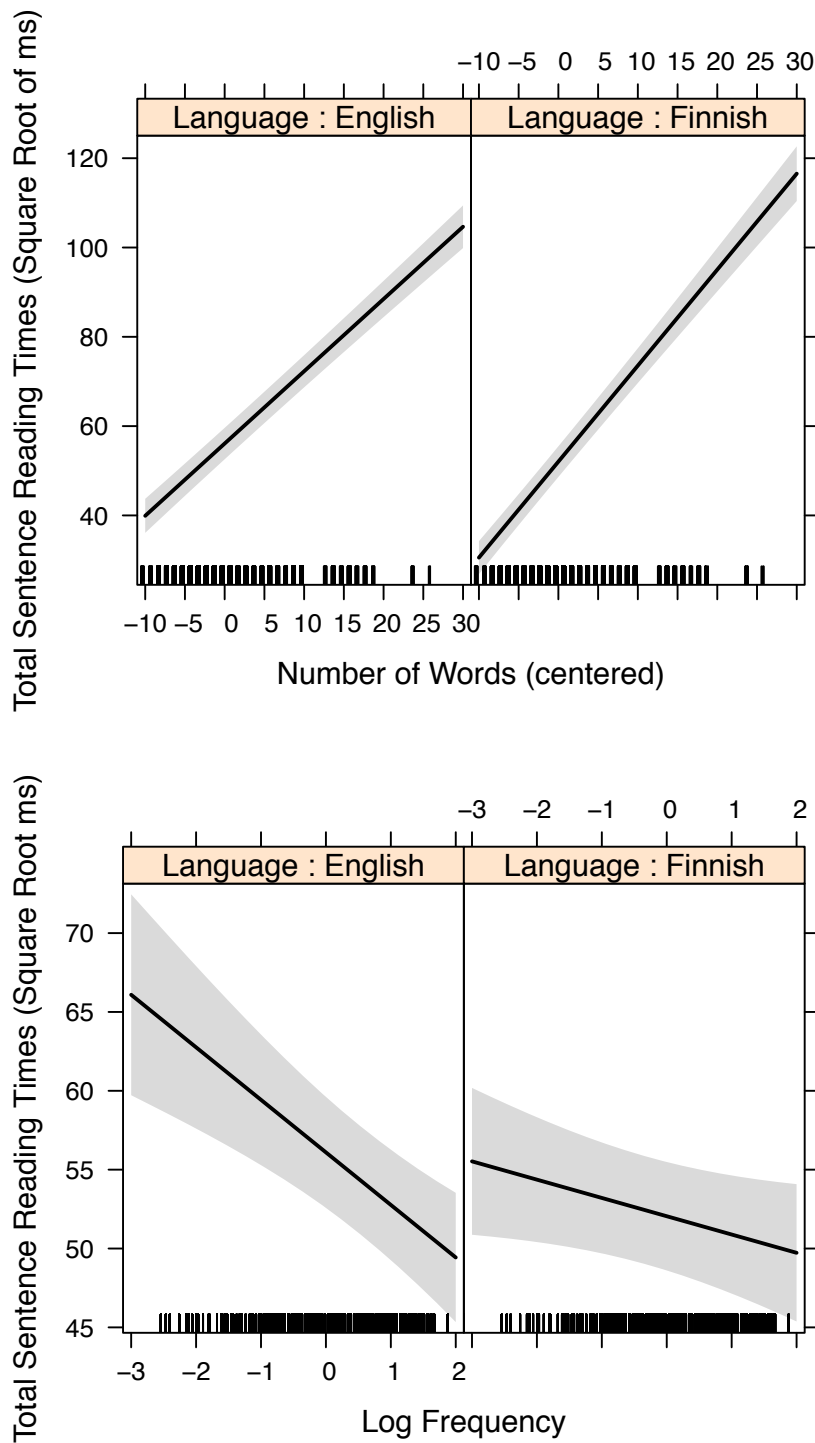
Fig. 3. Effect display for the significant interaction of Language and Number of Words (Top Panel) and Language and Log Frequency (Bottom Panel) in the Bilingual LMM for Total Sentence Reading Times. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

Because the separate models for the alphabetic languages were qualitatively very similar, we constructed a bilingual (English – Finnish) model (see Table 5). In both this bilingual model and the ones constructed for the other measures successive differences contrasts were used and the intercept represents the grand mean. The expected main effects of Number of Words, Average Word Frequency and Average Word Length were significant. The main effect of Language was marginal ($t = -1.64$, $p = .10$), however this was qualified by an interaction between Language and the average Number of Words and an interaction between Language and Average Word Frequency (see Figure 3). The small interaction between Language and Number of Words showed that an extra word in Finnish resulted in a longer increase in total sentence reading times compared to English. This is unsurprising given longer words in Finnish than English (the addition of a word will produce more additional fixations than in English. There was also an interaction between language and average frequency showing more pronounced frequency effects in English than Finnish. Probably, this occurs because of the presence of articles and pronouns in English but not Finnish (though note that frequency measures were standardized so they were coerced into a similar range).

3.3 Total Number of Fixations.

The average number of fixations per language is presented in Table 2, and the fixed factor estimates in Table 6. The average number of fixations is also presented as a function of the fixed factors in Figure 4.
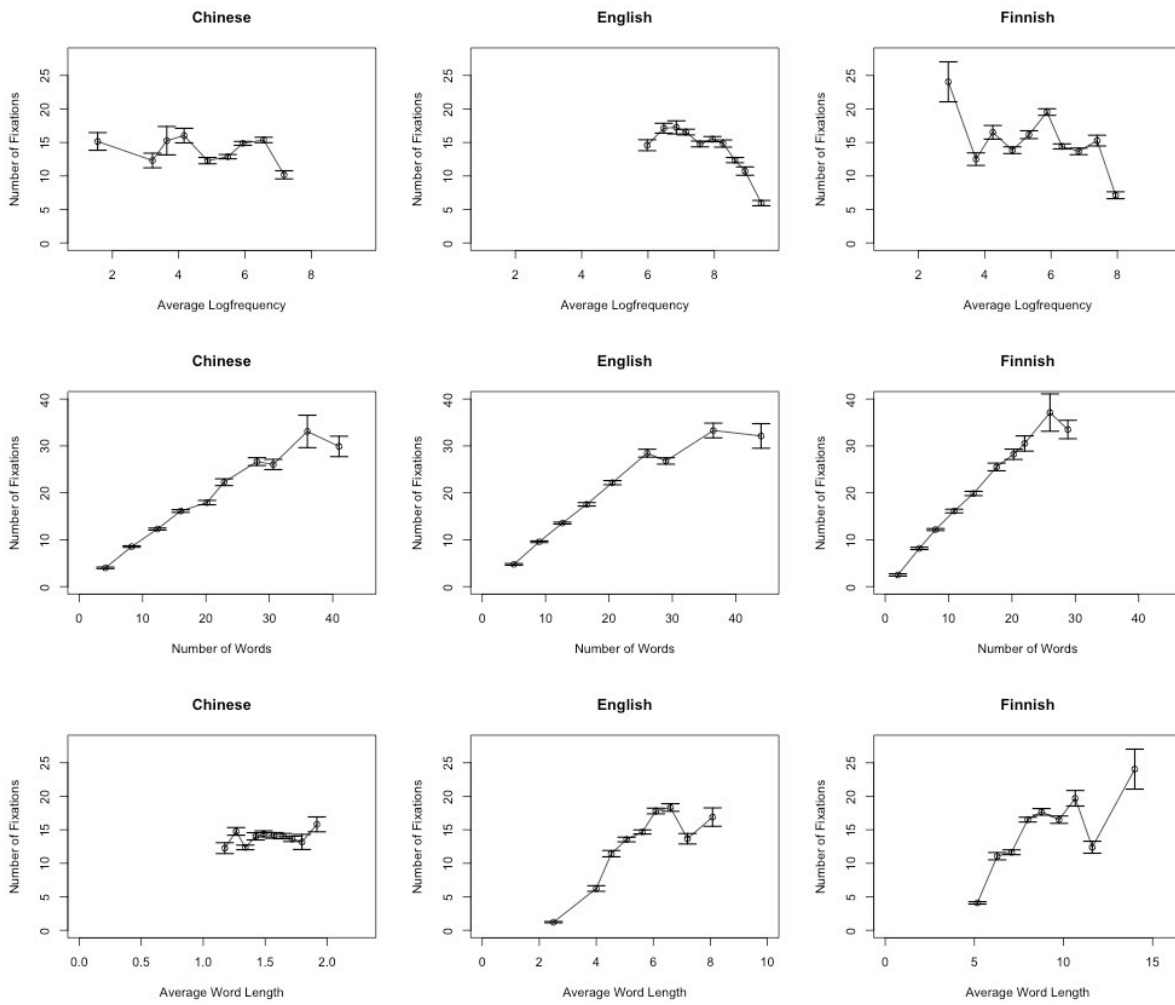
Fig. 4. Number of Fixations as a function of Average Logfrequency (Top row), Number of Words (Middle Row) and Average Word Length (Bottom Row) for Chinese (Left Column), English (Middle Column) and Finnish (Right Column). For each predictor, fixations were binned into categories containing minimally 10% of the data. Error bars are standard errors of the mean within-subject variance.

