beginning and the middle of the word, rather than an absolute value that is independent of word length. The reason behind leftward bias is still an open question. However, there is evidence to indicate that certain parts of the retina get trained during reading, and this training kicks into effect when the eye fixates on a word, but not on a non-word (T. A. Nazir, Ben-Boutayab, Decoppet, Deutsch, & Frost, 2004).

**Reading Models**

One of the key questions that reading models deal with is the question of word processing. Specifically, are words being processed sequentially one after the other, or can several words be processed at the same time? This would be what the literature refers to as serial vs. parallel processing, the most well known examples being the E-Z Reader and the SWIFT reading models. Another key question is: what drives eye movement? As briefly mentioned earlier, there are several competing theories for how to account for the empirical data that has been recorded in the various studies. These models differ on these issues: the eyes being driven by lexical processing (whether serial or parallel) or via oculomotor constraints (Rayner, 2009). In other words, the different reading models are trying to answer the question: What is driving eye movement (cognitive processes or the oculomotor system), and what is the nature of this drive (serial or parallel)?

One might wonder at how it is possible that the oculomotor system would be the one to drive eye movements instead of the content of what is actually being read. Such models refer to studies that show that the pattern of eye movements remains similar when reading actual text as when reading letter strings or when searching for specific letters (Vitu et al., 1995). They argue that eye movement is carried out via predetermined strategies (Vitup et al., 1995). However, subsequent testing has shown that in reality, a visual search or reading a string of letters resulted in longer fixations, shorter saccades, and more likelihood of word skipping (Rayner & Fischer, 1996). Rayner and Fischer (1996) argue that the previous study was able to account for some aspects of eye movement but that there are a small part of a much larger set of cases to be accounted for.

Of the oculomotor control reading models, the Competition-Interaction model is based on the following assumptions as listed by Yang and McConkie (2001): There is a race between the programming and cancelling of saccades based on the signals that are being fed in. When the eye is fixated on a word, it is being inhibited from movement by a “central-attention-related fixate center” which is in turn being affected by higher-level processing. This center keeps the eye in place, thus influencing the timing of saccades. When that competition or race is lost, with the activation level of a word dropping below a certain level, the order for a saccade is given. The movement-inhibiting activation increases during the saccade so that by the time it lands on the target word, that inhibition is fully in place. The reader has learned a strategy for effective eye reading, and this includes the activation of a movement command after a certain time has lapsed in a fixation. The word being fixated needs to have a certain set of visual characteristics in order for the inhibition to recede. If it does not, then the inhibition will instead increase with time. Similarly, in case the higher level processes are facing difficulties, inhibition signals stop the programming of another saccade. However, the fixation durations vary most of the times due to “physiological processes (time to resolve competition, random waiting times) that have little relation to the current cognitive activity” (S. N. Yang & McConkie, 2001, p. 3584).

The E-Z Reader model for eye movement in reading is a cognitive-control SAS model, short for sequential attention shifts. It is able to account for fixation durations (how long or how short of time needed for specific kinds of words), the probabilities for skipping and refixations, global eye movement measures, as well as local processing (Rayner, 2009). Its main assumptions are (a) that the lexical processing of words is strictly serial i.e. one word after the other, and (b) that this is the main driver behind eye movement though the model does allow that low level “pre-attentive” processing in the visual field happens in parallel (Rayner, Reiche, et al., 2006). This would benefit from low-level visual cues such as word spaces and word lengths, and this extends to an area that is larger than the one where lexical information can be extracted; cited evidence for that includes the ability to program roughly accurate return sweeps (Pollatsek, Reiche, & Rayner, 2006).

According to the E-Z Reader (Fig. 5.4), word recognition starts with a visual analysis stage (V) which analyzes the visual characteristics of all the words in the perceptual span (Schoter et al., 2012). This is then followed by lexical processing which is completed in 2 stages9. The first stage (L1) is a rapid indexing of the familiarity of a word, while the second stage (L2) is the extraction of the relevant orthographic, phonological, and/or semantic code for linguistic processing. The completion of the first stage triggers the programming of a saccade. The completion of stage 2 triggers a shift in attention on to the next word (Rayner, Reiche, et al., 2006a). Attempt for this case, does not refer to “spatial orientation...” but to the process of integrating features that allows individual words to be identified” (Reiche et al., 2003, p. 453) which in effect decouples attention from eye movement. This shift in attention is when the partial word C_s_c_l_e_d from (Reiche et al., 2003) and when the familiarity check for word n+1 begins (Pollatsek, Juhaz, Reiche, Machacek, & Rayner, 2008). The latest version of E-Z Reader (v10) specifies this shift in attention as a distinct phase that is assumed to take around 20 ms to complete (Pollatsek et al., 2008).

As explained by Pollatsek et al. (2008), the E-Z Reader 10 also accounts for postlexical word integration into the sentence construct by adding a specific stage (L3) for that, right after L2 is completed. This integration of a word's meaning into that of the sentence roughly calculated to take around 50 ms. This stage might be completed with the eyes still on word n, or when they have already moved on to word n+1. The authors argue though that such a stage is included in this version of the model as a “placeholder” with which to test the effect of postlexical integration. They see this as a starting point for addressing the effect of comprehension difficulties on eye movement. Reiche, Warren, and McConnell (2009) again explain that
this stage is a "placeholder for a deeper theory of postlexical language processing during reading" and so is a "tentative" step that goes to account how postlexical processing affects eye movement in reading (Reichele et al., 2009, p.6).

It is important here to note that saccade latency, which is the time needed to program a saccade, takes on average a minimum of 175-200 ms (Rayner, Quinn, Croll, & Clifton, 1991), which means that the trigger for saccade programming happens very early on in a fixation (Reichele et al., 2003). The division of lexical processing into an early and a late stage seems to accommodate that observation. Note, though, that the word identification process is not done in 2 stages but 3: early stage of visual processing, then the 2 stages of lexical processing (Pollatsek et al., 2006).

An important aspect of the E-Z Reader is that it does not offer a deep analysis of higher level linguistic processing (Reichele et al., 2003), though the addition of stage I in version 10 seems to go in the direction of accommodating that factor (Pollatsek et al., 2006). As Reichele et al. (2003) explain, the influence of that level of processing on eye movement is seen in the cases when things go wrong in text comprehension and the reader then stop or regress to earlier sections of the text. The model, then, is more concerned in the instances when comprehension is proceeding smoothly and is at "default" setting. A resulting shortcoming is that it cannot account for interword regressions (Reichele et al., 2003), nor can it fully explain what happens with long distance regresions (Reichele et al., 2009).

Another important aspect of the E-Z Reader is how it accounts for the programming of saccades. These are done in 2 stages: a lagile stage (M1) where saccades can still be canceled, and a non-labile stage (M2) where saccades cannot be canceled. Saccades are assumed to take 25 ms to execute and are targeted towards the middle of the word. The actual fixation positions are subject to a systematic motor error with the specific pattern of overshooting the target on short saccades and undershooting on long ones (Pollatsek et al., 2006). As to where to fixate, the E-Z Reader projects that this depends on "linguistic, visual, and oculomotor factors" (Reichele et al., 2003, p. 459).

The SWIFT reading model is a mathematical model that aims to give a feasible "psychological and neuropsychological" account for the known pattern of eye movement (Engbert, Nuthmann, Richter, & Kliegl, 2005). Engbert, Nuthmann, Richter, and Kliegl (2005) outline the core principles of SWIFT in the following points:

- Several words are being processed in parallel, and in competition based on their different levels of activation.
- The decision of when and where to move the eyes are "decoupled."
- The programming of saccades is controlled by an independent timer that is "modulated by a focused inhibition process" which allows for more time to inspect difficult words. This inhibition includes a time delay (this point is a new addition in version II).
- This program takes place in 2 stages: first is the labile stage during which a saccade is being programmed and can still be cancelled. The second is the nonlabile stage at the point after which saccades cannot be canceled. This is similar to what E-Z Reader also proposes.
- The oculomotor system intrinsically produces both systematic and random errors.

Saccades are often mislocated, and in these cases a new saccade will be immediately programmed.

The time needed to program a saccade (saccade latency) depends on the intended length of the saccade to be programmed. Both of these take place at the end of the labile stage, and so, only the duration of the nonlabile stage is affected by the length of the upcoming saccade.

Overall, the authors sum up the process as such: Words are processed in parallel with different level of lexical activation. This helps to keep track of the state of word processing, and to select the next target and the decision of when to move the eyes. These 2 decisions are made via "separate pathways." Foveal Inhibition is able to delay saccade timing in case more time is needed for processing. The programming of saccades happens in 2 stages, and the target is selected at the end of the labile stage (Engbert et al., 2005). The selection of the target is random and based on the activation level the multiple words being processed (Laubrock, Kliegl, & Engbert, 2006). As to the activation level, its maximum level is governed by word frequency only, and word predictability is assumed to function only within the lexical preprocesing step (Laubrock et al., 2006).

The SWIFT reading model is proposed to be able to "reproduce a number of well-established measures of eye-movement control during reading, average fixation durations and fixation probabilities, distributions of within-word landing positions, and interword regressions. Finally, the SWIFT model can explain the IFP effect of fixation durations based on error correction of mislocated fixations." (Engbert et al., 2005, p. 805). However, SWIFT is still unable to account for the number of fixations as seen in the body of evidence collected in eye tracking (Richter, Engbert, & Kliegl, 2006). The latest version of this model, SWIFT, is now able to account for the pattern of regressions and relocations, though not on a quantitative measure (Nuthmann & Engbert, 2009).

It is quite interesting to compare the serial and parallel accounts of reading and eye movement, and it is quite possible that the processing of words goes in parallel for tasks that do not require high level cognitive processing (such as in a visual search). However, researchers argue that word processing is serial in the cases where one is reading for meaning (Reichele, Vanyukov, Laurent, & Warren, 2008).

It is not the point of this review to go very deep into the debate surrounding the various reading models. However, given all the evidence reviewed in the previous sections regarding the effects of language on reading metrics, it is difficult to accept a reading model that does not account for these findings. When reading for comprehension, one would expect that the "comprehending" of the text is what is driving the eyes forward through the text.

A serial account of word processing seems more logical and reasonable, and in tune with how one understands the spoken language, or how one formulates thoughts. It is interesting that there might be a good chance that Chinese words are processed in parallel rather than serially (Ishodf & Wu, 2005), but that could very well be due to the absence of word spacing and so the parallel processing is that of different word components and not words themselves. In such a case, this is not too far from the attentional beam that is discussed in the E-Z Reader. In either case, the effects of language and script are very much seen and felt in eye movement measures and mechanics. It is then interesting for such research to be developed for a more varied and international approach to reading studies.
Effects of the Stimulus

Reading is an activity involving two parties: the reader and that which is being read. The previous sections have looked at the various effects of the reader's age and reading skills level as well as the linguistic characteristics of the text being read. The next sections now turn to address the effects of the visual characteristics of the reading material, including aspects such as its size, the medium on which it is displayed, the patterns and features involved, and the distance from the reader.

Manipulation of the Visual Characteristics of Text

In order to get a better understanding of the reading process and what sort of information is being processed at the various stages of a fixation, researchers often manipulate specific characteristics of the text. These studies are quite instrumental in drawing a full picture of the process of reading as will be later discussed in the chapter. Nevertheless, the results are also interesting in and by themselves. For example, Reingold and Rayner (2006) made a series of tests that showed that reduced contrast between the target word and its background significantly increases the fixation times on the target word but has only a negligible effect on the fixation durations of word n=1. Alternating cases, using a mix of uppercase and lowercase letters of the target word n, also increases fixation times on word n though to a lesser extent. It also increases the fixation times on word n+1, and to a much stronger extent than the decreased contrast. Increased typeface weight had weak influences on the processing of target word n but did exhibit stronger ones on word n+1 (Reingold & Rayner, 2006; Wang & Inhoff, 2010). This last finding is interesting in the context of typsetting. Every time one sets a word in bold, it increases the time needed to read the word that comes right after. Another interesting finding of the Reingold and Rayner study is that the effect of the quality of the visual stimuli has an impact on the first stage of lexical processing but not on the second one. The E-Z Reader model allows for 2 stages of lexical processing (Fryal M. Reingold & Rayner, 2006), and this study has pinpointed the time stamp of when the visual characteristics play a role in lexical processing.

