

On Developing Plain Language Principles and Guidelines

Karen A. Schriver, PhD

© 2014 Institute of the Estonian Language

Citation:

Schriver, K. A. (2014). On developing plain language principles and guidelines. In K. Hallik & K. H. Whiteside (Eds.), *Clear communication: A brief overview* (55-69). Tallinn, Estonia: Institute of the Estonian Language.

Contact Information:

KSA Communication Design & Research, Inc.
33 Potomac Street
Oakmont, Pennsylvania 15139 USA
412.828.8791
kschriver@earthlink.net

On Developing Plain Language Principles and Guidelines

Karen Schriver



Karen Schriver, Ph.D., is a researcher and consultant in information design and clear communication. She is president of KSA Communication Design and Research. Her book *Dynamics in Document Design* has been used in corporate, government, and academic environments around the world. She is currently working on a new book about evidence-based information design and plain language.

Plain language arguments can be difficult to make, particularly when we are trying to persuade people who are convinced they do not need plain language because to them clear writing and design are obvious and require little if any skill. A good way to develop cogent arguments about the nature of plain language activity is to draw on the empirical literature about how readers respond to the writing and design of texts and graphics. Here plain language advocates need to pay close attention to studies that inform textual choices at the word level, sentence level, and the whole-text level. We must also attend to research on the impact of visual design on reading, especially studies of typography, layout, graphics, and visual impression (Schriver, 2013a, 2013b).

Although the research up to this point does not tell us the whole story regarding what makes content comprehensible and usable, the good news is that there is already a considerable amount of empirical research on writing and design – important work that plain language advocates can draw on in nurturing their expertise (Schriver, 1997, 2012). I have been reviewing the research literature on information design and plain language – integrating the findings from hundreds of studies of reading, writing, and visual design (Schriver, in preparation). An important question lies in how to draw on this research in developing plain language principles and guidelines. Here I suggest a strategy.

Drawing on the existing research: “Consolidate and identify next steps”

As the international plain language movement progresses toward making recommendations about plain language principles and guidelines, it will be crucial to build on what is already known. This will require synthesizing the existing research and extracting from it useful advice. As we do so, it will be important not to oversimplify the inherent complexity of writing and design. My reading of the literature suggests we could develop two types of principles for plain language and information design: universal and conditional. Some research findings will allow us to develop universal principles (that is, guidance that applies to most cases) while other research findings will lead to conditional principles (that is, guidance that applies to particular contexts).

Consider universal principles. Here we want to know the essential characteristics of reading, writing, and design that need to be kept in mind as we write and design. For example, studies of human cognition tell us much about fundamental aspects of how people perceive, search, read, and interpret. Studies of cognition can help us derive universal principles for plain language and information design since this research reveals what people have in common as they read, write, design, or engage with communications. We can then interpret these principles and create guidelines suitable for practical action.

Similarly, there is a growing body of research that can inform the design of conditional principles. This work can help us specify guidelines about writing and designing for particular countries, languages, audiences, situations, genres, domains, or subject matters. For example, case studies of particular genres can shed light on strategies and heuristics for improving the effectiveness of communication for that genre. For some recent case studies, see Schriver (2014).

Of course, some of the issues we care about have not yet been studied (or have been studied only with native English speakers interpreting English language texts). Even so we can consolidate the available literature with an eye toward developing a provisional set of principles and evidence-based guidelines. We can also identify gaps in the literature and articulate the unresolved questions from our perspective. This will allow us not only to specify important research directions to address, but prompt us to form alliances for initiating new research on the complex issues we still face.

As a way of illustrating how such a “consolidate and identify next steps” strategy might work, I offer a snapshot of one area of the research literature: *word-level textual choices*. Next is a summary of the empirical findings on some of the word-level text features that influence how people read and understand. Then I suggest what is missing from this picture and what still needs further exploration. I conclude with some ideas about how to use this research to form universal and conditional principles that lead to evidence-based guidelines.

Case in Point: Consolidating Empirical Findings about Word-level Choices

Research on the impact of word-level features on reading has been carried out since the 1940s. Studies have focused on issues related to how word choice influences readers' understanding of content. Researchers have identified a number of word-level features that influence the clarity of writing, features that sometimes enhance understanding, while other times impede it. Among the most studied word-level textual choices include the following:

- Word length
- Word frequency
- Word difficulty
- Word concreteness
- Nominalizations
- Noun strings

Let's look at each in turn to see how they may impact readers' comprehension of text.