Table 6. Fixed Effect Estimates for Number of Fixations. Models for Chinese, English and Finnish.

	Chinese			English			Finnish		
	M	SE	t	M	SE	t	M	SE	t
Intercept	13.80	0.72	19.13***	14.73	0.73	20.17***	15.18	1.03	14.72***
Average Frequency	-0.61	0.25	-2.47*	-0.83	0.21	-3.94***	-0.61	0.28	-2.18*
Number of Words	0.88	0.03	26.41***	0.93	0.03	36.85***	1.29	0.04	29.17***
Average Word Length	4.21	1.53	2.75**	1.09	0.27	4.05***	1.42	0.18	7.70***
Average Frequency x Number of Words	-0.07	0.03	-2.04*	-0.10	0.03	-3.81***	-0.19	0.05	-3.65***

[+] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .00$

Table 7. Fixed Effect Estimates for Number of Fixations for the Bilingual Model (English – Finnish).

	Bilingual Model English - Finnish		
	M	SE	t
Intercept	15.26	0.63	24.34***
Language	-0.58	1.29	-0.45
Number of Words	1.11	0.02	45.43***
Average Frequency	-1.11	0.24	-4.62***
Average Word Length	1.36	0.14	9.39***
Number of Words x Average frequency	-0.22	0.02	-9.23***

+ p <.10, * p < .05, ** p < .01, *** p <.001
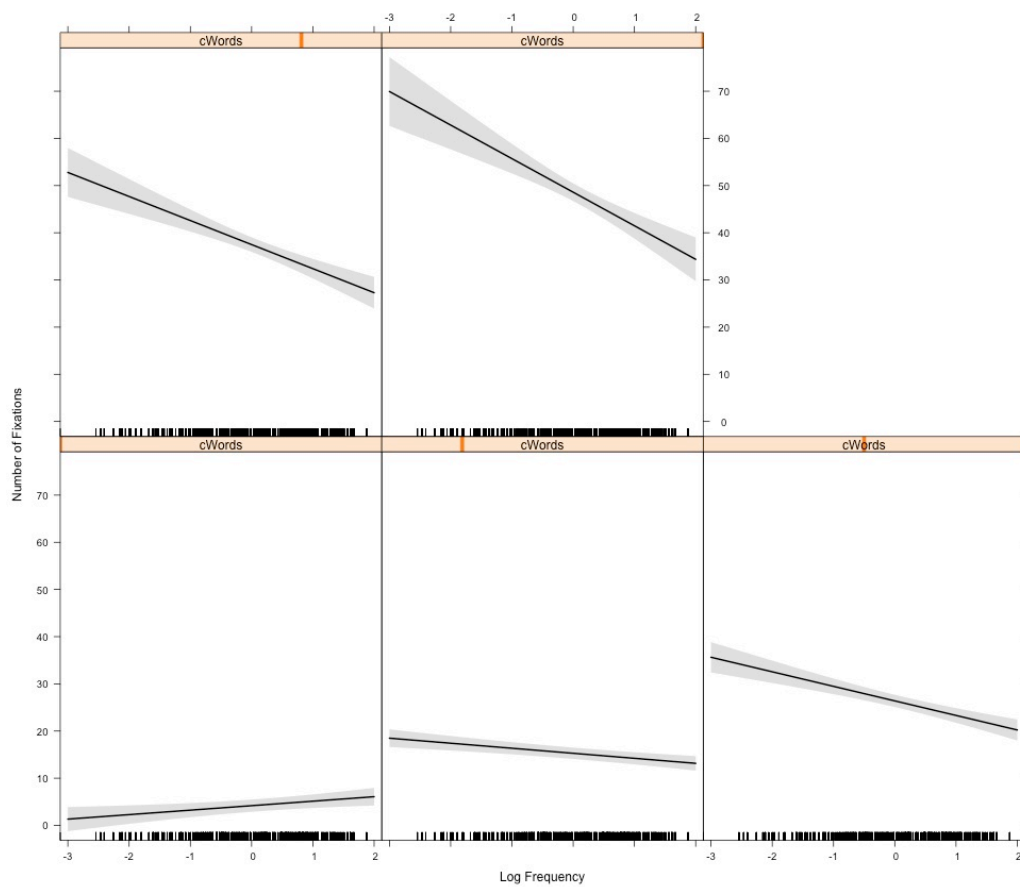
Fig. 5. Effect display for the significant interaction of Logfrequency and number of Words (centered) in the Bilingual LMM for Number of Fixations. The graphs represent different levels in the number of words, increasing from bottom left (sentences with few words) to top right (sentences with many words) as indicated by the bar at the top of the graph. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

The model comparisons resulting in the most parsimonious model with the best fit of the data led to identical models for each of the three languages. The models showed main effects from Average Word Frequency, Number of Words and Average Word Length with fewer fixations on sentences with higher frequency, shorter and fewer words. In all three languages, we also observed an interaction between average frequency and number of words such that the effect of frequency on the number of fixations was more pronounced when there were more words in the sentence, this

could be due to a speedup in a sentence with lots of high-frequency words, resulting in increased skipping behavior. Such an interaction might be more pronounced in a discrete measure such as total number of fixations than a continuous measure such as total sentence reading times, where trade-offs such as increased fixation durations before or after skipping are often observed (Reichle, Rayner & Pollatsek, 2003; see also Reichle & Drieghe, 2013 for a discussion). As with the total sentence reading time, the degree of similarity between the separate models is very striking and indicates that the variables that account for variability in the total number of fixations are similar across languages.

Once again, since the models for number of fixations were qualitatively identical for the alphabetic languages, we constructed the bilingual (English-Finnish) model, which is presented in Table 7. The bilingual model showed no effect of Language resulting in a model that was identical to the separate models for the two languages: Main effects of Word Frequency, Number of Words and Word Length with fewer fixations on sentences with higher frequency, shorter and fewer words. Unsurprisingly, as for the separate models, there was also an interaction between Number of Words and Word Frequency (see Figure 5). Thus, as for the three separate models, it appears that once again the same variables capture variance in the eye movement data similarly for the two languages, and more generally, although we were not able to formally combine the model for the Chinese data with that for the Finnish and English data (due to the different index of word length), it is again quite noteworthy that the three models that we constructed separately for each language were the same.