Studies have also found that any degradation in the visibility, as in blurring for example, of a few letters in a word will reduce the reading rate, which is the number of words read per a unit of time. The disruption is worst when the letters at word extremes are degraded (Jordan, Thomas, Patching, & Scott-Brown, 2003).

Reading on Screen

Early testing of reading on screen showed that it is slower than reading from paper (Gould, Alfaro, Barresi, et al., 1987). Certainly reading on a TV screen was possible but it was 28.5% slower than reading from paper (Muter, Latremouille, Treurniet, & Beatt, 1982). The intermittent light shining into the eyes was one possible reason for slower reading since the frequency of the light was disturbing the oculair motor control (Wilkins, 1986).

However, later research found that reading on screen could be as fast as reading on paper, and yielded similar comprehension levels if several factors were met (Gould, Alfaro, Finn, Haupt, & Minuto, 1987):

- The quality of the display needs to be good with high enough resolution.
- The typefaces used need to be similar to those used in print rather than dot matrix fonts used in CRT displays.
- There needs to be ample contrast of dark text on a light background, and the font should be anti-aliased (adding grey pixels to achieve a smooth effect rather than a jagged edge at the curves) (Gould, Alfaro, Finn, et al., 1987).

The smoothness of the outlines was further supported by the testing of the effects of ClearType hinting on LCD screens. ClearType employs a sub-pixel rendering technology that smooths the jagged edges on curves on PC computer screens. When faced with a choice of ClearType vs. grey scale rendering (as visible on Windows XP), readers had a preference for ClearType (J. Sheedy, Tai, Subbaram, Gowrisankaran, & Hayes, 2008). Studies have shown that there is an advantage to reading in ClearType, showing higher efficiency in word processing with lower reading times, fewer fixations and shorter fixation durations; however, these effects were more relevant for more "intricate" typefaces such as ones with curves or in italics (Slattery & Rayner, 2009). Another interesting finding was that this improved reading performance did not have any tradeoffs in reading comprehension (Slattery & Rayner, 2009).

Increasing pixel density (having more pixels to render characters on screen) also results in better performance in letter identification; this is most profoundly felt in low resolutions (J. E. Sheedy, Subbaram, Zimmerman, & Hayes, 2005). In fact, this specific study found that if conditions were equalized between LCD and CRT screens and paper, no main effect for display was found (J. E. Sheedy et al., 2005). The authors claim that letter identification could be equal among these conditions, but that cannot really be asserted with full confidence, as the failure to find a main effect for a display does not necessarily mean that there is no effect at all. Simply, it can only mean that no effect was found. The study found an improved performance with ClearType turned on (J. E. Sheedy et al., 2006), as has been demonstrated previously.

With regards to the typographic presentation of text on screen, the effect of line length and column width have been shown to affect reading speed. Short line lengths of 25 characters per line negatively influence the reading rates compared to the medium and long lines; long lines (100 characters per line) only held an advantage over the medium length (55 characters per line) when the subjects are asked to skim through the text (Dyson & Haselgrove, 2001). The study also found that the medium line length resulted in the highest comprehension rates of the three conditions (Dyson & Haselgrove, 2001). Comprehension was also higher when text was presented on screen in three columns in a paged design (requiring no scrolling) rather than in one column that did require the reader to scroll through (Dyson & Kipping, 1997). When the single column did not require scrolling, its reading rate was faster than that of the 3 columns design (Dyson & Kipping, 1997). This is consistent with an earlier study that also found that readers were more efficient in reading static pages than ones that required scrolling (Kolen, Duchaikey, & Ferguson, 1981) and a later study that showed that longer lines lead to better scanning but that shorter lines are more in favor (Ling & Schau, 2006).

Though reading on screen might proceed as fast as that on paper, recent studies have shown that an increasing number of people are suffering from what is being referred to as computer vision syndrome. This is a collection of symptoms that include "eyestrain, tired eyes, irritation, redness, blurred vision, and double vision" that are mostly due to dry eyes (Blehm, Vishnu, Khattak, Mittra, & Yee, 2005). One of the side effects of screen flicker is an increase in saccades that fall short of their target and premature triggering of saccades that results in inaccurate landing positions (Kenney & Murray, 1999).
Patterns and Features

Text is a series of letters strung together, repeated on and on, making a pattern out of these components. These letters make words, but can we recognize words if we were unable to recognize the component letters? A study has shown that a word cannot be read if its individual letters are not individually identifiable. Even the most common 3-letter English words are recognized as a pattern of features rather than as one feature. A word, therefore, is never one feature. Everything that one sees, is a pattern of features and this limits our efficiency in reading where efficiency is inversely proportional to word length (Pelli, Farell, & Moore, 2000).

In the Latin script, terminations, or the areas where strokes begin or end, have been shown to be the most important features in letter identification and differentiation from other letters; other features include curves, the direction these curves open to, lines, intersections etc. (Petti et al., 2008). A feature can be defined as an image, or image component, and suppose that there are several possible features, so that the signal to be identified can be described as a sum of independently detectable features” (Pelli, Burns, Farell, & Moor-Page, 2006, p.4647).

Efficiency is the unit of measurement when looking at letter identification and is calculated as the “ratio of thresholds of ideal” and human observers… (which is a pure measure of human ability) (Pelli et al., 2006, p.4649). Pelli et al. looked at various script systems and tested the efficiency of letter recognition. The scripts in order of most to least efficient in letter identification were listed as: Latin, Hebrew, Devanagari, Armenian, Arabic, and Chinese. Arabic and Chinese are the least efficient in letter identification and that is most likely because of the complexity of the letters themselves (Pelli et al., 2006). Note here that the test was carried out with individual and isolated letterforms so the role of Arabic orthography has not yet come into play. This study showed that complexity is inversely proportional to the efficiency of letter identification, and that it is an excellent predictor at that (Pelli et al., 2006).

Moreover, the study has shown that any modification that added complexity to letterforms also had an inverse effect on the efficiency of their recognition (Pelli et al., 2006). In fact, the authors concluded that “simple forms are seen efficiently, complex forms inefficiently, as though they could only be seen by means of independent detection of multiple simple features” (Pelli et al., 2006, p.4665). It is, however, perplexing how the study found Devanagari to be less visually complex than the isolated Arabic letters as the forms of the letter are much more intricate and dense.

An important factor to keep in mind regarding shape identification is the degree to which this particular shape is similar to other shapes with which it might be confused (Attnave & Arnould, 1956). This is possibly the reason why Arabic had such low efficiency marks since so many of its letters share the same basic shape and it is only the number and position of dots that differentiate them.

Within the same study by Pelli et al. (2006), one letter of a script typeface (Kuenstler) was calculated to be as complex as a common 5 lettered English word set in a typical serif typeface (ITC Bookman). Subjects were shown individual letters of various scripts, usually for 200ms and were then shown a line up of 26 letters including the one already shown. The task was to identify which letter was shown at first. Evidence showed that it took only hours to learn to master such a task and that years of reading in this script, up to reading a billion letters over the span of 40 years, did not add to the efficiency of letter identification. It is important though to add that the subjects did not need to remember which letter it was, but simply to pick it out of 26 possible options (Pelli et al., 2006).

Another interesting finding of the same paper was that letters and words are identified in the same way we identify everyday objects whether they are inert or living by detecting independent features, usually around 7 features per letter. This is usually referred to as the probabilistic summation, which posits that the probability of not identifying a stimulus made of several features is equal to the product of probability of not identifying its individual features (Pelli et al., 2006). It supposes that a stimulus is identified via “any one of its features, and that the features are detected independently” (Pelli et al., 2006, p. 4669). Indeed, this theory has become the “null hypothesis of visual perception, an extremely simple model that accounts for much that we know” (Pelli et al., 2006, p. 4647).

The Question of Word Shape

The role that the word shape plays in reading has been tested via the preview benefit as in a study by Rayner, McConkie, and Zola (1980). Words were presented paradoxically to the subjects, and when the eyes moved towards that target, the words were changed. The results were quite interesting when it came to the question of word shape and the integration of visual information across saccades as briefly mentioned earlier in this chapter. Of interest, here, is the finding that the words “ROUGH” and “rough” had equal facilitation for the target word “ROUGH,” even though the word shape and visual characteristics are different. Moreover, readers did not benefit from parafoveal previews of a word that had the same word shape but different letters (Rayner et al., 1980).

Actually, using the display change to test if the subjects were even aware of this switch, researchers found that subjects were more likely to notice a difference when the letter codes were changed, rather than when the lettercase was (Slattery, Angele, & Rayner, 2011). The one situation in which the word shape does play a role is in acronyms. Acronyms set in uppercase in a normal sentence case context were read as acronyms (i.e. read out by the initials), but these were mistaken for normal words (i.e. read out as whole words) when set in an all caps sentence (Slattery, Schotter, Berry, & Rayner, 2011).

Pelli and Tillman (2007) also looked at the reading process and questioned the methods used in word identification: 1. Letter by letter, 2. as a whole word or 3. based on the sentence context. They found that letter by letter identification accounted for 22% of the reading rate, identification of whole words accounted for 16%, and the context accounted for 22% but that rate varied across readers. They found that these 3 processes are dissociated from one another but that their effect is additive. The conclusion they get to is that words, like objects, are recognized by parts, as wholes, and depending on the context (Pelli & Tillman, 2007).

The Legibility of Type

Defining Legibility

A definition of legibility that deals only with the visual characteristics of text is inherently flawed, for legibility is a measure related to reading and recognition. Legibility operates in the process starting with the viewing of a linguistic symbol,
and obtaining meaning. It cannot, therefore, be contained in the pure visual characteristic of that symbol. As such, it is relevant and necessary to define legibility within the context of the reading process.

As mentioned earlier, it is in the first 50-60 ms of a fixation that the reader needs the visual stimulus to exist in order to extract relevant visual information, and that if the fixated word were to disappear after that period, reading would proceed as normal (Rayner et al., 1981). It is safe to assume that the visual characteristics of a fixated word show their effect in that early time frame of the fixation. Legibility effects, then, are felt early on in a fixation. This, however, is only one part of the overall legibility effects.

Starting with the EZ Reader model, there are different stages in word identification. An important stage is the shift in pre-attentive processing onto word n+1. Low level processing extracts abstract letter codes that contribute towards the preview benefit. The visual characteristics of a typeface are then very likely able to facilitate the preview benefit, in turn adding to the fixation times and increasing saccade lengths. Legibility effects, then, operate during two distinct processing stages: The first 50-60 ms of a fixation, and during the “pre-attentive” processing of parafoveal words.

This localization of legibility effects has already been demonstrated. One study has clearly shown that typographical manipulations have an effect during the stage L1 of lexical processing but not during L2 (Reingold & Rayner, 2006). Similarly, the use of bold weight on word n+1 gave higher fixation durations for n+1 (Reingold & Rayner, 2006), most likely because of lower performance in the parafoveal processing of word n+1.

As to the definition of legibility, the most convincing to date is that by Slattery and Rayner: “How easy the letters in a word are to encode” (Slattery & Rayner, 2009). In other words, legibility is how easy it is to extract the visual information of the fixated word in order for lexical processing to begin. The ease of encoding the visual characteristics of words via their letter codes manifests itself in faster fixation times. The facilitation of the preview benefit results in shorter fixation times (due to the facilitation of extracting letter codes from word n+1), a wider perceptual span, and wider saccades.