Word Length

Studies of reading tell us that word length is an important predictor of comprehension (Miller & Selfridge, 1950). Research shows that short words are easier to recognize, faster to interpret, easier to learn, and better remembered than long words (Carpenter & Just, 1981). Plain language advocates who want their readership to retain the content should avoid embroidering the key ideas with long words.

Implications. Research makes clear that choosing short words rather than long words has substantial benefits for readers. Short words are good for all people – whether they are skilled or unskilled in reading – and plain language advocates do readers a favor when they use them.

Word Frequency

Since the early 1900s reading researchers have been concerned with the impact of word frequency on understanding. Researchers hoped to assess the relative ease of texts by using word frequency to index the difficulty of the text's words. Some of this work led to the development of lists of the most frequent words in English (Leech, Rayson, & Wilson, 2001; Lorge, 1938; Thorndike, 1921; Thorndike & Lorge, 1943).

Over the years, psychologists and reading researchers have done many comparisons of high-frequency words and low-frequency words (Hudson & Bergman, 1985; Nagy & Anderson, 1984). Not too surprisingly, they find that high-frequency words are recognized faster and require less attention than low-frequency words (Carpenter & Just, 1983).

Studies show that the more memory consumed by recognizing the words in a sentence, the fewer cognitive resources the reader has for higher-level processes, such as making inferences about what the text means (Just & Carpenter, 1980). More recently, information architects such as Maurer (2006) have drawn on Lakoff's (1990) research on the cognition of categorization to suggest that short high-frequency words often represent fundamental categories and thus have good "information scent", helping people make interpretations more quickly (Lakoff, 1990; Mauer, 2006).

Implications. The research shows clearly that plain language advocates should use high-frequency words when they can. Of course, there will always be situations in which a sufficiently precise high-frequency word can't be found. In such situations, writers should choose the word with the highest frequency that expresses their meaning.

By implication, research on word frequency suggests that organizations should think twice before coining new words and creating acronyms because they are necessarily low frequency, and their unfamiliarity will likely slow people down, both in searching and understanding. Studies are consistent in suggesting that familiar words should be used in headings, labels, captions, and links to speed the recognition and retrieval of content.

Word Difficulty

Studies of word difficulty focus on differences between simple and complex words. Most researchers define simple words as "easy to pronounce"

or words with “few syllables”; and hard words as “difficult to pronounce” or words with “many syllables” (Bauer & Stanovich, 1980; Stanovich & Bauer, 1978). When readers encounter words they find difficult, their comprehension and speed of understanding suffers.

Implications. Writers can conclude that a simple word is almost always a better choice than a complex one. By implication, this research suggests that organizations should avoid complex multi-syllable words as well as words that native speakers may find tricky to pronounce. This does not mean avoiding precise words. Readers can only acquire an accurate understanding of the content when the text is comprised of words that render the content accurately. It means that writers should strive for lexical choices that are precise while at the same time simple and familiar to the general public.

Word Length, Frequency, and Difficulty

Fortunately, these three factors – short word length, high frequency words, and word simplicity – are positively related to each other. It turns out that the short words are usually the frequent words and the easy words. A consistent finding is that short words tend to be high-frequency words, those words native speakers of a language hear all the time (Dobbs, Friedman, & Lloyd, 1985). Put differently, words that appear frequently in a language are usually short words, helping people communicate more quickly (Zipf, 1949).

Consequently, short high-frequency words are recognized quickly by readers and require little attention to comprehend in comparison to low-frequency words (McNamara, Louwerse, Cai, & Graesser, 2005). In fact, short low-frequency words are easier to recognize than long low-frequency words (Hudson & Bergman, 1985).

Some research into word-level issues culminated in models of how frequently words appear in a language (Kucera & Francis, 1967). For example, Zipf demonstrated that in many languages there is a statistical relationship between the hard and easy words, such that easy words appear much more frequently, accounting for most everyday speech (this relationship is called “Zipf’s law”). Zipf showed that in everyday communication, people tend to choose words that require the least effort, coining the “least effort principle” to describe people’s tendency to take the shortest route to stating an idea (Zipf, 1949).

In addition, research shows that simple words also tend to be high-frequen-

cy words (Hudson & Bergman, 1985; Zipf, 1949). More recent cross-linguistic studies of the relationship between word length and word frequency were carried out using distributions of words in languages such as Spanish, Russian, Portuguese, Chinese, Swahili, Estonian, Czech, Maori, Turkish, and German with the same result (Calude & Pagel, 2011).