3.4 Rightwards Saccade Size

The average rightward saccade size per language is presented in Table 2 and as a function of the fixed factors in Figure 6. The fixed factor effect estimates are presented in Table 8.
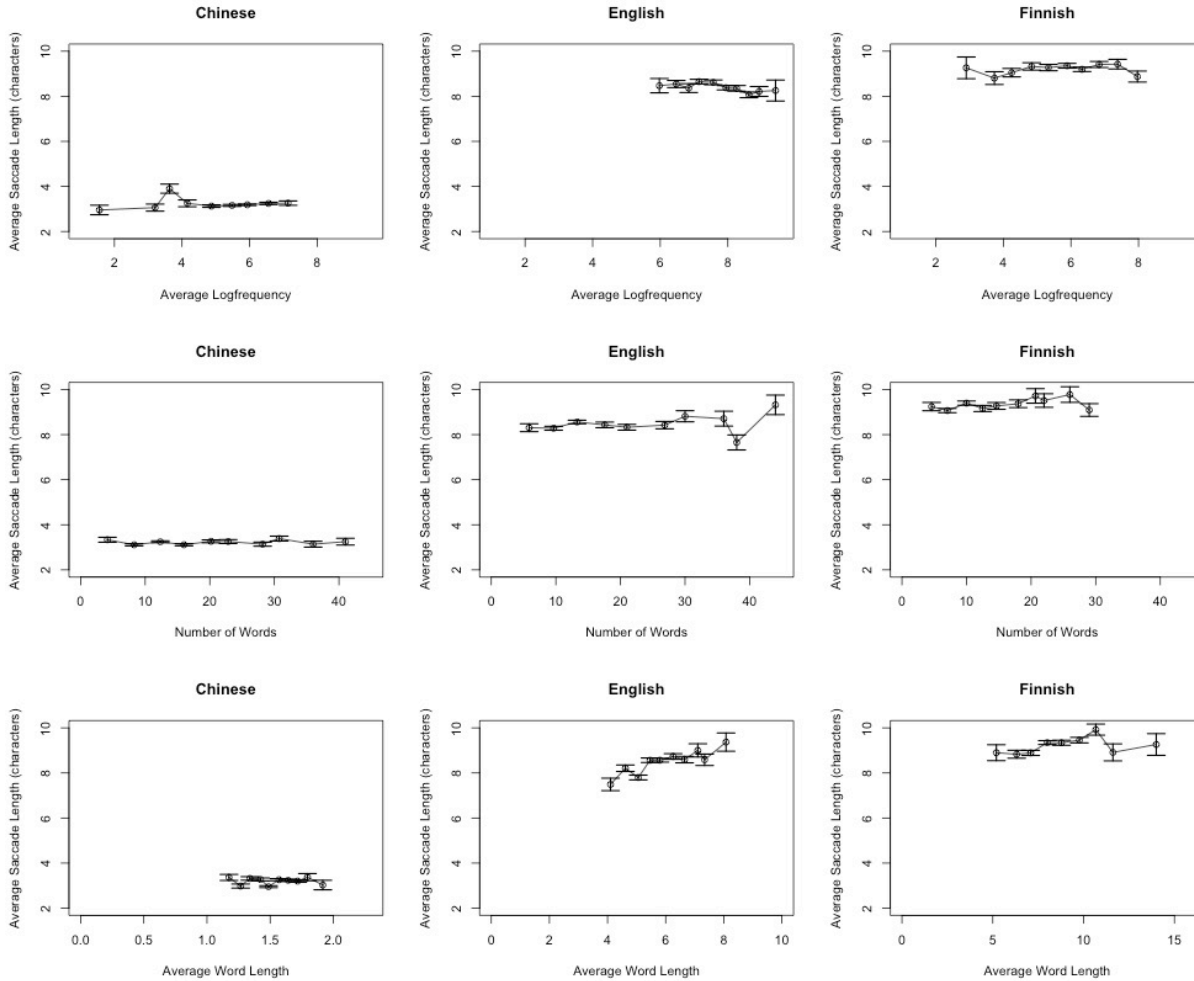

Fig. 6. Average rightwards saccade size (in characters) as a function of Average Logfrequency (Top row), Number of Words (Middle Row) and Average Word Length (Bottom Row) for Chinese (Left Column), English (Middle Column) and Finnish (Right Column). For each predictor, fixations were binned into categories containing minimally 10% of the data. Error bars are standard errors of the mean within-subject variance.

Table 8. Fixed Effect Estimates for Rightwards Saccade Size. Models for Chinese, English and Finnish.

		Chinese			English			Finnish		
	M	SE	t	M	SE	t	M	SE	t
Intercept	1.11	0.05	22.86***	2.12	0.04	53.53***	2.20	0.05	48.13***
Average Frequency	0.01	0.01	1.27	0.00	0.01	0.34	0.02	0.01	2.21*
Average Word Length	0.06	0.07	0.85	0.01	0.01	4.53***	0.02	0.01	3.49***
Average Frequency x Average Word Length	-0.15	0.07	-2.08*	0.05	0.01	2.69**			
Average Number of Radicals	-0.14	0.05	-3.06**						

$^+$ p < .10, * p < .05, ** p < .01, *** p < .001

Table 9. Fixed Effect Estimates for Rightward Saccade Length. Bilingual model
(English – Finnish).

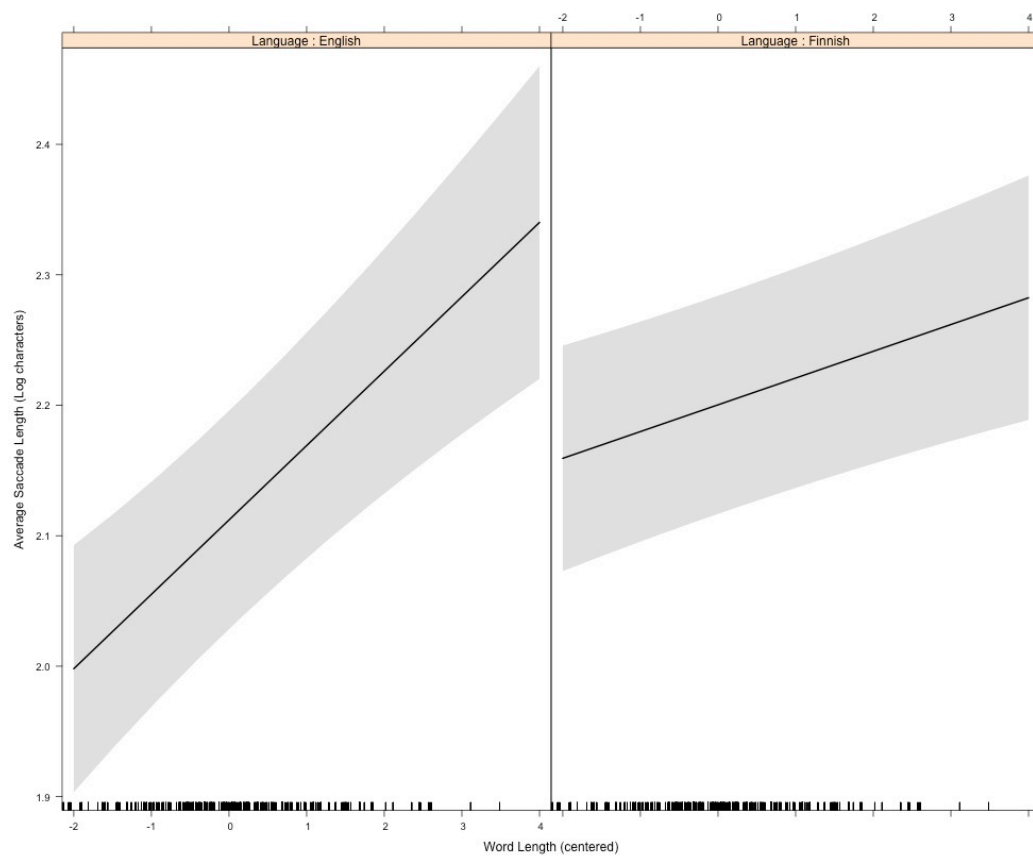|	Bilingual Model English - Finnish		
	M	SE	t
Intercept	2.16	0.03	70.84***
Language	0.09	0.06	1.48
Average Frequency	0.01	0.01	2.22*
Average Word Length	0.04	0.01	5.73***
Language x Average Word Length	-0.04	0.01	-3.30**

+ p <.10, * p < .05, ** p < .01, *** p <.001

Fig. 7. Effect display for the significant interaction of Language and Word Length in the Bilingual LMM for the Average Saccade Length. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

Table 10. Fixed Effect Estimates for Average Fixation Duration. Models for Chinese, English and Finnish.

	Chinese			English			Finnish		
	M	SE	t	M	SE	t	M	SE	t
Intercept	5.41	0.05	107.77***	5.32	0.032	166.57***	5.28	0.03	179.75***
Average Frequency				-0.02	0.00	-6.08***	-0.01	0.00	-3.48***
Number of Words	0.00	0.00	2.19*						
Average Number of Radicals	0.04	0.02	1.89[+]						

[+] p <.10, * p < .05, ** p < .01, *** p <.001

We can see that for this measure the predictor of Number of Words did not contribute to any of the three languages. This is unsurprising since we would not expect average saccade extent to be related to the number of words in a sentence (though Shu, Zhou, Yan & Kliegl, 2011 show that such a relationship can occur for normal unspaced Chinese text). Focusing on the alphabetic languages, in Finnish we see a main effect of Average Word Frequency and Average Word Length with high frequency words leading to longer rightward saccades, presumably due to easier lexical identification and increased word skipping, and also that longer words caused, on average, longer rightward saccade lengths. This latter effect reflects standard saccadic targeting effects during reading.