The wording here is interesting to dig deep into. Is it decoding or encoding? A dictionary definition gives: "Decode: To extract the underlying meaning from and "Encode: to convert (a nerve signal) into a form that can be received by the brain" (Forbes, 2012). Another definition that is more relevant to reading as a process:

"For the simple view, skilled decoding is simply efficient word recognition: the ability to rapidly derive a representation from printed input that allows access to the appropriate entry in the mental lexicon, and thus, the retrieval of semantic information at the word level" (Hoover & Gough, 1990).

One might be tempted to use the word “decode” in such a case, since the purpose of reading is to extract meaning. However, a more accurate term is encoding for legibility has more to do with the recognition of letter code (and making this information available for higher level cognitive processes) rather than the ability to put these together into meaning. In other words, legibility is not about word recognition as in the Simple View of Reading by Hoover and Gough. It is about the ability to recognize the abstract letter codes so that lexical processing can begin. In the Simple View, reading is the product of word decoding and linguistic comprehension:

"In the simple view of reading, linguistic comprehension is the ability to take lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations. Reading comprehension involves the same ability, but one that relies on graphic based information arriving through the eye" (Hoover & Gough, 1990).

This again drives home the point that decoding involves lexical access at the word level, which is more than the pure level of what legibility is. As a counter example, imagine that a subject is presented with a word in Spanish and asked to read it aloud. If the subject does not understand Spanish, then the stimulus is not decoded. However, the subject will still be able to read it aloud to a certain degree as he/she will recognize the letterforms. The word would be legible, but not comprehended. Similarly, if legibility is the ease of decoding, i.e., word recognition, then testing for legibility with non-words is not valid. However, this is again not the case. In addition, the different stages of reading in the EZ model split it into stages where orthography, phonetics, etc. are being processed, and higher level processing (comprehension) comes later. As discussed in the paper, legibility effects come into play early (first 50–60 ms) and not late in the process. Hence, legibility will enable comprehension, but the ease of comprehension is not a measure of legibility.

Methods for Measuring Legibility

Researchers have used various methods to study the legibility of typefaces. The following list is a brief overview of the types of testing used:

- Speed and comprehension: Reading speed and a comprehension check are quite common methods, as favored by Tinker (1963). The logic in such tests is that more legible type will facilitate the reading process and result in faster reading. Speed can be measured as in total reading time or the number of words read per minute. This is usually a test of continuous reading.

- Naming: This usually involves the presentation of a word or nonword either in the Invern or parafovea. The faster or more accurate the words are named, the more legible the typeface is. This is a very pure method of measuring legibility as higher-level processes are not involved.
Threshold visibility: This is a form of a naming test. The threshold is either that of exposure time (the stimulus is presented for a very short period of time, and then masked), or that of distance (the stimulus is presented at a distance from the reader). The shorter the exposure rate, or the farther the distance, the more legible a typeface is supposed to be (Lund, 1999). Another less frequent threshold is that of luminance, with higher legibility for the lower luminance at which characters are still identifiable (Yager, Aquilante, & Plass, 1998).

Eye tracking: The use of eye tracking to test legibility gives very accurate speed tests. When looking at global reading measures, a more legible design is characterized by shorter fixation times, longer saccades, and a smaller total number of fixations (Slattery & Raynor, 2009).

Blink rate: How often the reader blinks can be indicative of eye fatigue. Less legible text would hypothetically cause more fatigue, which results in increased blinking (Luckiesh, 1947).

Search tasks: The reader is given a search task to complete and the faster that is the more legible the design (Lund, 1999).

These methods measure quite different things. Speed of reading and the various eye tracking data are concerned with the reading of continuous text. As long as the different typefaces are being tested with paragraphs of equal difficulty and familiarity levels, then the linguistic effects should not interfere, and the global reading measures are good indicators of legibility.

Naming and threshold visibility are used in the identification of letters or words and as such are in many ways different from reading as a process. On the one hand, they do not involve higher level cognitive processes and are therefore, paper measures of legibility, especially if nonwords are used as that neutralizes the effects of word frequency. However, the distance threshold test is a test of visibility and how a typeface degrades over distance. It is then very different from the typical reading process in the fact that both type size and distance are quite different in these settings. This is not to say that either of them is not valid at a test, but rather that they are measuring different aspects of letter identification, and this is to be expanded upon in the following sections.

Letter Identification and Different Script Systems

Letters have a raison d'être, and that is to be identified. They are a visual code that is the building block of written language, and that is true of any kind of writing in any kind of language. It is, therefore, quite relevant to note that letterforms i.e. how many strokes are actually needed to differentiate one letter from another. Averages, for example, were: Arabic 2, Cyrillic 3.69, Devanagari 3.27, Greek 1.71, and Latin 2.08. The overall average for all 115 scripts combined was roughly 3 strokes per letter (2

for numerical systems) and with around 50% redundancy. This average of 3 strokes was irrespective of the size of the character set of the respective writing systems. The authors offer a very interesting hypothesis: this commonality across various writing systems is actually due to the need of letters to be seen. Therefore, the physiology of the human visual system is influencing the development of writing systems. It is their view that the existing visual system for object identification could be behind the selection and developmental trend of such forms (Changizi & Shimojo, 2005).

This view is certainly logical when one looks at the commonalities of human inventions that have sprung up across different cultures. Like the wheels are round and walls are more or less straight, there is a certain level of efficiency to the way letters are drawn: a reasonable number of strokes per character and enough differentiation to keep letter identification more or less simple. Even in the most complex of writing systems, strokes are generally either straight lines or circular in form. Very rarely does one come across squiggly lines or zigzagged strokes. In such case, it is then logical to assume that there is a universal human drive towards efficiency in form, but only up to a certain point. The authors point out that 50% redundancy is the cost of improved differentiation (Changizi & Shimojo, 2005), and that is certainly in keeping with a selective drive for efficiency in visual communication. As shown by Pelli et al. (2006), some script systems are more efficient than others, but it is as interesting to note the differences between scripts as it is to note that which is common.

Indeed, these common elements of the different script systems stem from several important factors, as Changizi, Zhang, Ye, and Shimojo (2006) expand in a follow up paper. The basic principle that they build upon is the various configurations in which the number of strokes can combine to form different shapes. For example, 2 strokes can intersect in a T shape, an L shape, or an X shape. These configurations are topologically identified in such a way that the lengths of the strokes, or the orientation of the shape can vary, but the basic topology remains the same. Having noted that the 115 scripts that they analyzed had an average of 3 strokes per character, the authors developed a topological system of the possible configurations of a 2- and a 3-stroked letter. These configurations were seen to be similar across human visual symbols, in a non-random fashion. The authors also compared these configurations with shorthand symbols (signs developed to support the motor skill). They assessed the efficiency of the consecutive strokes and pen lifts and found no correlation between motor complexity and the distribution of topological configurations that letterforms take which lead to the conclusion that these shapes were "not strongly selected for the motor system" (Changizi et al., 2006, p. 123).

The authors also looked at the relationship of these configurations and the perception of visual complexity. They compared them to trademark symbols (these developed for the visual aspect rather than the motor one). Their results found a strong correlation between visual complexity and the visual configurations and concluded that these signs used in script systems are selected for the visual system and not for the motor one (Changizi et al., 2006).

Next they compared these topological shapes found in script systems to those that are found in nature and found a relationship there, leading to the conclusion that the selection and evolution of these shapes have developed in accordance to what the human eye is best trained to recognize (Changizi et al., 2006).

The remaining point to be argued, is that the written symbols need to be simple enough for any given person to be able to learn how to write them. Of this Unger (2012) writes:

"... Legibility research is really a study of ergonomics... While learning to write and read, such basic elements [such as elementary graphic entities, that
Letter identification and frequency channels

The recognition of letterforms is carried out via specific frequency channels that are dependent on the scale of the letter, and are influenced by the stroke frequency of the letter itself (Majaj, Pelli, Kurshan, & Palomares, 2002). The authors defined stroke frequency as the average number of strokes that intersect with a horizontal cross section of the letter; the division of that average by the average letter width yields the spatial frequency of the letter (Majaj et al., 2002). The authors found that the selection of the channel for identifying letters did not vary across different sightings of the same character. In fact, the observer was using the same channel to identify letters, rather than choose different channels that might help to reduce noise. They conclude then that the choice of which channel to use is dependent on the letter itself rather than the observer, making letter identification a bottom-up process (Majaj et al., 2002). The spatial frequency channel also does not change if attention was overt, as in a direct fixation, or covert, as in the pre-attentive phase of letter identification (Tsap, Pelli, & Carrasco, 2004). It is also the same frequency (1.6 octaves wide) whether the reader is looking at a letter or a word (Majaj, Liang, Marrelli, Berger, & Pelli, 2003).

Another interesting finding from the same paper reveals that different aspects of letterforms play different roles according to the size in which the type is set: “Large letters are identified by their edges, small letters by their gross strokes” (Majaj et al., 2002, p. 1181). In other words, when type is set in large sizes, the reader will react to the details of outline treatment. On the other hand, type set in small sizes is identified via the strokes themselves and how they relate to one another rather than how their outline is treated.

Letters Identification and Brain Functions

Dehaene (2009) reports on several interesting findings related to specific brain areas related to letter identification. He gives the example of a patient who, after suffering a stroke that affected his right visual field, was no longer able to recognize letterforms or words. The patient’s linguistic abilities and writing skills were not affected and he was still able to recognize numbers which led to the conclusion that “reading digits relies on anatomical pathways partially distinct from those used for reading letters and words” (Dehaene, 2009, p. 56).

This loss of the ability to read is referred to as pure alexa and the patient mentioned was the first recorded example by neurologist Joseph-Jules Déjérine in 1892. The point to make here is that there are specific areas of the brain that are essential for reading and letter identification, and that an injury there could very much oblitrate the possibility of the patient to ever read again.

Indeed, the region of the brain critical for reading is located in the lower back area of the left hemisphere, or more accurately “a few centimeters to the front of the occipital pole, on the bottom side of the left hemisphere” (Dehaene, 2009, p. 62). Research has also shown that this region is strategically placed to act as a channel through which all visual word information flows which is then dispersed to many other areas of the left hemisphere (Dehaene, 2009). This "letterbox" area is also immune to changes in uppercase and lowercase and will show similar activation when one or the other is used as a prime. So the abilities of this region are not purely visual in nature but are able to function on a more abstract level. Moreover, this region remains the same for readers of right to left scripts as well and the location is almost identical in readers of Chinese.

With the help of brain imaging research, it is also possible to detect the time course of activation in the two hemispheres in response to specific reading tasks. stimuli presented in the right visual field are transmitted to the left hemisphere and vice versa. However, the letterbox region in the brain that specializes in recognizing words is in the left hemisphere. It turns out that for words that are presented in the left visual field, their initial transmission goes to the right hemisphere but after 40 ms, this signal is transmitted onwards to the left hemisphere to this same letterbox. Within 200 ms, these words are processed just like the ones presented in the right visual field (Dehaene, 2009).

Going back to the issue of the commonalities among world scripts, neuroscience offers further evidence to the concept of universality in writing systems. Dehaene (2009) offers these observations:

- Writing systems involve concentrated black on white marks that are probably the most efficient way to transmit visual information to the retina.
- All scripts use a small number of basic shapes which are hierarchically combined to create words, word-parts or letters.
- All the script systems make the assumption that size and location can vary, but rotation cannot. This is most likely due to limitations in the ability of the visual neurons to tolerate only 40 degrees of rotation.

This is further support for Changizi’s conclusions that script systems evolved to fit what the human eye, and by extension the human brain, is able to see. It is also what the eye and the brain are able to process efficiently.