Word length, word frequency, and word simplicity have all been incorporated into readability formulas. Klare points out that word frequency plays such a central role in what makes text difficult that it became a basic part of readability formulas (Klare, 1968); for a discussion, see (DuBay, 2004). Of course, the validity of readability formulas for predicting the understandability of text has been shown to be seriously problematic. For example, reversing the order of words in a sentence makes it completely unintelligible but does not change its readability score. Clearly, there is more to intelligibility than just word-level factors. Research makes clear that usability testing provides a much more valid metric than readability formulas (Huckin, 1983; Redish, 2000; Schriver, 2000).

Even so, that does not mean that research on word frequency and its correlation with text difficulty should be dismissed. Because the correlation has not gone away, and hard words still play a role in comprehension, word frequency has been incorporated into recent comprehensibility formulas that take into account text features derived from current psycholinguistic research and traditional readability measures (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara et al., 2005). Although word length, word frequency, and word difficulty contribute to the comprehensibility of text and certainly need to be considered while we compose, there are other factors beyond the word level that contribute to how usable a text will be for its readership.

Word Concreteness

An important line of research on how people understand words is concreteness versus abstractness (e.g., the difference between understanding “apple” and “liberty”). This research compares abstract and concrete words and finds that learning and remembering concrete words is easier (Graesser et al., 2004; McNamara, Crossley, & McCarthy, 2010; Schwanenflugel, Harnishfeger, & Stowe, 1988). Studies show that because concrete words (usually nouns) often evoke more synonyms and more visual imagery than abstract words, readers have an easier time retrieving their meaning (Kroll & Merves,

1986; Sadoski & Paivio, 2001). Here is a quick example:

Abstract: To achieve excellence in plain language, advocates must work hard in acquiring knowledge of the stylistic norms of written language.

Concrete: To excel as a plain language advocate, you will need dedicate yourself to mastering the conventions of a plain writing style and develop an ear for vivid and memorable prose.

Concrete words also give writers more opportunities to make ideas more vivid because keywords can be substituted with synonyms over consecutive sentences. If the keywords that comprise a text's main points are concrete, the idea can be embroidered over paragraphs with other concrete words that are semantically linked, making it much more likely that readers will understand and remember the main points (Cox, 1978; Paivio, Yuille, & Madigan, 1968).

Implications. If the subject matter is inherently complex, detail the main ideas with concrete words that are as vivid as possible. In elaborating an abstraction, do so near the point at which the abstraction appears; that is, at the point the reader needs it. (In composing online, use a "roll-over" for the detail, particularly if the elaboration interrupts the text.) Research also suggests that if writers cannot find a way to make the abstraction itself more concrete, they might generate a concrete context in which to frame the abstraction (Bransford & Johnson, 1972). For example, when an idea is abstract, designers can provide visual or verbal examples that make the context clear.

Nominalizations

Nominalizations are nouns that have been derived from verbs or adjectives, often with Latinate suffixes such as "ize" or "tion" (e.g., *customization* instead of *customize* or *cessation* instead of *cease*). Nominalizations tend to make texts wordy and may make it difficult for readers to comprehend the main idea quickly. Studies find that readers often have trouble understanding nouns made from verbs (e.g., *amortization*) and find it much easier when the verb is used (e.g., *amortize*) (Charrow & Charrow, 1979; Coleman, 1964, 1965; Coleman & Blumenfeld, 1963).

Implications. In general, plain language advocates should avoid using nom-

inalizations, but there are exceptions. Some nominalizations are useful since they may not be hard to understand. They may even reduce sentence length (e.g., “failure”, “discovery”). Other nominalizations may be easy to understand if the underlying idea was presented in a previous sentence. For example, consider this sentence: “Researchers’ arguments focus on the cognition of interpretation and on how nominalizations slow reading and add to text density.” In this case, “researchers’ arguments” (used instead of “researchers argue”) is a nominalized subject referring to a previous sentence that would be obvious in context. For a thoughtful discussion, see Williams (2004).

Noun Strings

Strings of nouns (noun + noun + noun) make it hard for readers to parse ideas, as in the following (from a letter from a bank):

“You have exceeded the federal banking regulations maximum transaction number. An excessive items fee has been posted for excessive money market pre-authorized automatic debit transactions.”

Noun strings often slow readers’ efforts to make sense of the syntax of the sentence (Gagné, 2001; Levi, 1978; Murphy, 1990), especially if one of the nouns is also a nominalization – a noun made from a verb or adjective, for example:

“The chief loan officer controls the allocation of pre-screened amortization candidates and the refusal of your application suggests non-qualification or a history of late payment.”