For English, although we see a similar model in several respects, the main effect of Average Word Frequency is qualified by an interaction with Average Word Length such that the effect of frequency was more pronounced for long than short words. We suspect that this effect is driven by the refixation behavior of the English readers. Long infrequent words are likely to receive two fixations in English, thereby reducing the saccade length on average. Short, and long high frequency words are comparatively less likely to receive a refixation, resulting in less reduction in the average saccade length. This effect is unlikely to occur for Finnish since average word length is greater than for English (see Table 1), and consequently, Finnish words are far more likely to receive a refixation by default.

A considerably different model appears for Chinese. Here, both the main effects of Word Frequency and Word Length are qualified by an interaction between the two with a negative fixed effect estimate indicating reduced average rightward saccades when the words are long and high frequency. This surprising finding will need to be confirmed in experiments but it could be due to low frequency words being

more constrained by their preceding context, thereby inducing slightly more skipping as a consequence of increased predictability. In line with this suggestion, recent work has shown that initial character frequency can constrain the extent to which later characters in a Chinese word are parafoveally processed (see Cui, Yan, Bai, Hyönä, Wang & Liversedge, 2013). We also observed an effect of number of radicals such that an increase in the number of radicals was accompanied by a decrease in average saccade length. The number of radicals is correlated with the number of strokes in a sentence and as such this could reflect the visual density of the text. If this effect is related to visual density, then it suggests that Chinese readers' saccadic targeting is based on the visual density of the parafoveal text (shorter saccades for more dense text).

The bilingual model for the alphabetic languages (see Table 9) shows a main effect of Frequency with slightly longer saccade sizes observed in sentences with higher frequency words. There is also a main effect of word length that is qualified by an interaction (see Figure 7) such that an increase in word length leads to an increase in saccade length and this to a lesser extent in Finnish compared to English. This is not too surprising; a one character difference between 14 and a 15 letter word (which will feature more often in Finnish) will have a reduced influence relative to a character difference between a 6 and a 7 letter word.
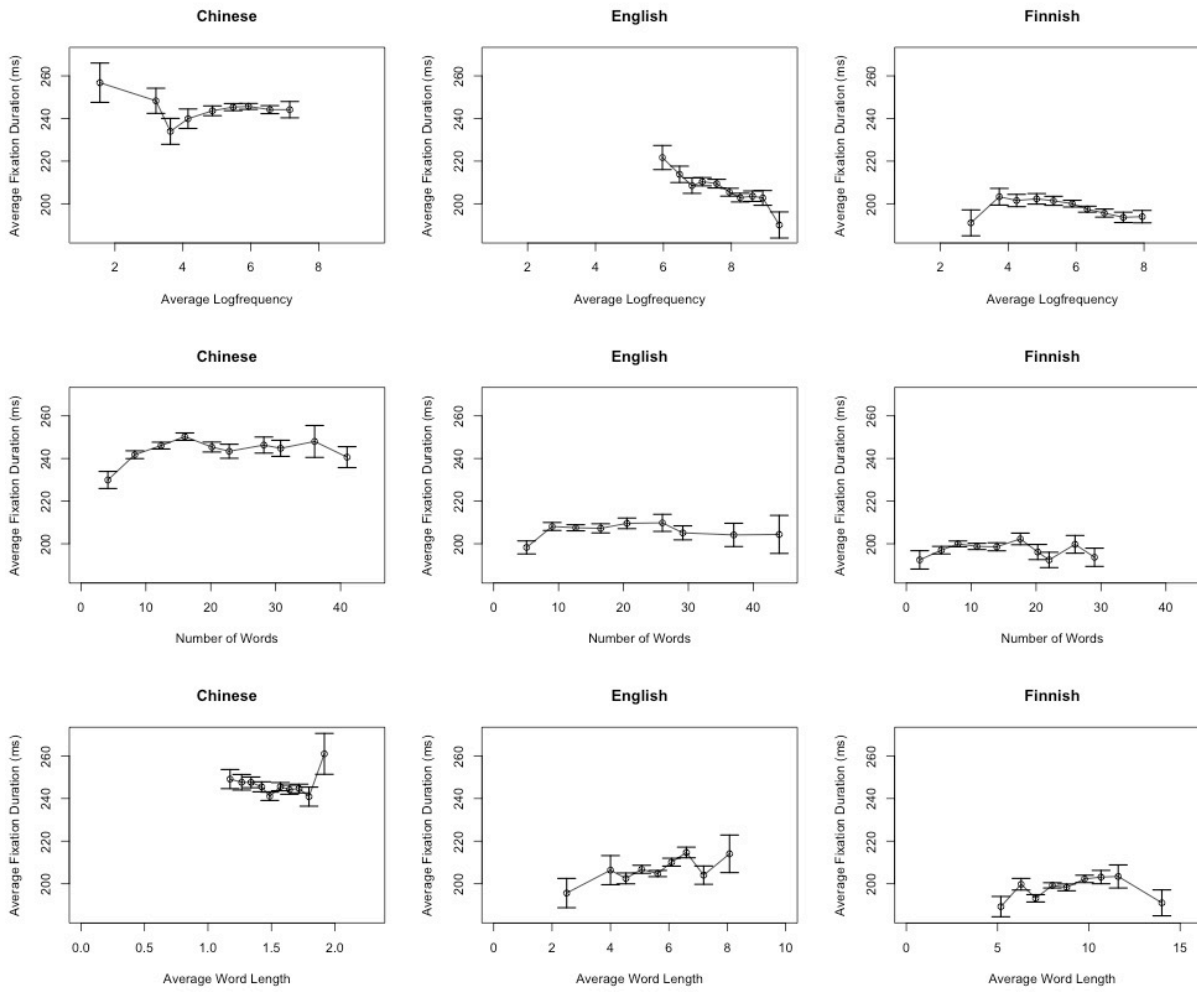
Fig. 8. Mean Fixation Duration (ms) as a function of Average Logfrequency (Top row), Number of Words (Middle Row) and Average Word Length (Bottom Row) for Chinese (Left Column), English (Middle Column) and Finnish (Right Column). For each predictor, fixations were binned into categories containing minimally 10% of the data. Error bars are standard errors of the mean within-subject variance.

3.5 Average Fixation Duration

The average fixation duration as a function of the fixed effects are presented in Figure 8. For the data on average fixation duration for which the means are presented in Table 2, and the fixed effect estimates in Table 10, we can see that the main effect of word frequency is the only significant effect for the alphabetic languages with a slightly shorter average fixation duration in sentences with comparatively high

48

frequency words. For Chinese quite a different model appears with a main effect of Number of Words and Average Number of Radicals such that sentences with more words and radicals received longer fixation durations. The average number of radicals could be related to visual density with a cost associated with processing characters that contain a lot of radicals (and therefore also strokes). The effect of number of words is more surprising, however it is important to note the very small effect size. Indeed, this factor could also be related to visual density with (slightly) higher levels of visual crowding in sentences with a high number of words. Again, this is a finding that needs to be examined more closely, preferably in an experiment allowing tight control over these factors. The bilingual model (see Table 11) again shows a main effect of frequency but it is qualified by an interaction with Language such that the effect is less pronounced for Finnish (see Fig 9). This could be simply due to the fact that in Finnish there are slightly more fixations in a sentence distributing the processing load over more fixations.

Table 11. Fixed Effect Estimates for Average Fixation Duration. Bilingual Model (English – Finnish).