Letter Identification as a Measure of Legibility

Several studies look at the question of legibility via the task of letter identification. In these studies, the experiments test the relative legibility of typefaces via how easily their individual characters can be identified, as well as the internal legibility of the typeface itself where tests check for the accuracy of character identification and probabilities for misidentification. One study of this sort was carried out by Chaparro, Shalih, and Chaparro (2006) which looked at the relative legibility of Cambria, Constantia, and Times New Roman. Commonly confused pairs of letters were (l, i), (z, z) and (o, o). Of these 3 typefaces, Cambria had the highest number of correctly identified letters. A typeface in which letters are not confused with another is most likely a more legible typeface. The test for identifying letters is a good measure of the ease with which one can extract the visual characteristics of what is being read. A word cannot be identified if its individual letters are not. Since the probability of not recognizing a word is the product of the probability of its letters not being recognized (Pelli et al., 2006), this sort of test can give clues as to what constitutes a more legible design.
Another comparative study looked at the relative performance of the letter e and number 0 in 20 typefaces (Fox, Chaparro, & Merkle, 2007). Character identification tasks were run and the typefaces were listed in the order in which these characters were confused with others. Though such a listing is very common in legibility studies (which typeface performed better than another), it is the lesser interesting aspect of this paper because the typefaces were tested at 10 pts. each and so the optical size was varying. This is a possible confounding factor as some typefaces will appear larger and size has an effect on legibility performance (Webster & Tinker, 1935). The really interesting aspect of this study is that the authors calculated different factors and analyzed the letterforms to arrive at specific characteristics that improve the legibility of these characters. That is definitely more informative and relevant than the relative performance of a group of typefaces. This sort of information can help designers to improve the legibility of their design by introducing characteristics that aid in correct character identification. For example, the study found that the lower the crossbar of the lowercase e is, the better it is at being recognized (Fox et al., 2007).

The use of legibility studies to inform type design choices seems to be picking up. A recent paper examined the variations of letter shapes and their effect on letter identification both at a distance and in paraverbal processing (Beter & Larson, 2010). The results, such as better performance if narrow letters were wider or x-height letters took up more of the ascender or descender space, are quite interesting for several reasons: 1. Modifications were made on the same typeface being tested and so the variables were well controlled. 2. Modifications were done systematically and were targeted towards testing specific features. 3. Therefore, it is possible to generalize the findings thus contributing to the external validity of the research.

As to which features are most important in letter identification, and as mentioned earlier in this chapter, letter identification is dependent on specific features of letters. The study by Fiet et al. (2008) showed that stroke terminations were the most important feature in the identification of Latin characters and the ones that carried the most valuable visual information. In terms of lower case, the three most important features were: terminations, horizontal, and slant tilted right; for the upper case, they were: terminations, horizontal, and intersections (Fiet et al., 2008). These are then the most critical features that operate in letter identification.

Structure and Word Identification

Readers have certain expectations in terms of the regularity of a typeface design that produces efficiency in letter identification; irregularity in typeface characteristics have the opposite effects since the readers are tuning in to these characteristics and the details that they expect to see (Sanocki, 1987). Indeed, mixing fonts results in lower reading performance which suggests that readers develop a schema of characteristics that help them to identify letters more efficiently (Sanocki, 1988). Chinese readers showed the same font tuning process when reading non-Chinese readers, however, did not font tune when reading Chinese leading to the conclusion that font tuning is specific to the task of reading and that the processing of letters as shapes is different from letter identification processes (Gauthier, Wong, Hayward, & Cheung, 2006).

Recent evidence points to the active role that structure plays in word identification: readers use visual cues based on a structural description of letterforms during word processing (Walker, 2008). The structure of a letterform is the spinal cord around which the letter is drawn, and can be arrived at by drawing the median line in between the 2 edges of the letter outline. The structure of one letter is, by conventions of type design, related to other letters via various relations of size and proportion. The end result is a system of structural relationships around which the whole typeface is built. For example, if the letter is tilted forward (therefore we have an italic or oblique structure), one would expect that the whole typeface is tilted forward with the same angle. This knowledge comes from the exposure to typographic practices, and is referred to in this paper as the "translation rules." The reader, Walker explains, carries forward these translation rules for the typeface of text. This appears to be the case in an underlying structure across letterforms. These translation rules are time dependent if there is enough of a time lapse between the presentation of stimuli then the font tuning effect disappears (Walker, 2008). The reader then carries forward expectations of what the structure of the typeface is supposed to be, but if enough time passes, these expectations are no longer present. The Reader can retain the font translation rules in working memory for up to 750 ms, though this seems to be an unconscious decision rather than a conscious strategy (Walker, 2008).

Therefore, it seems that the features needed for letter identification are not set loosely but rather fixed onto an underlying structure that is modulated by the design of the typeface that the text is set in. In other words, the features that are critical to letter identification such as terminal endings are not ones that operate in a free grid. These features (or rather body parts when one is referring to letter anatomy) need to be set in the right position that is dictated by the internal structure of the typeface itself.

To give a human related example, if the face is the most important feature for recognizing people, it needs to be placed in its right position: the front of the head, rather than the back of one's feet. Moreover, these features need to conform to the Gestalt law of good continuation and the law of grouping, i.e., that the features are placed close enough in order to be recognized as one whole object (Pelli et al., 2009).

Crowding

Researchers observed an interesting phenomenon: A letter is less likely to be identified in peripheral vision if it is surrounded by other letters; the farther it is, the higher the "eccentricity" and the less likely it will be identified (Pelli et al., 2007). Crowding, then, is this "excessive feature integration, inappropriately including extra features that spall recognition of the target object" (Pelli et al., 2007, p. 2).

An interesting aspect of crowding is that it is able to predict the size of the visual span: that the visual span is in effect the span of text that is not crowded (Tillman et al., 2007). Tillman et al. (2007) found that the minimum letter spacing needed for letter identification (so as to avoid crowding where letter identification is no longer possible) predicted the minimum letter spacing as well as the visual span. As to what the visual span is, Pelli (2009) describes it as the "number of letters in a line of text, that one can identify without moving one's eyes."

Letter Superiority vs. Word Superiority: Redefining Legibility

At first glance, there seems to be a conflict between two observations related to letter and word identification:

1. Word superiority effect, which says that a letter can be better identified in the context of a word than when it is on its own or in a non-word (Reicher, 1969).

2. Letter superiority, which says that a letter is easier to identify when set in the context of a word than when it is in isolation.
The letter superiority effect, which says that in threshold testing, an isolated letter has better visibility than a word (J. E. Sheedy et al., 2005).

This conflict is similar to the apparently opposing results regarding the effect of weight on typeface legibility. As will be discussed later in the chapter, a bold typeface resulted in longer reading times (Slatté & Rayner, 2009), and yet Sheedy et al. (2005) found that weight improved the performance of a typeface in threshold testing. However, the results need not be contradictory. For these to be reconciled, we need to look at the definition of legibility again.

Legibility is the ease with which the visual features can be extracted. So, what is a good measure of "ease"? To answer that, we first need to look at the process of extracting visual information. As mentioned earlier, the viewer uses different frequency channels to identify letterforms based on their size, this channel is dictated by the signal itself and not by the viewer, and that size can be manipulated by changing the viewing distance (Majač et al., 2002). Thus, it is possible to conclude that text seen at a distance is inducing a different frequency channel than the test at a close range one, as it is generally larger than typical sizes used in body text. So back to the original question, how can one measure "ease"?

It stands to reason, that in the context of typical reading settings as reading from a printed publication or from screen, the ease of letter encoding would translate into shorter fixation durations, lower reading times, and less locations in total. However, in the case of reading from a distance, which is reading under duress, it is a different scenario. It is not about the speed with which letters can be encoded, but rather the question if they can be encoded at all. The threshold test is then a valid test for reading under duress just as reading measures are good indicators of the legibility of type in continuous reading. The apparent contradiction in the results of word vs. letter superiority, as well as the different indicators for the use of bold, can be reconciled by the observation that these different frequency channels are being used to analyze letterforms, and by Majač et al. finding that different aspects of letterforms (outline edges vs. gross strokes) are informing letter identification.

If legibility is about the ease of extracting visual information, then in the case of normal reading, speed is a good measure. In the case of constraints on visual acuity (as in reading from a distance), then the threshold measure is good. Both are compatible because the relative size of the text affects the frequency channel used to see it. In such cases, then legibility is relative to the distance from the stimulus, as well as the task required.

It follows then that legibility is relative rather than absolute, and depends on relationship of the viewer to the stimulus in terms of distance, task required, and familiarity with the content in both visual and linguistics aspects. Specific aspects of text styles such as italics are less legible whether viewed up close (Slatté & Rayner, 2009) or at a distance (J. E. Sheedy et al., 2005). Other aspects change roles, so text set in bold and viewed at close range will read slower that text set in regular (Slatté & Rayner, 2009), but when viewed from a distance, a bold letter is better distinguished from a regular one (J. E. Sheedy et al., 2005).

J. E. Sheedy et al. (2005) do address the apparent conflict between letter and word superiority and also find that these results are reconcilable, though they offer a somewhat different explanation:

"The letter superiority effect has been identified at threshold sizes and seems attributable to early visual-processing factors. The letter superiority effect may or may not apply to text presented at suprathreshold size, subject to the same discussion presented earlier about the amount by which text is above threshold. Letters at typical reading size are more above threshold than are words of the same size; hence it can be argued that letters should also be more legible than words at typical reading sizes. However, this requires further study. The word superiority effect, on the other hand, refers to several related findings that individual letters are more accurately read when they appear in a word than when they appear embedded in nonsense strings. A familiar context (a word) makes individual letter recognition easier than does an unfamiliar context (nonsense letter string). This effect is more likely attributable to higher level cognitive factors than those required for simple letter identification. The present results show that individual letters are more legible than words at the smaller and earlier processing stage of letter and word identification." (J. E. Sheedy et al., 2005, p. 812)

There is a further argument to support their claim, and that is the issue of efficiency of feature detection. As mentioned earlier, the efficiency of feature analysis needed for word identification is inversely proportional to the number of letters in a word (Pei et al., 2003). Therefore, it follows that identifying a single letter is more efficient than identifying a whole word.

Moreover, the word superiority effect is not claiming that the word is superior in being recognized as the same implies, but rather that a word provides a context which aids in faster recognition of letterforms. There is then no conflict between these findings, but rather that they paint a picture that is more complex and nuanced than what one might initially expect.

As to the relative nature of legibility, the conclusion of Beier and Larson (2010) supports this changing aspect of letter identification:

"The study confirmed the notion that the performance of letter shapes varies according to the situation in which it is presented, and that some features are most important in distance viewing and others are most important in the paradoxical view" (Beier & Larson, 2010).

The changing dynamics of text is nothing new and was expressly addressed almost one century ago:

"The factors which make it possible to read one style of print at a greater distance than another may not be the same as those which lead to the reading of one type face faster than another under ordinary reading conditions." (Webster & Tinker, 1935)
Webster and Tinker (1935) think that it is the context clues that are facilitating reading text in paragraphs and that might be the reason behind the different results. Though that might help in the reading process itself, it does not explain why one typeface is more legible than another. The reasons for different performances of typefaces depend on the design of the typeface, the distance from the viewer, and the setting of type itself.

### Uppercase vs. Lowercase

A typeface is not an end product but rather an ingredient in the design process. In terms of legibility, what one does with the typeface (how it is set, which case is used, etc.) can be as important as the design of the typeface itself. A few studies looked at the role of case in Latin typesetting. Arditi and Cho found that in threshold sizes, uppercase text performed better than lowercase in both threshold testing and in the speed of reading tests (Arditi & Cho, 2007). The study was done with both uppercase and lowercase text set at the same point size so the overall area and the size of the counters of uppercase letters are bigger than those of lowercase ones. The experiment is then relevant to question of optical size. It confirms the consensus amongst typographers that uppercase letters appear larger and more visible than lowercase ones that are set at the same point size and that this is advantageous in the case of very small type (Fig. 5.7). The general convention regarding the style of Latin text is that text set in mixed (sentence) case is easier to read than text set in uppercase:

"Lower-case letters have more 'character' in terms of variation in shape and the contrasting of ascenders and descenders with short letters. This leads to characteristic word forms that are much easier to read than words in all capitals" (Tinker, 1963, p. 34).