Implications. Writers should avoid noun strings for they slow down the reading process and often confuse readers. Noun strings may have more of an impact on readers without topic knowledge about the text’s main ideas and on readers who have difficulty with reading. Readers with topic knowledge about the text’s content may be familiar with noun strings in that topic area and read them at the same rate as other text.

Overall, even though noun strings are intended to save space (to say the idea faster and make it shorter – usually good ideas), in many cases, they appear to do more harm than good. Moreover, they make the text seem dense, pon-

derous, and at times pretentious.

Identifying what's next: Gaps in research on word-level features

We have seen in this brief summary that word-level features such as length, frequency, difficulty, concreteness, nominalizations, and noun strings can have a significant impact on how readers process text and on whether they understand the main point. Of course, ideas about the use of word-level features may seem to be obvious or just common sense, but it is important for plain language advocates to confirm their beliefs with data. In this way, we can make evidence-based decisions as we write and design, and importantly, we can also defend our decisions with data when confronted with arguments about our choices.

Moreover, when we take a closer look at what the research tells us, we discover issues we would like to see studied further. For example, in looking over the research on word-level features we find gaps in our understanding of the following:

- Does the use of high-frequency words help skilled readers as much as people who have difficulty with reading or who are reading in a second language?
- Is word length comparable to symbol complexity or character complexity in non-Western languages, for example, in Chinese or Japanese?
- Which is better: an abstract yet precise word or a simple but less precise word? What are optimal techniques for handling subject matter that is inherently abstract? How can plain language advocates make abstract subject matter clear yet maintain the integrity of the content?
- Do some noun strings and/or nominalizations convey meaning better than other forms of expression? What are the conditions and situations in which a noun string or a nominalization might be the best strategy? Do noun strings cause problems for readers in languages other than English?
- What is the relationship between the concreteness of a word and the reader's ability to visualize the word? What role does visual imagery

play in the understanding of text? Are some subject matters or genre better comprehended if presented using easy-to-visualize concrete words?

Summary

By taking a “consolidate and identify what’s next” approach, we can move more rapidly toward evidence-based guidelines for plain language. And as the international community identifies the persistent problems of language and visual design that most countries share, we will be able to say with greater precision what research we still need. The research summarized here was carried out mainly with English speaking participants and focused for the most part on the English language. Principles and guidelines based only on English language research should be considered conditional until we establish that they apply to other languages as well. Thinking about the big picture and embracing what we have in common and how we differ will allow us to be more strategic as the international community works together on shared goals.

Let’s consider two examples from the research reviewed here that illustrate what universal and conditional principles and their attendant guidelines might look like.

Universal principle: There is a strong relationship between word length, word frequency, and word difficulty. The longer the word, the less frequent it is, and the harder it is to pronounce, the more likely it is to slow readers down and give them difficulty.

Evidence-based universal guideline: To make text easy to understand for most people, favor short words, high-frequency words, and simple words over long words, low-frequency words, and hard words.

Conditional principle: Professional groups may have their own subculture for language and prefer to use words and acronyms that are high frequency for them but low frequency for those outside of their group.

Evidence-based conditional guideline: If you are a member of a group with a special language subculture (e.g., physicians, engineers, lawyers)

and writing for the general public, avoid using words that are high-frequency within your group but not for people outside your group.

As these examples show universal principles apply to most writing and design situations while conditional principles apply to particular contexts. Evidence-based guidelines that derive from universal principles need to be based on fundamental ideas about what people have in common as they engage with writing and visual design. Those that are developed from conditional principles need respond to the uniqueness of local contexts – situations, languages, domains, or subject matters.

With principles and guidelines based on research about how people actually engage with writing and design, we will be better able to make arguments about what works. We will also be better positioned to establish international standards for what constitutes clear and effective communications.

Bibliography

- Bauer, D. W., & Stanovich, K. E. (1980). Lexical access and the spelling-to-sound regularity effect. *Memory and Cognition*, 8, 424–432.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 717–726.
- Calude, A. S., & Pagel, M. (2011). How do we use language? Shared patterns in the frequency of word use across world languages. *Philosophical Transactions of the Royal Society: Biological Sciences*, 366(1567), 1101–1107. doi: 10.1098/rstb.2010.0315
- Carpenter, P. A., & Just, M. A. (1981). Cognitive processes in reading: Models based on readers' eye fixations. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 177–213). Hillsdale, NJ: Erlbaum.
- Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). NY: Academic Press.
- Charrow, R., & Charrow, V. R. (1979). Making legal language understandable: Psycholinguistic study of jury instructions. *Columbia Law Review*, 79, 1306–1374.
- Coleman, E. B. (1964). The comprehensibility of several grammatical transformations. *Journal of Applied Psychology*, 48(3), 186–190.
- Coleman, E. B. (1965). Learning of prose written in four grammatical transformations. *Journal of Applied Psychology*, 49, 332–341.