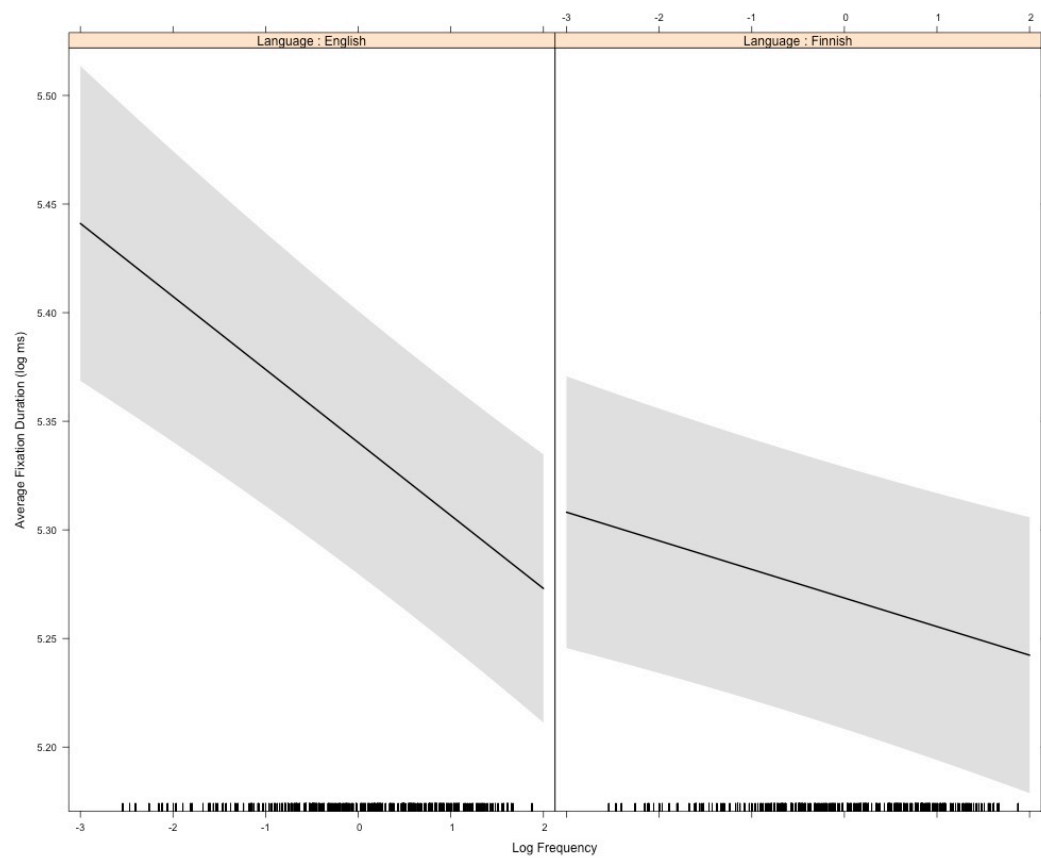|	Bilingual Model English - Finnish		
	M	SE	t
Intercept	5.30	0.02	241.28***
Language	-0.07	0.04	-1.66[+]
Average Frequency	-0.02	0.00	-6.84***
Language x Average Frequency	0.02	0.01	3.00**

+ p <.10, * p < .05, ** p < .01, *** p <.001

Fig. 9. Effect display for the significant interaction of Language and Log Frequency in the Bilingual LMM for the Average Fixation Duration. A 95-percent confidence interval (the grey shaded region) is drawn around the estimated effect.

4. General Discussion

Let us start by considering the basic descriptive data from the texts that were translated from English into Chinese and Finnish. Recall that the translation process was carried out quite painstakingly to ensure that sentential constituents were maximally comparable across languages. Even so, due to quite marked differences across languages, it was still the case that there were far from word-to-word correspondences. Most strikingly, the number of words in the Finnish stimuli was substantially below the number of words in the Chinese and English stimuli, which themselves contained approximately the same number of words (this held for the total number of words in all the texts, as well as for the number of words per sentence). The reduced number of words in Finnish is due to the fact that the language does not contain as many function words as English, and that it is agglutinative. Often, multiple words in English or Chinese correspond to single morphologically complex words in Finnish. It might appear somewhat more surprising that there was approximately the same number of words in the Chinese and English stimuli, and this in turn might suggest a potential one-to-one relationship between the words in these texts. However, this was far from the case. As mentioned earlier, articles do not appear in Chinese but appear quite frequently in English and a variety of characters that represent function words in Chinese do not appear in English. Additionally, it was the case that often single words in one language required multi-word counterparts in the other language in order to achieve translational equivalence. Overall these differences counteracted each other, resulting in Chinese and English texts that had very comparable numbers of words.

In accord with the differences in the number of words in the stimuli, we found that words were substantially longer in Finnish than in English, again, largely due to

agglutination and the reduced number of function words. Note that whilst we were able to directly compare word lengths in the two alphabetic languages, the metric of comparison differs between Chinese and the alphabetic languages, thereby ruling out direct comparison in the same way. Nonetheless, Chinese words were shorter than words in the alphabetic languages. Finally, the finding that the average frequency of words in English was higher than in Finnish and Chinese, for which average word frequencies were approximately the same, is due to the prevalence of determiners (*a* and *the*) and prepositions (*in*, *on*, etc.) in English, that do not exist in Chinese and Finnish, and for which frequency counts are very high. Frequency counts for these words in English weighted the mean frequency value upwards. Therefore we used standardized log frequency measures for the three languages.

Next let us consider the global descriptive eye movement data for the three languages. In contrast with the findings of Sun et al. (1985) and Feng et al. (2009), it was clearly the case that there were robust differences in eye movement behavior between the different languages. Briefly, Finnish readers made more, shorter fixations than Chinese readers who made fewer but longer fixations. The fixation number and duration data from the English readers were intermediate to the Chinese and Finnish data. Furthermore, saccade extent varied substantially between languages contingent on the length of the words in the language; saccades were shortest for Chinese (the language with the shortest words), longer for English (words longer than Chinese, but shorter than Finnish) and longest for Finnish.

The question arises as to why we found differences in eye movement behavior between Chinese and English readers (and of course Finnish readers too), whilst Sun et al. (1985) and Feng et al. (2009) did not. There are several likely reasons for this. First, it is very difficult to directly compare the results reported by Sun et al. with the

present results because in their analyses they applied a scaling factor assuming that three Chinese words was equivalent to two English words. The application of this scaling factor, of course, has the effect of reducing any differences between the languages. In relation to the results reported by Feng et al., there are two critical differences between those and the current findings. First, Feng et al. found that saccades were shorter in English than in Chinese, whilst contrastingly, we found that (rightward) saccades were longer in English that in Chinese (and indeed, longer in Finnish than in English). These contradictory findings can likely be explained by the fact that Feng et al. approximately equated the size of the Chinese and English text passages in terms of the space that they covered on the computer screen. The consequence of this would have been to make the size of the Chinese characters very large in comparison with the size of the letters comprising the English words in their stimuli. As Morrison and Rayner (1981) established, saccade extents during reading are best indexed in relation to word units rather than the physical size of text (the amount of letters that cover a degree of visual angle), and for this reason, given the physical characteristics of the Feng et al. stimuli, we might reasonably expect longer saccades for their Chinese than their English texts. The second direct inconsistency between the current results and those of Feng et al. is that we found that average fixation durations were longer in Chinese than in English (and indeed that fixation durations were longer in English than in Finnish), whereas Feng et al. found no reliable differences in average fixation durations between Chinese and English readers. This discrepancy cannot be attributed to physical differences in the size of the Chinese and English text stimuli used by Feng et al. However, there is another important aspect of the Feng et al. stimuli that might explain this difference. In their study the great majority (86%) of the text used to examine Chinese reading differed in

54

content to the text used to examine English. That is to say, the texts were not direct translations of each other. In contrast, the stimuli used in the present study were very carefully translated to minimise content differences and maximise comparability of the stimuli across languages. Critically, the semantic content of text is known to directly influence the duration of fixations during reading (Rayner, Warren, Juhasz & Liversedge, 2004), and therefore, similarity in fixation durations for Chinese and English readers in the Feng et al. study could have arisen due to content differences across languages. In contrast, when differences in semantic content are minimised across stimuli, as was the case in the present study, then language specific differences in fixation durations emerge.

If we consider the descriptive results for the linguistic stimuli and those for the eye movement behavior together, we have the most striking, and arguably important, result of the whole study. It is clear that despite substantive differences in both the eye movement behavior and the written forms of the linguistic stimuli, overall there was no difference in the total reading times for the texts. Note, similarly, that there were no differences in comprehension levels across the different languages. It took the Chinese readers the same amount of time to read the texts as the English readers, who themselves took the same amount of time to read the texts as the Finnish readers. In other words, despite variability both in the written form of the linguistic stimuli and variability in the characteristics of the reading behavior of the participants across languages, the time taken to encode and construct a representation of the meaning of the information conveyed in the text did not differ for Chinese, Finnish and English readers. Note that the level of expertise in reading was also comparable across participant groups. This finding of parity across languages is consistent with the suggestion that there is a basic limit to the rate at which humans are able to assimilate

fundamental aspects of meaning from written text, and that this is not determined by the particular linguistic form in which the information is conveyed. To be clear, whilst languages may dramatically differ in terms of their phonological, morphological and orthographic structure, they are quite similar in terms of their propositional content. Consequently, regardless of the nature of the written form of a language, the convergence on a similar semantic representation for a sentence is essentially similar across languages. Tentatively, this finding of cross-cultural consistency might, potentially, be indicative of a universal aspect of human information processing.