Webster and Tinker have also tackled the same issue of different results between visibility from a distance and continuous reading and came away with this conclusion:

"It was found that material in lowercase letters had more definite word form and was read faster in connected discourse than material in all capitals (uppercase). The latter, however, was perceived at a greater distance from the eye than the lower-case print. The larger outlines of the upper-case letters undoubtedly caused the difference" (Webster & Tinker, 1935)

Later studies have shown that word shape is not facilitator in parafoveal processing but rather it is character code information (Rayner et al., 1980). However, even when the word shape yields no preview benefit, the greater distinction between lower case shapes and features should facilitate the extraction of visual information.

### The Effect of Style

The style of the typeface used has an effect on reading performance. It is though less pronounced than the effect of the difficulty level. A study looking at the different reading measures when reading in Old English vs. Times New Roman found that reading in Old English resulted in a larger number of fixations, higher fixation times and shorter saccades. The font effect was also larger for the older group (almost 54 years older on average) who fared worse with Old English (Fig. 5.8). There was a significant interaction between age and font style for sentence reading time and fixation duration and marginally significant for number of saccades and regressions. Readers were more likely to skip words when reading in Times (Rayner, Bolch,G. et al., 2006) which is again indicative that it easier for them to read in Times.

When reviewing legibility studies, the question of external validity comes up. An often cited paper is by Mansfield, Logge, and Baue (1996) which looks at the relative legibility of proportional vs. fixed width typefaces for readers with normal or low vision. The testing is done with Times Roman and Courier Bold. As a comparative analysis of two unrelated typefaces, the findings are relevant. However, it is at the point where one tries to extrapolate these results that drawing conclusions becomes that much harder.

The authors admit that there are many differences between the typefaces and that these could be the cause of any results found (Mansfield et al., 1996, p. 1493). Of these differences, the most obvious is that one is in regular weight while the other is in bold. Another noticeable difference is the very large difference in word space width. Both of these factors have an effect on reading speed as shown earlier in the chapter (Bold weights increase fixation durations, and word spaces play a very important role in the ability to program saccades). More subtle differences include the contrast and modulation of forms in Times, the heavy slab serifs in Courier Bold, and the large difference in the character widths of the two typefaces. As such, the results of this study cannot be generalized, and can be only seen as a comparative analysis of the relative legibility of two typefaces.

Another study looking at the relative threshold performance of typefaces was the already cited study by Sheedy et al. (2005) which tested 6 typefaces (Georgia, Times...
One of the shortcomings of the studies regarding the role of serifs in reading is that they were more of comparative performance analysis rather than an isolation of the factor to be studied. So if one was to compare Arial and Times, whatever results generated cannot be easily nailed down as due to the presence of serifs because there are several other variables that have not been controlled in the experiment testing. While such research is informative for graphic designers in the process of selecting one typeface or another, it is less informative for type designers in the process of designing a new typeface.

The legibility effect of specific features is an area that easily invites further research. A study that finds Times to be more legible than Arial cannot be generalized to mean that all serif typefaces are more legible than sans serif ones. At the end of the day, it is an issue of external validity and the ability to build theories that can be verified by empirical data. What it does is to actually offer support rather than final proof. In such an avenue of research, this might be the best-case scenario.

On the other hand, there is the other approach, and that is to design a system of typefaces where possible variables are controlled, and serifs are the only factor to change in the design. One such study designed a typeface family specifically set to control parameters that might vary between a typical serif and sans serif typefaces, and found that sans serif typefaces were 20% faster to read in very small sizes (Morriss, Aquilante, Yager, & Bigelow, 2002). They also found that this advantage disappeared in large sizes.

The question of serifs poses a question for its place in this paper. The narrative being built here tries to connect findings that are less script specific and rather universal in nature. The question of serifs is one very specific to the Latin script, and Greek and Cyrillic by extension, and is very far from the research purpose of this paper. As such, given the questionability of external validity, supported by Lund's findings (1999), and the wealth of confounding factors in the majority of studies, also reviewed thoroughly by Lund (1999), it was best to learn from the errors committed in these studies while building the experiment, but not to go deep into this topic in this literature review. A few studies stood out, and are briefly covered in the paragraphs below.

One of these studies is a paper by Robinson, Abbamonte, and Evans (1971) in which they outline a proposal as to why serifs are important, and then conduct computer simulations of letter identification that echo the visual detectors in the retina. Of the reasons they initially propose, one finds the typical reasons that one hears in typography and type design education: The serifs lead the eye across the horizontal. Serifs preserve the shape of the letters by emphasizing the terminals, etc. These they relate with the arguments that the eye makes only a low fixations per line so horizontal continuity cannot be such a big factor, and that instead of preserving the shape, serifs can very well be added noise. However, it is their conclusion that really stands out:

"Serifs are only important in letters which are small enough to be perceived by line detectors: most ordinary print in the texts of books, periodicals, and typewritten material. Larger and/or thicker letters are probably perceived by a different part of the system—the edge detectors. However, the image of large letters are of line-form when viewed from a distance. Serifs are useful not only when the letterforms are physically small, but may also be functional when large letters form a small image on the retina—as, for example, when a billboard is seen from a distance" (Robinson et al., 1971, p. 358).

Though one may question the validity of using a computer simulation to verify the way the human eye detects visual stimuli, this paper stands out especially when
viewed with the information that we know today. As discussed earlier, the reader uses different frequency channels when looking at different sizes of text, and letterforms are identified by their gross strokes in small sizes and by their edge in large sizes (Majaq et al., 2002). So, if one were to review paper from the point of view of research available in the 70s, than yes, there are many questions still open, but it seems that the authors were on to something with their theory. The interesting thing is that their theory provides a neat explanation to a very common typographic practice that students of typography learn very early on: use serif typefaces for text, and sans serifs for headlines.

Legibility studies, though, usually revolve around a human reader. Another exception was a study that looked at the legibility of various typefaces by a machine rather than human reader (Zhang, 2006). It used different OCR readers to analyze texts set in 18 different typefaces. The study shows better results for Sans Serifs and identified commonly misidentified pairs of letters. Though the approach is interesting, one wonders if OCR text analysis is analogous with the reading process.

Connolly (1998) found that the presence or absence of serifs did not play a role in legibility as the top performing typefaces of several tested were serif and sans serif designs; the study did though find better performance and subjective rating for typefaces with open design such as large counters and wide spacing (Connolly, 1998). This is another hint that the relative legibility of serif to sans serifs might be dependent on factors unrelated to the actual treatment of stroke terminals.

In either case, the role of serifs is a valid research construct and definitely an endeavor that will support typographic design using the Latin script. However, as mentioned earlier, this is a topic that is tangential to the core of this paper and so we leave it here.

The Effect of the Complexity of Style

Though the effect of the complexity of typeface style on legibility is under represented as a research topic, there are still some very interesting studies that have gone to show that the more complex the style, the less legible it is. For example, Pelli et al. (2006) looked at the effect of complexity (as evident in script typefaces) and defined "perimetric complexity" as the "inside-and-outside perimeter, squared, divided by ink area" (Pelli et al., 2006). Their results showed lower performance for those typefaces. Another study also found lower performance for a script typeface vs. a Roman (Rayner, Reiche et al., 2006), and in this case the use of a script typeface was specifically done to test font difficulty. A later study by Slattery and Rayner (2009) also looked at the relative legibility of a serif [both in roman and italic] and a script typeface, and found lower performance for script. Comparing Times New Roman to Harrington, and Script MT: Times New Roman had a smaller number of fixations, and with shorter average fixation duration; testing Roman against Italian and Bolds also showed longer reading times for Italics (Slattery & Rayner, 2009).

It is possible then that these effects, worse reading times for script and Italic (Fig. 4,10), are due to complexity of form rather than the familiarity of typefaces as the authors mention. It is especially so given Pelli’s findings, as well as the fact that the same paper found longer reading times for Italics and that are arguably almost as familiar as the Roman.

As mentioned earlier on, all visual information needed for reading to continue smoothly is extracted in the first 50-60 ms of a fixation (Rayner et al., 1981). This implies that the role of typography and its effect on eye movement processing are limited to that time frame. It is possible then that a complex visual will increase that duration, similar to the effect of a low-frequency word (lshoff & Rayner, 1986), and will also reduce the preview benefit resulting in shorter saccades and longer fixation durations. An interesting study looked at the effects of age on typeface legibility as well as the comparative performance of various typefaces. The results were independent of the age of the reader but specific to small sizes in print. They too showed consistent low reading performance for typefaces that were condensed and complex in form and design (Connolly, 1998).

Perhaps the best summation of the effect of the complexity of the style was put by Rayner et al. (2006). They wrote that as long as text looks normal, typographic effects are very small, but "It stands to reason that a font that is more difficult to encode should cause readers to look at words longer than fonts that are easier to encode" (Rayner, Reiche et al., 2006). This confirms what Webster and Tinker stated in 1935 that the simplicity of outline improves legibility (Webster & Tinker, 1935).

An interesting aspect of complexity of form is the increased number of strokes per character width, leading to higher stroke frequency as mentioned earlier. The reason why this higher frequency negatively impacts reading can very likely be due to crowding. Pelli (2008) claims that for objects to be recognized and differentiated they need to be at least 3 mm in the visual cortex at a specific minimum distance (6mm to the radially direction, or 3 mm in the circumferential direction); if the objects are closer than that, they will be perceived as a jumble (Pelli, 2008). It follows then, that if the strokes are too close to one another they stand a higher risk of being indistinguishable from one another. This is very likely one of the main reasons why complex shapes require more time to encode.

The effect of Typographic Variables

A common sentiment echoed by designers is that legibility is not just about the typeface but what one does with the typeface. Typographic variables (Fig. 5.11), such
as type size, line length, word spacing, letter spacing, leading, color and contrast, are all variables that can interact or individually influence the legibility of a typeface. For example, type size can affect legibility when text becomes very close to minimum acuity levels (Mansfield et al., 1996), and research has shown that common text sizes used in print today and historically have been influenced by the properties of our visual system (Legge & Bigelow, 2011). This becomes even more critical as we age, and older readers prefer larger text sizes (Bernard, Chaparro, Mills, & Halcro, 2003), most likely due to the drop in their visual acuity.

Of the body of work related to legibility research, the work of Miles Tinker stands out in terms of both quality and quantity. Tinker used speed of reading as a measure of legibility, and used comprehension checks to ensure that text was being read for meaning, all while testing the effects of various typographic variables. His view was that more legible text has a wider perceptual span, longer saccades, and less fixations. He defined perceptual span as the number of words in a line divided by the number of fixations on that line. This has since been proven to be wrong (R. E. Morrison & Inhoff, 1981). Morrison and Inhoff reviewed Tinker’s findings and showed that Tinker’s perceptual span are directly correlated with saccade length, and therefore mapped his findings of the perceptual span to saccade length (R. E. Morrison & Inhoff, 1981). This enables an analysis of the data with the benefits of more recent findings. Therefore, the following points follow their interpretation of Tinker’s data.

Tinker proposed that threshold tasks were not necessarily indicative of higher legibility in reading tasks. For example, uppercase letters were better discriminated at a distance, but functioned poorly in text settings. Text set in all caps resulted in longer reading times, and 12.4 more fixations. The saccades were longer in physical span, but in reality covered a smaller number of letters (7 vs. 8.1) (Tinker, 1963).

Increasing text size from 10 to 14 pts. resulted in reduced reading speed and higher number of fixations. The average fixation duration was reduced, but that was offset by the increase in fixations and reduction in the number of characters covered in a saccade (Tinker, 1963). Decreasing text size to 6 pts also had negative effects with longer fixations and shorter saccades (Tinker, 1963).