- Coleman, E. B., & Blumenfeld, P. J. (1963). Cloze scores of nominalization and their grammatical transformations using active verbs. *Psychological Reports*, 13, 651–654.
- Cox, W. (1978). Problem-solving as a function of abstract or concrete words. *Contemporary Educational Psychology*, 3, 95–101.
- Dobbs, A. R., Friedman, A., & Lloyd, J. (1985). Frequency effects in lexical decisions: A test of the verification model. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 81–92.
- DuBay, W. H. (2004). The principles of readability. 1–76. Retrieved from Impact Information website: <http://www.impactinformation.com>
- Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 236–254. doi: 10.1037/0278-7393.271.236
- Graesser, A., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 193–202.
- Huckin, T. N. (1983). A cognitive approach to readability. In P. V. Anderson, R. J. Brockmann & C. R. Miller (Eds.), *New essays in technical and scientific communication: Research, theory, practice* (pp. 90–108). New York: Baywood Press.
- Hudson, P. T. W., & Bergman, M. W. (1985). Lexical knowledge and word recognition: Word length and word frequency in naming and decision tasks. *Journal of Memory and Language*, 24, 46–58.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Klare, G. R. (1968). The role of word frequency in readability. *Elementary English*, 45, 12–22.
- Kroll, J. G., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology*, 12(1), 92–107.
- Kucera, H., & Francis, N. W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in writing and spoken English based on the British national corpus*. London: Longman.
- Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
- Lorge, I. (1938). *The semantic count of the 570 commonest English words*. NY: Bureau of Publications, Teachers College, Columbia University.
- Mauer, D. (2006). Lakoff's "women, fire, and dangerous things" – What every IA should know. OZIA Podcast.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality.

- Written Communication*, 27(3), 57–86.
- McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005). Coh-Metrix version 2.0.
- Miller, G. A., & Selfridge, J. A. (1950). Verbal content and the recall of meaningful material. *American Journal of Psychology*, 63, 176–185.
- Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, 29, 259–288.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19, 304–330.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement*, 76(3, Part 2).
- Redish, J. C. (2000). Readability formulas have even more limitations than Klare discusses. *ACM Journal of Computer Documentation*, 24(3), 132–137.
- Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Erlbaum.
- Schrivier, K. A. (1997). *Dynamics in document design: Creating texts for readers*. New York, NY: John Wiley & Sons.
- Schrivier, K. A. (2000). Readability formulas: What's the use? *ACM Journal of Computer Documentation*, 24(3), 138–140.
- Schrivier, K. A. (2012). What we know about expertise in professional communication. In V. W. Berninger (Ed.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (pp. 275–312). New York, NY: Psychology Press.
- Schrivier, K. A. (2013a). Plain by design: Evidence-based plain language. PLAIN2013.
- Schrivier, K. A. (2013b). What do technical communicators need to know about information design? In J. Johnson-Eilola & S. Selber (Eds.), *Solving problems in technical communication* (pp. 386–427). Chicago, IL: University of Chicago Press.
- Schrivier, K. A. (2014). Special Issue on plain language and information design. *Intercom*, (February), 4–30.
- Schrivier, K. A. (in preparation). *Information design moves for print and web: Evidence-based practice*. Manuscript in preparation. Pittsburgh, PA.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27, 499–520.
- Stanovich, K. E., & Bauer, D. W. (1978). Experiments in the spelling-to-sound regularity effect. *Memory and Cognition*, 6, 410–415.
- Thorndike, E. L. (1921). *The teacher's word book*. New York: Teachers College, Columbia University.
- Thorndike, E. L., & Lorge, I. (1943). *The teacher's word book of 30,000 words*. NY: Bureau of Publications, Teachers College, Columbia University.
- Williams, J. M. (2004). *Style: Ten lessons in clarity and grace*, 8th ed., New York: Addison-Wesley

Longman.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

Author's note: Special thanks to **John R. Hayes** and **Janice (Ginny) Redish** for their very helpful comments on this article.