Let us next turn our discussion to the statistical models that we constructed for the four different global measures of eye movement behavior during reading (total sentence reading time, average fixation duration, number of fixations and rightward saccade length). To briefly recapitulate, we constructed linear mixed models for each language separately to investigate the extent to which three key variables, word frequency, number of words, and word length captured variance in these measures. Furthermore, if the resulting models for Finnish and English were structured similarly, then we undertook analyses to build a combined model for the two languages. The most parsimonious models were always adopted, and our overall goal was to establish whether the models that resulted from this process were comparable in structure, and if so, which variables (or interactions between variables) consistently accounted for variance in the eye movement data. Our assumption was that any such variables might represent cross-linguistic commonality in cognitive representation and process consistent with the suggestion of universality in written language comprehension.

For the total sentence reading time we obtained very clear results. In the model that we constructed for the individual languages, there were effects of word

frequency, number of words and word length (though this effect was small for Chinese). These effects were exactly as we would have expected with shorter reading times for sentences comprised of fewer, shorter words of higher frequency. The model for Chinese also included a small interaction term reflecting an effect of the number of words and frequency, such that the integration of additional words produced longer increased reading times when the average frequency of the words in the sentence was low compared with high. We offered two explanations for this effect; first, reduced total times for sentences with lots of high frequency words could (partially) be due to increased prevalence of word skipping; second, the interpretation of a word with respect to preceding sentential context is more difficult when the sentential context is comprised of low frequency words (that are more difficult to process than high frequency words that are easier to process). The models for the different languages were extremely comparable, and the variables that we included in the models appear to have a very similar influence in each language. Furthermore, although we were unable to construct a statistical model for all three languages together (due to the different metric of word length for Chinese relative to the alphabetic languages), for the combined model for the alphabetic languages, Finnish and English, there was only a very marginal effect of language, and this was qualified by an interaction with the number of words in a sentence, such that an additional word in a Finnish sentence produced a greater cost than in an English sentence. This effect fits very neatly with the fact that Finnish is an agglutinative language in which the words are much longer than English, as can be seen from the descriptive data derived from our stimuli. It is quite unsurprising that an additional word in Finnish will result in a greater cost in total sentence reading time relative to English. Finally, a further interaction in the combined Finnish-English model requires discussion, namely, the

joint effect of frequency and word length such that frequency influences were reduced for sentences with words of increased length. This effect is likely due to a reduced probability of high frequency word skipping when those words are longer. In sum, for the total sentence reading times, we obtained statistical models that were very similar for Chinese, Finnish and English, and whose structure indicated comparable influences of word frequency, word length and the number of words in the sentence. Clearly, these variables account for variance in this measure across languages to a similar degree. To the extent that total sentence reading time reflected the time taken to comprehend text that was quite comparable in content across the three languages, then our findings are suggestive of commonality in relation to the influence of these variables in each language. To be clear, our findings are consistent with the suggestion that word frequency, word length and the number of words in a sentence are variables that capture common aspects of representation and processing across languages, and it appears that they do this to quite a similar degree.

We will next consider the statistical models we constructed for the total number of fixations measure. As with the total sentence reading time measure, the statistical models for each of the languages for the total number of fixations included robust effects of word frequency, number of words and word length. Also, as for Chinese in the total time analyses, we obtained a reliable interactive effect of frequency and number of words, but this time the effect occurred for all three languages. Total number of fixations was increased when sentences included more low frequency words. As above, our explanation for this is twofold, arising due to increased word skipping when sentences contained more high frequency words, and greater difficulty when incrementally interpreting additional words in relation to preceding sentential context comprised of low frequency words. Note that this

interaction was observed in all three languages for total number of fixations (unlike the sentence reading times). The reason why this effect is clearly present in all 3 languages could lie in the fact that this measure is more sensitive to effects of word skipping that are associated with high-frequency words. That is, for total sentence reading times there will be a trade-off such that increased word skipping will be accompanied by slightly longer fixation durations (for instance due to reduced visual acuity when saccades are launched from more distant launch sites). In the discrete measure of total number of fixations, this trade-off will not arise, hence the effect of an increased number of high frequency words will show up more clearly.

Two things are clear from the individual models; first, the models for total number of fixations were very similar to the models for total sentence reading time. In some sense, this is to be expected, but nevertheless, it is reassuring that the consistency is there. Second, as with the models for total time, the consistency of the models across languages is very striking. The variables that we selected as likely factors accounting for variance in eye movements across languages with very different written forms seem to do this in a quite similar way. Once again, formalising this suggestion by combining the models for Finnish and English showed no effect of language, indicating that the models were indeed very comparable. The combined model also showed, once again, the interactive effect of the number of words and word frequency. To reiterate, the variables of word frequency, number of words and word length do a very good job of consistently capturing variance in eye movements across languages.

The third measure that we considered was rightward saccade extent. It is important to note here that this measure relates to *where* readers targeted their eyes, not *when* they decided to move them. The *where/when* distinction is critical in

relation to eye movement control (Findlay & Walker, 1999). In terms of conceptual

units over which linguistic processing operates, words as abstract descriptors relating

to the temporal aspects of cognitive processing, as we have already shown, relate

quite comparably across languages. However, given the stark differences in the

physical characteristics of the written form of Chinese, Finnish and English, it was to

be expected that there would be differences in the basic structure of the models for

saccade length, as well as the degree of comparability in relation to spatial aspects of

processing, across languages. Our finding that the number of words in a sentence did

not capture significant variance in average saccade extent is perfectly reasonable.

Increasing the number of words in a sentence would effectively result in increased

numbers of saccades and fixations, but would not modulate the actual magnitude of

those saccades. Differences in average saccade extent, whilst different across

languages, should remain fairly constant regardless of how many words comprise a

sentence. The models for rightward saccade extent in the alphabetic languages were

quite similar, showing effects of word frequency and word length (this being qualified

by an interaction between the two for English). These effects can be explained in

terms of patterns of saccadic targeting, word skipping and refixation behavior in the

two languages. In particular targeting saccades to the middle of words (Rayner, 1979)

likely produced the word length effect. Also, since high frequency words are easier to

process than low frequency words, more word skipping for high frequency words

would have occurred and saccades would have been longer generally. Finally,

increased refixations on long infrequent words in English could explain the interactive

effect obtained there. On the other hand, this would have occurred to a lesser degree

in Finnish because most words in this language are long anyway. The combined

model for Finnish and English produced a main effect of word frequency similar to

the separate models for each language, along with an interactive effect of word length with language. A comparable increase in the length of a word in Finnish had less of an effect on saccade amplitudes than in English, which given average word length differences in the two languages, again, was entirely to be expected.

Our findings for the Chinese model of rightward saccade extent were far less transparent, with interactive effects of frequency and length such that saccades were shorter in sentences containing many long high-frequency words. Our only explanation for this effect was tentative and centred on the idea that lower frequency characters may constrain the identity of counterpart characters within a word, and this in turn could have impacted on skipping behaviour. Current understanding of saccadic targeting in Chinese reading is limited. Some researchers have suggested that targeting of parafoveal Chinese words may be based on segmentation of the upcoming text (Yan, Kliegl, Richter, Nuthmann & Shu, 2010), others have argued that initial targeting may be algorithmically based, with refixations being made contingent on non-optimal landing positions (Li, Liu & Rayner, 2011), and still others have suggested that, in fact, a multitude of factors might contribute to the decision of where to move the eyes in the upcoming text in Chinese (Zang, Liang, Bai, Yan & Liversedge, 2012). Clearly, this is area of eye movement control in Chinese reading that still requires extensive research.

One effect that was unique to the Chinese data, and for which there may be a ready explanation, was that of the shorter saccades for sentences including more radicals. This effect is almost certainly an effect of the visual density of the text; sentences with more radicals will also have increased numbers of strokes. A recent study by Liversedge et al., (2014) demonstrated an influence of the visual complexity of Chinese text based on the number of strokes in a character on word skipping, and

this effect occurred whilst the parafoveal extent of the stimulus was held constant. Thus, there were good, *a priori*, grounds to assume that effects of the visual density of characters would affect saccade extents during Chinese reading.