In terms of line length, very short line lengths resulted in longer fixations and a higher number of fixations. Very long lines resulted in decreased accuracy in executing return sweeps (Paterson & Tinker, 1942). Extra leading improved the reading of long lines (Tinker, 1963). When type is set with the optimal line length and leading, there was no difference between 9, 10, 11, or 12 pts. (Tinker, 1963). Reduced contrast between text and background also has a negative effect on legibility; it increases fixation times, number of fixations, and regressions (Tinker, 1963). Typefaces, then, have improved legibility "when the size of the letter is increased, when the lines in the letter are widened, when the area of white space around or within the outline of the letter is increased, when the contrast of shading and hair lines is lessened, and when the outline of the letter is made simpler" (Webster & Tinker, 1935).

Legibility, Complexity, and the Mark of a More Legible Design

This chapter is too long and with too much information for one to be able to adequately summarize all that it has covered. There are, on the other hand, several conclusions to be drawn and are essential to carry forward into the next chapters. These are the corner stones on which the next two chapters are built:

- Reading is affected by the linguistic qualities of the text, as well as the reading abilities of the reader, and the visual characteristics of the text.

- Typefaces have been shown to have a statistically significant effect on various reading measures.

- Legibility is relative, rather than absolute and it is affected by the visual characterstics of text, as well as the distance of the viewer from the stimulus and the task at hand.

- Legibility is the ease with which the reader can extract the visual information of text in order for lexical processing to begin.

- The marks of a more legible design are shorter fixation times, longer saccades, and a smaller total number of fixations.

- Legibility is negatively influenced by the complexity of the writing system and typeface style.

And so with this chapter, we are half way in drawing the arc of our narrative. We have seen how the Arabic script has developed from manuscript to typographic forms, and where it stands today. We have looked at the reading process and what makes a more legible design. Next, we look at how these concepts map out to the Arabic language and script. For that we turn the page.
Chapter 6
Arabic Language and Reading

The reading process is an interaction between the brain, the eyes, and a visual stimulus, which, for Arab readers, is Arabic language text on a page or a screen. The Arabic text in turn presents both visual and linguistic content. The first few chapters dealt with the nature of the visual presentation i.e. the Arabic script with its characteristics, components, and its evolution to its present day status. The previous chapter looked at the process of reading, and the marks of legibility in typeface design. This chapter deals with the linguistic content, the way it interacts with the visual stimulus, and the few studies that have been done to investigate the reading of Arabic. This chapter is the field where Arabic design and language interplay. As will be argued later on, the orthography and morphology of the Arabic language have a strong influence on reading. With that in mind, the chapter starts with a short introduction to the Arabic language that will turn out to be very handy once we get into the details of how it is all affecting reading.

Whether by design or twist of fate, Arabic text presents a rather puzzling combination of elements. To put it shortly, many Arabic words are formed as derivatives of a consonantal root. These derivatives are formed by changing vowel signs and the occasional addition of a few consonants. However, these same vowels are usually dropped out in written Arabic. As a result, a significant portion of a word's phonology is lost. Vital information is left out. So how does one read Arabic then? In short, it is mostly educated guesswork. And that is what this chapter is all about.

The Morphology and Orthography of the Arabic Language

There is a popular Arabic saying amongst linguists and it goes as: in order to read Arabic, you need to know what you are reading. Some sayings are exaggerated or sometimes even false. This one is not. It sums up in one sentence the hard truth about reading Arabic. Why is it so? It all comes down to how words are formed, i.e. morphology, and the fact that short vowels are dropped out of everyday texts. Excluding foreign words and transliterations, the majority of the Arabic lexicon of Modern Standard Arabic is built via two forms of structures: derivational morphology and inflectional morphology.
Derivational Morphology: Root + Pattern Model

Derivational morphology is the structure through which verbs and nouns are formed. These are based on triliteral or quadriliteral roots, also referred to as three or four consonantal roots respectively. The first term is more accurate as it is possible to have a root that includes consonants and a long vowel such as the root *kab* (related to saying). Words are formed when a phonological pattern is applied on a root. There are three types of patterns:

1. The first is a series of short vowels that are interspersed with the root consonants/ vowels (Abu-Rahia & Awwad, 2004). For example, take the pattern *maa* which indicates the action of doing for the past tense of third person male singular and apply it on the root *kab* which is the semantic field of writing. This yields the verb *kabab* which means, “he wrote.” This morphological structure and the resulting phonology do not break the orthographic order of the root since the short vowels are usually not indicated (in everyday texts it would be written instead of *kab*).

2. The second type of pattern is a combination of short and long vowels inserted within the root (Abu-Rahia & Awwad, 2004). One example is the pattern *maa* (with a long a) meaning the person who did the action. When applied on *kab* it gives *kabab* meaning writer. This type of pattern breaks the orthographic order of the root consonants just as in this example the Allt is inserted in between the first and second consonant.

3. The third type of pattern is a combination of short vowels with long vowels and/or certain consonants that are inserted between, in front of or after the root (Abu-Rahia & Awwad, 2004). One example is the pattern *maa* indicating the place of action which when applied on *kab* gives *makaab* meaning office or desk. This type of pattern also breaks the orthographic order of the root consonants. Derivational morphology in Arabic is then called non-concatenative since the morphemic units are interwoven into one another.

These patterns can be alternatively categorized as verbal vs. nominal patterns. There are fifteen frequently used verbal word patterns and nine nominal ones; while nominal patterns retain semantic consistency, different verbal patterns applied on the same root could convey different semantics (Abd El-Moneim, 1987). For example, nouns based on the root *kab* such as *kab* (letter), *makab* (desk), *kab* (small book), *kab* (book), *kab* (see), are all semantically consistent. On the other hand, verb patterns can form transitive (*kabab* - he wrote), intransitive (*makab* - something was written), active, and passive (*makab* - it was written) verbs from the same root (Abu-Rahia & Awwad, 2004).

Luckily, patterns are consistent in the meanings they generate whether it is an action, an event, a place, an object, or a person. Similarly, roots have embedded semantics in them. *Kab* is always related to writing, *Haa* to whispering, *Kar* to breaking, and *Kar* to studying etc. This is the key to learning Arabic. Even dictionaries are organized via roots. Once students learn the meanings implied within roots and patterns, they are then able to extrapolate that knowledge and apply it to learn new words. Thus, the root communicates more semantic information than the phonological pattern, and this points to the basic meaning of any given word; the actual pattern applied points to a word class as described earlier (Abu-Rahia & Awwad, 2004).

As for roots, there are strong and weak ones, the latter accounting for 10% of total Arabic roots (Maryati, 1987). Weak roots usually have a long vowel

1 Also referred to as non-linear or discontinuous morphology (Boudelaa & Marslen-Wilson, 2004b)

Inflectional Morphology

The second way words are formed is through inflectional morphology. Instead of roots and patterns, this is a process of attaching prefixes and suffixes to existing words. This can be applied to both verbs and nouns. In verbs, the system is dependent on person, number, gender, and time (Abu-Rahia & Awwad, 2004). For example, the verb *kabab* (she wrote) becomes *kababat* (she wrote) once you add the unvocalized t that denotes both the female gender and a singular number. Contrary to Latin, verbs and pronouns in Arabic also include a case for pairs. So, to conjugate the verb *kabab* in the past tense one adds a series of suffixes (coloured here in grey) to the past tense of the third person masculine singular version of the verb.

Third Person Masculine:
- *Kababa* - he wrote
- *Kababti* - (long a, Allt) - the pair wrote
- *Kababta* - they wrote

Third Person Feminine:
- *Kababth* - she wrote
- *Kababthu* - the pair wrote
- *Kababtna* - they wrote

First person:
- *Kababa* - I wrote
- *Kababna* - we wrote

In the present and future tense, verbs take on both prefixes and suffixes:

Third Person Masculine:
- *Yakabba* - he writes
- *Yakabht* - (long a, Allt) - the pair writes
- *Yakabba* - they write

Third Person Feminine:
- *Yakabba* - she writes
- *Yakabht* - the pair writes
- *Yakabba* - they write

First person:
- *Akabba* - I write
- *Akabba* - we write

(Where *b* or *y* is one of the components. These would act as a consonant in the basic root form (e.g., *Sya* meaning agreeing) but could disappear when some patterns are applied (e.g., *Itay*).) This fusion of one of the letters of the root and the letters of the pattern is referred to as gliding, which is part of an ablautic instance of several morphemes combining together to make more complex ones (Boudelaa & Marslen-Wilson, 2004b). The effect of such morphological play on lexical access and reading will be discussed later in the chapter.)
This delineation of gender and number is also indicated in the addition of pronouns to nouns. Pronouns are usually added as suffixes and usually follow a similar conjugation to that of verbs. For example, if one takes the noun kithā (book):

Third Person Masculine:
Kištāhu – his book
Kištāhumā – the pair's book
Kištāhum – their book

Third Person Feminine:
Kištāhā – her book
Kištāhumā – the pair's book
Kištāhum – their book

First person:
Kištā – my book
Kištānū – our book

These are some examples of inflectional morphology in Arabic verbs and nouns. Going back to the first paragraph in this section, one is now able to see how readers of the Arabic language are able to deduce the meaning of a word even though the vowels are not indicated. This is possible once one is familiar with the verbal and nominal patterns. This identification process kicks off once the context clarifies if the word is a verb or a noun, and one is then able to guess which pattern best fits the written consonants and vowels. This is why it is said that one needs to know what one is reading in order to be able to understand the meaning. The role of the root and pattern in lexical access, and their subsequent effect on reading and word recognition will be discussed later on in this chapter.

Etymons

Though the general consensus amongst linguists bears that the smallest morphological unit in the Arabic language is the triliteral root, a few authors have argued that the smallest unit, called etymon, is actually comprised of only two unordered consonants. The argument for the etymon points to the existence of a similar semantic reference in different roots that share two of their three consonants. Examples cited include batta (cut off), batar (separate), batata (separate) (cut off), Bata (separate), Bata (separate) (cut down) (Boudela, 2001). These words have only two consonants (d and b) in common and the root-pattern theory would classify these as being based on five different roots (bta, btr, bta, bat, bat), even though it is obvious that they share a phonological and semantic reference.

Bohas (1997) argues that this is a commonly occurring paradigm in the Arabic lexicon and the root-pattern as the smallest morphological unit fails to explain this regular occurrence. Boudela and Marslen-Wilson (2001, p. 68) expand on how the process forms:

"In order to have a surface form the bi-consonantal etymon morpheme is morphologically expanded by the addition of an ephemeristic segment (i.e. a segment inserted as a result of a phonological process) specified as non-syllabic, which covers the 27 consonants of the language."