Overall, it is fair to say that the statistical models for rightward saccade extent for each of the languages, as well as across languages are the least consistent and involve the least straightforward patterns of effects. We believe that this is a direct consequence of the saccade extent measure being an index of spatial aspects of eye movements during reading, as contrasted with each of the other measures that quantify temporal aspects of eye movements. We will expand on this issue below. Put simply, there appears to be reduced common variance across in relation to the variables that comprise our statistical models within the rightward saccade extent data, and it is very likely that this is because we selected the languages for this study to maximise those differences. This said, in our view, the saccade extent data are still meaningful and interpretable.

Finally, we must consider the average fixation duration data. For Finnish and English, there were exclusively effects of word frequency, and in the bilingual model these interacted with language such that the effect was less pronounced in Finnish than English. We argued above that this effect arose because words are much longer in Finnish than English, and therefore, effects are distributed across multiple fixations of shorter duration (often on the same word) to a greater degree. The effects in the Chinese model were quite different, with number of words and number of radicals producing longer fixation durations, and we argued that this arose due to increased density of Chinese text in sentences with more words containing more radicals (since number of strokes in a word is correlated with number of radicals).

Having considered each of the aspects of the results individually, it is pertinent to now take something of a meta-perspective. Does the current study provide evidence for universality in processing? This was the central question we set out to address in this study. To our minds, the degree of similarity in the statistical models was compelling, and suggests strongly that the number of words, the average word frequency and word length all impact similarly on the total time they take to read a sentence in Chinese, Finnish and English. As we have argued, saccadic eye movement behaviour underlies normal reading in all cultures, and eye movements are known to reflect the moment-to-moment cognitive processes that occur as we read. In our view, the present study demonstrates compellingly that these variables reflect representations and aspects of processing that are common across three languages with very different written forms. However, we are somewhat cautious in making the claim that this is a demonstration of universality in written language processing. We fully acknowledge that we have demonstrated such effects only across three languages, and therefore, at best, our effects can be considered to be consistent with notions of universality.

It is useful to consider what our effects might mean in terms of the psychological processes underlying language processing. As a reminder, the three variables that we included in our statistical models were average frequency and length of the words in each sentence, along with the number of words in each sentence. The first two of these variables have been demonstrated to index the ease with which words are lexically identified during reading, whilst the number of words quantifies the cumulative processing cost associated with lexical identification (and subsequent linguistic processing) of all of the words comprising a sentence. Thus, for the statistical models of measures that showed the most consistency, that is, total sentence

reading time and total number of fixations, our results demonstrate (1) that word representations are central to language processing in each of the languages we investigated; and (2) that the nature of processing occurs similarly in relation to two key lexical characteristics in each of the languages, namely word length and word frequency. Put simply, word identification is common across languages and is a key aspect of processing in reading across languages, and further, word length and word frequency both similarly constrain this process in Chinese, Finnish and English. In relation to these aspects of written text comprehension we believe our findings are consistent with universality of representation and process during reading.

Why did we not observe a similar degree of consistency across models for the average fixation duration and forward saccade extent data? Here the statistical models for each language were often quite different, markedly so for Chinese relative to the alphabetic languages. Of particular importance here is that fact that these two measures are both qualitatively different to total sentence reading time and total number of fixations measures. The average fixation duration is a measure of mean performance across all of the fixations made as a sentence is read (as distinct from a total performance measure, as in total number of fixations or total time). Individual fixations during reading are ordinarily considered, primarily, in relation to local, word based, reading measures (e.g. first fixation duration, single fixation duration, gaze duration, etc.). It is in these more sensitive, fine grain measures that we often observe the most immediate effects of linguistic variables that occur on a fixation-by-fixation basis, rather than in global measures of reading performance. And since our unit of analysis in this study was the sentence (a quite global region of analysis), it follows that more variability would occur across languages for average fixation duration

relative to total number of fixations and total sentence reading time. This is exactly what we observed.

As we have already pointed out, the forward saccade extent measure, for which we saw the least consistency across models, is a measure of spatial, not temporal characteristics of eye movement behavior (unlike all the other three measures). Temporal measures of eye movements are ordinarily taken to reflect information processing aspects of comprehension, that is, linguistic processing, whereas, spatial measures are ordinarily associated with visual aspects of processing (where to position the eyes to allow for optimal uptake of visual information necessary for comprehension). Overall, our results can be argued to indicate that while aspects of linguistic processing might exert a common influence on eye movements during reading across languages, aspects of visual processing, particularly those associated with saccadic targeting, may not exert a similarly common cross-linguistic influence. For example, it may not be reasonable to suggest that saccadic targeting is as fundamentally word based in Finnish or Chinese as it is in English. Finnish contains numerous very long morphologically complex words, within which multiple saccades will often be made (e.g., Bertram et al., 2004). It seems reasonable to suggest that saccadic targeting in this situation may well be based on sub-word units of information (e.g., morphemes), rather than whole words (Hyönä & Pollatsek, 1998). Furthermore, Chinese has an unspaced, character based format and an abundance of word boundary ambiguities. Again, a default assumption that saccadic targeting would be word based in this language could be erroneous. Given these marked differences in the written forms of the languages, the commonality that we observed for word based linguistic processing will obviously not be there across

languages to the same extent in relation to spatial aspects of eye movement behaviour, that is, saccadic targeting.

At this point, we should return and consider the relationship that exists between the visual and linguistic processing systems. As we have argued, the extent to which a reader can linguistically process the words within a sentence is contingent upon the rate at which visual representations of those words on the page are delivered. Taking the counter-side of this relationship, it is also the case that the rate at which linguistic processing proceeds may constrain the rate at which the reader moves their eyes to advance through the text. In this sense, it is also possible that the linguistic processing system has the potential to act as something of a processing bottle-neck. It is certainly the case that the visual and linguistic processing systems have to operate in close synchrony during reading, with the visual system working to deliver visual information in an effective and timely fashion, whilst the linguistic processing system attempts to incrementally interpret text, attaining comprehension rapidly on a fixation by fixation basis. The relationship between the visual and linguistic processing systems can be considered to be an antagonistic one, with each system having the potential to constrain the other. The nature of this antagonistic relationship is very relevant to the present results. Specifically, the visual and informational density of the language being read will be critical to the balance that exists in the relationship between the two processing systems. In comparatively sparse languages such as Finnish, the visual system must deliver representations of the text very frequently and across a substantially spatially extended horizontal range in order to furnish the language processing system with the information it requires at a rate that permits comprehension to occur effectively and efficiently. To characterise this in simpler terms, in Finnish reading, the visual system must work very assiduously, pushing

ahead in the text to acquire and deliver new visual information rapidly to meet the demands of the linguistic processing system for which processing demands are relatively light due to the paucity of linguistic information available during any particular fixation. In this sense, the balance in the relationship between the two systems is such that processing load is relatively light for the language processor, but quite heavy for the visual processing system, and in this situation the rate of delivery of the visual information acts as a stricture on the rate at which linguistic processing can proceed. The behavioural consequence of this state of affairs is that the reader makes more, shorter fixations, along with longer saccades.

In contrast, for a comparatively dense language such as Chinese, the balance lies in the opposite direction. Here, during each fixation, the visual system delivers a very substantial amount of information upon which the linguistic processing system may operate, and because this information is densely packed across a limited horizontal spatial extent, only limited saccadic movements are necessary for the efficient delivery of such information. Furthermore, in this situation, the linguistic processing demands are high. With each new fixation a substantial amount of linguistic information arrives, and the language processing system must endeavour to process this as quickly as possible. In this situation, the demands on the visual processing system are relatively low, whilst the demands on the linguistic processing system are comparatively high. Thus, during Chinese reading, the rate at which the linguistic processing system operates acts as a constraint on the rate at which the visual processing system can proceed through the text. Once again, there is a behavioural consequence of this balance in the relationship between the two systems, namely that Chinese readers make fewer, but longer fixations and short saccades. Of course, the balance is more even for English, which has an intermediate density

relative to Chinese and Finnish. Both the global descriptive analyses of our data, as well as the statistical models that we constructed for the different reading measures, reflect patterns of eye movement behavior that are entirely consistent with this account. It should also be clear that only through the simultaneous consideration of reading behaviour for Chinese, Finnish and English texts, each translated to be maximally comparable, is it possible to observe so directly how the balance in the relationship between the visual and linguistic processing systems shifts in the different languages.