Furthermore, the authors claim that the etymon theory manages to explain a feature of the Arabic language where two semantically similar roots share two of their three consonants but in a different order. This allomorphic variation occurs in examples such as māta (do) (perish) and tannīs (come to an end). The respective roots are [m]a and [t]n and the suggestion is that the etymon [t]n is common to both and that the reverse order does not change the basic meaning of the etymon. (Boudela & Marslen-Wilson, 2001). This, it is argued, holds true for 135 of the possible 235 etymons in Modern Standard Arabic (G. Bohas & Darfool, 1993). Though this might seem logical at first glance, further arguments start to border on being somewhat far-fetched:

"A second instance of allomorphy is in cases where the segmental structure of the surface form is different across the realizations of a family of related forms, argued to be linked back to the same underlying etymon morpheme. An example is the set of forms [Salbara] (blind), [Bhaba] (keep under lock), [baba] (tie up), [baba] (knob), [Habasa] (hold back), [haba] (bind), [haba] (a rope), [Affa] (refrain), where the etymon consonants are in bold, and where they all share a core meaning related to notions such as (restraint) and (tying up). The underlying phonological commonality between these forms is that they all consist of a featural combination of a [labial] consonant and a [pharyngeal or pharyngealized] consonant, suggesting that the abstract specification of the form of the etymon is in featural rather than segmental terms. (Boudela & Marslen-Wilson, 2001, p. 68)

The above argument seems to try too hard to find supporting arguments for the existence of the etymon. Indeed, the etymon is no longer composed of two common consonants but of two phonetically related consonants. The authors give an example of how this system works:

"Thus, in the forms [haba] (cut down), [batara] (cut off) the etymon is [b] rather than [t] or [h], because the meaning (cutting) recurs in other forms containing the two consonants [b] and [t] as in [haba] (cut) and because this is consistent with the featural specification of the etymon as [-labial] and [-dentate]. By the same token, the forms [basa] (tie) and [baba] (bind) are morphologically related because they share the consonant [b] and their respective [s] and [t] consonants are homorganic in that they are both pharyngealized." (Boudela & Marslen-Wilson, 2001, p. 68)

Opponents of the etymon theory point to the fact that the above paper by Boudela, and the study that it is carried, is only concerned with the psychological reality of the existence of the etymon but does not offer any theoretical framework in which the etymon could be isolated or identified (Benin & Frost, 2001). Indeed, it turns out that the identification of the etymon in any given word is a quite complicated and difficult task, which is in strong contrast to the ease and simplicity with which the trilateral root can be identified.

Benin and Frost's argument goes on to say that the "there are no a priori and clearly defined rules for morphological decomposition that would result in the unequivocal stripping or isolation of the etymon letters or phonemes" (Benin & Frost, 2001, p. 114). Given that, the alternative approach was to test if native speakers were sensitive to the etymon and able to extract it. Benin and Frost (2001) put this to the test and the result was that only 68% of the subjects were able to identify the
etymology, this was considered to be a low figure especially given the fact that there are only three possible two-letter combinations in a three-letter root. Bentin goes on:

"The weakness of this performance is even more conspicuous considering that native speakers of Semitic languages can easily report without error what letters of the word belong to the root, even at the first grades of primary school" (Bentin & Frost, 2001).

Boudela's study does offer interesting evidence to support the existence of the etymology. The study found that when testing word pairs, the lexical decision performance was significantly faster when the target and the prime shared a common etymology. Bentin acknowledges these results but offers an alternative theory that "these effects reflect knowledge accumulated implicitly by exposure to the statistical regularity that exists between orthographic and phonological sublexical units and semantic features" (Bentin & Frost, 2001, p. 115). Given the lack of a supporting theoretical construct, or wide ranging and supporting field data, the question of the etymology and its role as a morphological unit remains to be confirmed. The trilateral root remains as the widely accepted smallest morphological unit in Arabic.

Homographs

As mentioned in previous sections, Arabic texts in every day situations are usually un-vocalized. It has also been demonstrated how any given root can be expanded into many different words depending on the pattern that is applied on it. The pattern, as already shown, is a sequence of vowels, as well as possible consonants that are added onto the root. There are several patterns that affect the consonants of the root similarly, and only differ in the vowels they deploy. Once these vowels are dropped out of everyday text, the effect is that Arabic texts have a large number of homographs: words that look similar, but are vocalized and therefore pronounced differently.

Some example patterns that do not add any consonants to the root are:

|a-s-a| he did an action
|a-e-s| an action
|a-s-a| a noun resulting from the action

Indeed, one can find many examples of this sort, more than there is space for here. A good example of homographs related to the root *mlik* is *milkan* (king), *milkan* (property), *malaka* (he owns), *malaka* (he gave property to someone), *malaka* (it was owned), *milka* (he was given property). Without any vocalization, all these words are written as: *ملك*.

In the previous examples, two of the patterns actually do affect as the consonants as they double (stress) the middle consonant. However, the doubling of the consonant is noted as a mark above the character. This is part of the vocalization system. Though it is linguistically incorrect not to show it, this is often the case.

Orthography

The Arabic script is partially connected, and as mentioned already, the vowels are of 2 kinds: short and long. Short vowels are usually dropped out in everyday texts and are included in texts for children, beginning readers, poetry and liturgical texts. Long vowels, of which there are three, are usually pronounced as long vowel sounds though they do on occasion function as consonants as well. Every consonant corresponds to one sound only (though sound can vary across regional accents) and, with the exception of the "Tej niqba" and the "Laam" in the "Ali-Lam" article with certain nouns, these consonants are always sounded out, and the relation of grapheme to phoneme in the Arabic language is consistent. The Arabic language, then, employs shallow orthography when fully vocalized and deep orthography when it is not.

Another important aspect of Arabic orthography is related to the shape of the script itself, and that is the abundant use of dots and the reliance on very few basic shapes in the script make-up. This makes the Arabic script quite challenging to read for beginning readers. However, this is an aspect of the script that cannot be tested or so easily changed. It would not be possible to set up an experiment where one typeface is dotted and the other is not. Because of that, the issue of the dots does not come up in psycholinguistics research and is absent from the following pages.

**Literature Review of Findings Related to the Reading of Arabic**

**The Nature of the Arabic Spoken and Written Language**

As mentioned in chapter 1, spoken Arabic differs significantly from the formal written version. Dialects across different regions are so different that it is sometimes very difficult to understand what another Arab might be saying. However, the written Arabic language, Modern Standard Arabic (MSA), is the same across the entire Arabic speaking region. Therefore, while it might be difficult for a Lebanese to understand what an Egyptian is saying, they will both read the same books and the same newspapers. MSA is used for written communication as well as formal occasions such as political speeches and the news on TV and radio.

The spoken Arabic dialects are different from one another and very different from MSA. A study that looked at the ability of MSA and spoken Arabic to prime one another found that these two are functioning as two separate languages (Ibrahim & Abaran-Perez, 2005). Children going into school would have been exposed to spoken Arabic at home and in daily activities and, as such, it is their first language. In preschool years, they are exposed to MSA through TV programs or books read aloud to them, but the first language they learn is the local dialect of Arabic. In fact, in the study cited above which was done with Arab Israeli readers, MSA and Hebrew had the same priming effects for spoken Arabic leading to the conclusion that MSA is retained as a second language in their cognitive system (Ibrahim & Abaran-Perez, 2005). Furthermore, MSA and spoken Arabic are stored as distinct lexica, as was shown in a naming study, though there exists strong connections between the two languages due to the perception that they are different forms of the same language (Ibrahim, 2006). Given this situation, Arabic is then a case of diglossia due to the differing nature of the spoken and written forms and because the specific "socio-functional" situations in which these two forms are used (Saleh-Haddad, 2005, p. 562). Not only are these two forms different in their lexicon, but also in their phonetic make-up: spoken Arabic has fewer consonants and more vowels (Saleh-Haddad, 2005). The nature of syllables is also different: in spoken Arabic, words can start with a silent
consonant or can have two silent consonants in a row, both of which are not allowed in MSA. As to the difficulties facing a child learning to read what is practically a new language, Salem-Haddad explains:

"In order to identify a word, the beginning reader must be able to discover the linguistic relatedness between the two forms of the word and to recover the linguistic distance between them. This is a formidable task, especially given the fact that phonological distance is usually compounded by morphosyntactic distance. Also, because almost all function words and many of the high frequency content words that (s)he encounters have a phonological form in MSA that is completely different from their form in the child’s spoken vernacular." (Salem-Haddad, 2005, p. 563)

This leads to a situation where a child is reading in a shallow orthography (fully vocalized Arabic) that nevertheless contains uncommon phonemes, which lead the author to hypothesize that "phonological processing for MSA phonemes would be more difficult for children than that for SAV phonemes, and that this would be related to their reading fluency." (Salem-Haddad, 2005, p. 564). The author also points out that the new sounds that a young child encounters in MSA makes vocalized Arabic less of a shallow orthography than what one initially expects. As it turns out, "the strongest predictor of reading fluency in vocalized Arabic was letter recoding speed. Letter recoding speed was predicted by memory, rapid naming, and phoneme isolation" (Salem-Haddad, 2005, p. 559). In other words, even though Arabic is a case of diglossia, the most reliable factor that influences reading is the ability to connect the letter to its sound. Luckily for young Arab children, studies have shown that exposure to MSA in preschool years can yield to better reading comprehension once they start school (Abu-Rabia, 2000). It is fortunate, then, that with parental support at home, young children will have an easier job once confronted with the task of reading what is a new language to them.

Another study that tested Arabic and Hebrew monolingual readers vs. Hebrew-Russian bilinguals found that Arabic readers had similar reading abilities to bilingual readers which is further support for the argument that MSA is at the level of a second language to Arabic speakers (Ibrahim, Estivariz, & Abaron-Perez, 2007). In fact, the authors argue that exposure to literary Arabic (MSA) requires the same levels of linguistic analysis as the exposure to two languages that are as different from one another as Hebrew and Russian. Furthermore, in another study, Arab children showed similar linguistic abilities like the Hebrew-Russian bilinguals: better phonological skills, but lower on vocabulary scores (Eviatar & Ibrahim, 2000).

Learning to Read

In spite of all the odds, and the pessimism one feels when faced with the idea that one's native reading language is still a foreign language, Arab children still manage to learn to read and write. As to how they manage to do that, one study looked at whether the dual route approach to reading is also applicable to Arabic. The results showed that Arab children go through a similar process as other children do in learning how to read English (Taouka & Coltheart, 2004):

- Discriminative-rect phase: This is the stage where children know a certain number of sight words that they recognize via specific visual features (Taouka & Coltheart, 2004).

- Phonological-rectreading phase: In this stage the children have a much larger vocabulary that they can actually read so they will read words letter by letter, using their knowledge of grapheme-to-phoneme conversion to arrive to the correct pronunciation that will help them recognize the word itself (Taouka & Coltheart, 2004).

- Orthographic phase: By this stage, children will have arrived to the point in time where they are able to directly rely on orthography in order to read (Taouka & Coltheart, 2004).

The study also showed that the ability to use the correct contextual form of a letter is a skill that is learned late in the reading stages, but once learned this knowledge becomes so "deeply engrained" that it is difficult for adult readers to read Arabic text that does not employ the correct contextual forms (Taouka & Coltheart, 2004).

In terms of difficulties for beginner readers, it seems that there are certain types of misreadings and misspellings that are consistent through the early years of education. In effect, errors are often caused by misreading the short vowels, by writing the inappropriate contextual form of a letter, and by omitting or adding extra letters in writing exercises (Azzam, 1993). The frequency of the errors diminishes as children grow, but the types of errors remain (Azzam, 1993). Other studies have also found that at least 50% of spelling errors are phonological in nature (Abu-Rabia & Taha, 2006).

As to bilinguals, a study found a "cross-linguistic relationship between phonological awareness" in bilingual Canadians who spoke both Arabic and English, even though these languages are written in completely different script systems (Salem-Haddad & Geva, 2008). The same study did not find a relationship between morphological awareness between Arabic and English leading to the conclusion that "morphological awareness is primarily a language-specific linguistic skill that emerges as a function of language proficiency, and is therefore relatively independent in the two languages of bilingual children." (Salem-Haddad & Geva, 2008, p. 15).

Effect of Morphology on Reading

Are words recognized as whole entities or via the root? As mentioned in chapter 5, the role of morphology in lexical access is dependent on language itself and that it is often the case that there is a race between whole-word access and decomposition into the route. If words are stored in the mental lexicon as whole entities, then there should be no difference in the identification of morphologically simple or complex words (Abu-Rabia & Awad, 2004). Abu-Rabia and Awad tried to see if the root would act as a prime for other words derived from the same root, or if the word pattern would prime other words derived from the same pattern and used the naming of high-frequency words and non-words as procedure. It found that neither the root nor the word pattern acted as a prime for their derived nouns leading to the conclusion that "the nominal derivational morphology of Arabic words is represented in the

---

2 SAV stands for Spoken Arabic Vernacular
mental lexicon as separate whole words, and the nature of the morphology exerts no influence on the process of word recognition" (Abu-Rahla & Averd, 2004, p. 332).

As to why this is a different case from Hebrew, the authors note that the partially connected nature of the Arabic script makes it different from Hebrew and a complex cognitive task to perform and so readers overlook the decomposition of a word into its root (Abu-Rahla & Averd, 2004). However, as discussed in chapter 5, highly frequent words are accessed as whole even in languages that are not as reliant on morphology as Arabic, and this study used high frequency words in its setup. In such a case, their statement that "The main conclusion of this study on the morphology of Arabic is that roots and word patterns have no essential role in word organization in the mental lexicon" (Abu-Rahla & Averd, 2004) would only be applicable for the recognition of high frequency words. This is an issue that will be further elaborated on in the next paragraphs.

But before we do that, there is another element of morphology in Arabic that is the less commonly discussed which is the CV-structure, where C and V refer to consonants and vowels respectively. The CV-structure is the specific pattern of consonants and vowels that result from the application of a pattern on a root (McCarthy, 1981). McCarthy lists an inventory of the possible canonical forms in Arabic:

1. C V C C V C
2. C V C C V C
3. C V V C V C
4. C V C V C V C
5. C V V C V C V C
6. C V C V C V C
7. C V C V C V C
8. C V C V C V C

A study (Boudela & Marslen-Wilson, 2004a) investigated the priming of words that share the same vowels (as dictated by different patterns), words that share the same CV-skeleton, and words that share both (in essence the same word pattern). It found that the priming by words that share the same CV-skeleton was stronger than that by words that shared the same vowels, and had the same strength as that induced by the word pattern. The authors point to the fact that this CV-skeleton priming was done by words that did not share the same consonants or had any semantic relationship (Boudela & Marslen-Wilson, 2004a).

The same authors further investigated the changing role of morphological priming over time of display (Boudela & Marslen-Wilson, 2005). With the use of masked priming with display durations (in milliseconds) of 22, 48, 64, and 80 ms, the study found reliable root priming at all SOAs. They also found priming effects for the word pattern but only at 48 and 64 ms for nouns, and 48 for verbs. This led the authors to conclude that there is a likely difference in the time needed to extract word pattern vs. root information. The reasons proposed are very interesting to note:

"In other words, while the visual event presents fully specified information about the consonantal root, it presents only partial information about the root pattern. This means that while accessing the meaning of an orthographically presented root can be direct, the mapping of a visually presented word pattern onto its morpho-syntactic meaning is indirect and may require mediation through access to phonology. If this is correct, then it is phonological mediation that results in word pattern priming kicking in after root priming. Another factor underlying the differential priming outcomes is that root patterns and roots may be simply the nature of information conveyed by the two units.

In particular, the meaning conveyed by the root is arguably more constraining than that conveyed by the word pattern. Surface forms sharing a root make up a more coherent morphological family than those sharing a word pattern, and this may also facilitate access" (Boudela & Marslen-Wilson, 2005, p. 231).

In other words, the root is always a prime because its letters are almost always preserved in Arabic orthography. On the other hand, the pattern is often a combination of short vowels, and a few long vowels and consonants. Since the short vowels are usually not indicated in normal texts, the reader will often need to rely on these long vowels and consonants to guess what pattern is applied and therefore how to correctly read the word. If the pattern is composed of only short vowels, then it is only the sentence context that will clarify which pattern is applied.

As to orthographic and semantic priming, the authors found evidence of that from only 80 ms onwards, which again further supports their conclusion that "morphological effects in Semitic languages represent distinct structural characteristics of the language" (Boudela & Marslen-Wilson, 2005, p. 207).

As mentioned earlier in this chapter, Arabic roots can be either weak or strong depending on if one of its letters undergoes allomorphic changes or not (ٌ/١٠/). Interestingly, weak roots have the same priming capabilities as strong ones, and roots show priming abilities even when the semantic relationship between prime and target is not very obvious (Boudela & Marslen-Wilson, 2004b). Patterns are also able to act as primes, with the exception of the cases when they induce allomorphic variations that change their underlying CV-skeleton (Boudela & Marslen-Wilson, 2004b). The same study also found that Arabic words that are neither semantically related nor modeled in the same pattern but are phonetically similar nevertheless do not have priming effects (Boudela & Marslen-Wilson, 2004a).

An interesting study by the same authors looked at priming effects of roots and patterns in terms of their productivity (or family size), both of which are used to indicate the number of possible words this root or pattern will produce. As part of the experiment description, the authors list some interesting statistics of Arabic roots and patterns. There are around 6000 roots in Arabic and the average productivity is quite low (12 derived words for the most frequent 3000 roots) and the highest productivity ranging between 30 and 40 derived words. As for patterns, there are 155 nominal patterns (in the sample selected by the authors), but their productivity is much higher and is on average around 60, but can go up to 434 (Boudela & Marslen-Wilson, 2004b). They tested the ability of the word pattern to prime, and tested that against the productivity of both roots and patterns.

What they found was that there were strong priming effects for less productive word patterns that were applied on productive roots. Also, if the root is not productive, then the word pattern would not prime, irrespective if that pattern was productive or not (Boudela & Marslen-Wilson, 2011). In other words, if a word is derived from a root that has a large set of possible derivatives, and its pattern is infrequent, then its priming power is stronger. Also, if it is part of a small set of possible derivatives, its priming power is weak, and the productivity of its pattern is not relevant. As to why this is so, it is possible that a more productive root is simply more frequent as it is the entry point for so many words in the Arabic lexicon.

Furthermore, the results show that the amount of pattern priming is dependent on the family size of the priming root, but not of that of the target root (Boudela

4 Roots that have a large number of word derivatives would be highly productive and vice versa.
A Marj-leen Wilson, 2011). This led the authors to conclude that the processing of the word pattern is dependent on the processing of the root and a delay in root processing will delay the processing of the pattern. This is also supported by their previous investigations as to the SQA time frames of when the root and pattern are able to appear.

Going back to the issue of the root frequency, is the root, which is an abstract string of consonants with a specific semantic reference, stored as a lexical unit, or is the Arabic lexicon made up of only whole words? Evidence for the first came from an Arabic-French bilingual who suffered a stroke that resulted in a language deficit where the subject’s comprehension and ability to read aloud were impaired (Prunet, Beland, & Frisul, 2000). A study of the errors that this subject made led to some very interesting observations. His impairment was characterized as deep dyslexia, and his language tests showed a consistent consonant metathesis error; two consecutive consonants of the 3 consonants found in the root would be switched around, but the pattern applied remained unchanged. These errors occurred at 23 times the rate of errors while reading French. The interesting aspect of these errors in Arabic is that the consonants that are added in via the word pattern were never switched around and that the CV skeleton remained unchanged. This led the authors to conclude that the metathesis is happening only with the root which means that the root is stored as an independent unit in his lexicon (Prunet et al., 2000).

Another support for the role of the root in Arabic morphology came through a study that looked at name shortening or hypocoristics, such as the name Hassan being turned to Hassoun (Davis & Zawdeh, 2001). The study points to the prevalence of the consonantal root in these formations, at the time that the extra consonant added by the word pattern do not surface. An example of that is Salmah (root is /slm/) being shortened to Sallam where the n added in by the pattern does not appear in the hypocoristic (Davis & Zawdeh, 2001). Moreover, the ability to correctly identify which words share a common root (morphological identification), and the ability to name more words that share the same root (morphological production) have been shown to be reliable indicators of reading comprehension (Abu-Rabia, 2006).

Looking again at bilinguals, Arabic and English bilingual children who had high proficiency in decomposing morphologically complex Arabic words had higher oral reading proficiency in Arabic, while this relationship was not present in the reading of English (Saleh-Haddad & Geva, 2008). Bilingual children who had poor skills in reading English and Arabic still showed better spelling and pseudoword reading skills than monolingual children who only read English (Abu-Rabia & Siegel, 2002).

Given the limited amount of research regarding Arabic morphology, it is exciting to research done for Hebrew. The similarities between the two are many: they are both Semitic languages with similar root and pattern based morphology (Deutsch, Frost, Pollatsak, & Rayner, 2000), both written from right to left, and both drop out vowels in every day texts. There are of course obvious differences: the Hebrew script is not attached, and the gap between spoken and written is not as wide as in Arabic.

The findings related to Hebrew are also similar to those in Arabic: The root, or any word derived from it, has a strong priming effect in both naming and lexical decision tasks of words derived from that same root (Deutsch, Frost, & Foster, 1998). The word pattern is also able to act as a prime, though only for verbs and not for nouns (Deutsch et al., 1998). These results led to the conclusion that that the Hebrew lexicon includes:

"... a multiple system of connections between a whole-word level (nouns and verbs) and a sub-word morphological level, which consists of root and verbal-pattern morphemes. By this view, all word units, whether nouns or verbs, are connected to root morphemic units. In addition, verbal forms are also connected to verbal pattern units. This organization is independent of semantic factors" (Deutsch et al., 2000, p. 491).

Moreover, if the root was shown in the parafoveal, it is able to give a preview benefit for the target word n+1 (Deutsch et al., 2000). When readers were fixating on word n, the following word would be the root. However, once the readers move their eyes towards that word, the display changes, and they will see the target word, which in this case is derived from the root shown before. This is the typical boundary technique and is a good test of the priming abilities of the root in Hebrew. Even when the readers did not move their eyes, and the word currently fixated was replaced by a word derived from a root previewed in the parafoveal, the root facilitated lexical decision making (Deutsch et al., 2000). The authors also find on a very important note: the root gave a preview benefit, even when its letters were interspersed all over the target word (Deutsch et al., 2000). In essence, this is not about two words that start with the same letters and thus act as primes for one another; this is a unit in a non-concatenated morphology demonstrating its role in lexical access.

This key role was further demonstrated in a later study by the same authors. Using a similar experimental setup and using eye tracking to measure the amount of preview benefit, the study investigated the ability of the root to act as a prime, though this time in a sentence context (Deutsch, Frost, Pelleg, Pollatsak, & Rayner, 2003). Also different in this experiment was that the priming word was not the root, but rather another word derived from it. In the effect, the test was looking to see if two words derived from the same root can prime one another, and the results were positive (Deutsch et al., 2003). Another interesting aspect that the authors point to is that the priming effect of the derived word happens early on in lexical access. The preview benefit, as discussed in chapter 5, happens in the pre-attentive phase of sentence reading. That is to say, this parafoveal preview is taking place before the eyes actually fixate on the target word. As such, and as the authors have concluded, morphological effects are very early in the Hebrew word encoding process (Deutsch et al., 2003).

However, further investigation into the type of derivational morphology that is able to give a preview benefit showed that words derived from the same pattern are able to facilitate the processing of one another, though only in the case of verbs but not nouns (Deutsch, Frost, Pollatsak, & Rayner, 2005). The sensitivity to the word pattern extends to the syntactic role that the derived word fits into. In a sentence context that requires a noun, a verbal preview delayed the reading times of a noun derived from the same root, thus demonstrating that the morphological information gained in the preview is modulated by the sentence context expectations (Deutsch et al., 2005).

While comparing the results of parafoveal previews of roots and verbal patterns, the authors arrive to the conclusion that the root is analyzed at an earlier time frame than the verbal word pattern (Deutsch et al., 2000), a result which is similar to that found for Arabic and discussed in previous paragraphs.

Morphology, then, is a big player in the process of reading semantic languages such as Arabic and Hebrew. The role of roots and patterns is very clear in terms of their ability to prime, in the time frames they operate in, and the extent with which they define the Arabic lexicon. This is not to say that all questions have been answered or that all issues are clearly settled. However, within the field of study of Arabic reading, the influence of morphology is the one that has been most thoroughly tested.