To summarise, we undertook an investigation of eye movement behaviour during reading in Chinese, Finnish and English. These languages differ markedly in relation to their linguistic characteristics and their written form. The texts we used in our experiments were translated very carefully across languages to ensure that they were maximally comparable, and we computed our reading time measures across sentences, again, to ensure comparability. We computed total sentence reading time, total number of fixations, average fixation duration and forward saccade extent. We adopted a LMM approach and constructed statistical models of our data to evaluate the extent to which word length, word frequency and the number of words in a sentence accounted for common variance. For total sentence time and number of fixations, we found evidence that lexical representations corresponding to words are important, and that word length and frequency are key indices of lexical processing in all of the languages. We argue that these effects are consistent with the notion of universal aspects of representation and process during reading. We also argued more generally that the data provide insight into how the nature of the written form of a language determines the balance in the relationship that exists between the visual and linguistic processing systems during reading.

References

Bai, X., Yan, G., Liversedge, S.P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1277-1287.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17, 364-390.

Bertram, R. & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language,* 48, 615 – 634.

Bertram, R., Pollatsek, A., & Hyönä, J. (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language, 51,* 325-345.

Blythe, H.I., Liang, F., Zang, C., Wang, J., Yan, G., Bai, X., & Liversedge, S.P. (2012). Inserting spaces into Chinese text helps readers to learn new words: An eye movement study. *Journal of Memory and Language, 67,* 241-254.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.

Clifton, C., Ferreira, F., Henderson, J.M., Inhoff, A., Liversedge, S.P., Reichle, E.D., & Schotter, E.R., (2015). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. Journal of Memory and Language, in press.

Coltheart, M., & Crain, S. (2012). Are there universals of reading? We don't believe so. Invited commentary on "Towards a universal model of reading". *Behavioral and Brain Sciences, 35,* 20-21.

Contemporary Chinese Dictionary. (2005). Beijing: Commercial Publishing House.

Dehaene, S. (2009). *Reading in the brain: The science and evolution of a human invention*. New York: Penguin, Viking Adult.

Durgunoglu, A.Y., (2006). How the language's characteristics influence Turkish literacy development. In M. Joshi & P.G. Aaron (Eds), *Handbook of orthography and literacy* (pp. 219-230). Mahwah, NJ: Erlbaum Associates.

Ehrlich, S.F., & Rayner, K., (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 6410, o.

Ellis, N.C., (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, N.C., & Hooper, A.M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics, 22,* 571-599.

Feldman, L.B. & Moscoso del Prado Martin, F. (2012). Does a focus on universals represent a new trend in word recognition? *Behavioral and Brain Science. 35,* 285-285.

Feng, G., Miller, K.F., Shu, H., & Zhang, H. (2009). Orthography and the development of reading processes: An eye movement study of Chinese and English. *Child Development, 80,* 720-735.

Findlay, J.M., & Gilchrist, I.D. (2003). *Active vision: The psychology of looking and seeing*. Oxford, UK: Oxford University Press.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Science*s, *35,* 263-279.

Gelb, I.J. (1952). *A study of writing: The foundations of grammatology.* Chicago: The University of Chicago Press.

Gilchrist, I.D., Brown, V., & Findlay, J.M. (1997). Saccades without eye movements. *Nature, 390*, 130-131.

Gregory R.L. (1966). *Eye and brain: The psychology of seeing.* London: Weidenfeld & Nicolson.

Hanley, J.R., Masterson, J., Spencer, L. & Evans, D. (2004). How long do the advantages of learning to read a transparent orthography last? An investigation of the reading skills and reading impairment of Welsh children at 10-years of age. *The Quarterly Journal of Experimental Psychology, 57,* 1393-1410.

Huey, E. B. (1900). On the psychology and physiology of reading. I. *American Journal of Psychology, 11*, 283-302.

Hyönä, J., & Pollatsek, A. (1998). The role of component morphemes on eye fixations when reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1612-1627.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612-624.

Javal, L. É. (1878, 1879). Essai sur la physiologie de la lecture (in several parts). Annales d'Oculistique, *79*, 97, 240; *80*, 135; *81*, 61; *82*, 72, 159, 242.

Just, M. A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87,* 329-354.

Li. X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General*, 143, 895-913.

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research, 51,* 1146-1156.

Liversedge, S.P., Blythe, H.I,. & Drieghe, D. (2012). Beyond isolated word recognition. *Behavioral and Brain Sciences, 35,* 31-32.

Liversedge, S.P., & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4,* 6-14.

Liversedge, S.P., Rayner, K., White, S.J., Vergilino-Perez, D., Findlay, J.M., & Kentridge, R. (2004). Eye movements when reading disappearing text: Is there a gap effect in reading? *Vision Research, 44,* 1013-1024.

Liversedge, S.P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition*, 22, 441 - 457.

McConkie, G.W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17,* 578-586.

Morrison, R.E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics, 30,* 395-396.

Perfetti, C.A., & Harris, L.N. (2013). Universal reading processes are modulated by language and writing system. *Language Learning and Development, 9*, 296-316.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. Cognitive Psychology, 7, 65-81.

 Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner K. (2009). The thirty-fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Rayner, K., & Liversedge, S.P. (2011). Linguistic and cognitive influences on eye movements during reading. In S.P.Liversedge, I.Gilchrist & S.Everling. (Eds.), *The Oxford handbook of eye movements* (pp. 751-766). Oxford University Press.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1188-1200.

Reichle, E.D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445-476.

Reichle, E.D., & Drieghe, D. (2013). Using E-Z Reader to examine word skipping during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39,* 1311-1320.

Sainio, M., Hyönä, J., Bingushi, K., & Bertram, R. (2007). The role of interword spacing in reading Japanese: An eye movement study. *Vision Research, 47,* 2575-2584.

Seidenberg, M.S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, J. Ren & O. Tzeng, (Eds.), *Dyslexia across languages: Orthography and the gene-brain-behavior link* (pp.151-174). Baltimore, MD: Paul Brooke Publishing.

Shen, E. (1927). An analysis of eye movement in the reading of Chinese. *Journal of Experimental Psychology, 10,* 158-183.

Shen, D., Liversedge, S.P., Tian, J., Zang, C., Cui, L., Bai, X., Yan, G., & Rayner, K. (2012). Eye movements of second language learners when reading spaced and unspaced Chinese text. *Journal of Experimental Psychology Applied, 18,* 192-202.

Shu, H., Zhou, W., Yan, M., & Kliegl, R., (2011). Font size modulates saccade-target selection in Chinese reading. *Attention, Perception & Psychophysics*, 73, 482-490.

Spencer, L. & Hanley, J.R. (2003). The effects of orthographic consistency on reading development and phonological awareness: Evidence from children learning to read in Wales. *British Journal of Psychology, 94,* 1-28.

Sun, F., Morita, M., & Stark, L.W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics, 37,* 502-506.

Tinker, M. A. (1936a). Reliability and validity of eye-movement measures of reading. *Journal of Experimental Psychology, 19,* 732-746.

Tinker, M.A. (1936b). Eye movements in reading. *The Journal of Educational Research, 30*, 241-277.

Wang, H-C., Schotter, E.R., Angele, B., Yang, J., Simovici, D., Pomplun, M. & Rayner, L. (2013). Using singular value decomposition to investigate degraded Chinese character recognition: Evidence from eye movements during reading. *Journal of Research in Reading*, *36*, 35-50.

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2012). Lexical and sublexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning Memory and Cognition, 38,* 1069-1075.

Yan, M., Kliegl, R., Richter, E., Nuthmann, A., & Shu, H. (2010). Flexible saccade target selection in Chinese reading. *The Quarterly Journal of Experimental Psychology, 63,* 705-725.

Yang, H., & McConkie, G.W. (1994). Eye movement control in Chinese reading. *Bulletin of the National TaiNan Teacher's College, 29,* 193-229.

Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S.P. (2013). Inter-word spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception & Performance, 39,* 720-734.

Ziegler, J.C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychology Science, 21,* 551-559.

Footnotes

1. The full set of stimuli can be obtained by emailing the first author.

2. Throughout this paper we will use the terms *bilingual* and *trilingual* to refer, quite appropriately, to experimental aspects (data, analyses and statistical models) respectively pertaining to two or three of the languages we chose to investigate. We acknowledge that these terms are most frequently used to refer to individuals who speak two or three languages, but to be clear, this issue was not the focus of the present study.

Acknowledgements