Bachelor information

Name: Emil O. W. Kirkegaard
Student ID: 20103300
Discipline: Linguistics
Advisor: Peter Bakker (?)

English resumé:

In this thesis, I discuss language reform with a focus on orthographic/spelling reform. I briefly review the some prior cases of language reform. Then I discuss reasons why one may want to or not want to reform spellings. In doing so, I draw upon works from psycholinguistics, educational linguistics (phonics), experiments with simplified spelling systems and cross-country comparisons. Then, I review the Danish orthography and its problems. Finally, I review my proposal for a danish spelling reform, *Lyddansk*.

Dansk resumé:

I denne opgave diskuterer jeg sprogreform med et fokus på ortografisk/stavemæssig reform. Jeg gennemgår kort nogle tidligere tilfælde af sprogreform. Derefter diskuterer jeg grundene for og imod en reform af stavemåderne. I min diskussion inddrager jeg værker fra psykolingvistik, uddannelseslingvistik (phonics), eksperimenter med simplificerede stavesystemer og tværnationale sammenligninger. Efter det så gennemgår jeg dansk retskrivning og problemer med den. Til sidst, så gennemgår jeg mit forslag til at reformere danske stavemåder, *Lyddansk*.

Why spelling reform? A review of the reasons with a focus on Danish

Emil O. W. Kirkegaard[1]

1. Introduction

In 2009, before starting university, I stumbled across a book called *Cut Spelling* (Upward, 1992) which has had a lasting impact on my career choice and interests. The book consists of a review of reasons for and against reforming English spellings, with its own proposal after which the book is named. The name of the proposal comes from the fact that it 'cuts away' quite a lot of letters in the spellings of English words. As a result of this, *Cut Spelling* is about 10% shorter than the traditional orthography (TO).

Already when I was reading this book, it was immediately apparent to me that many of the problems with English spellings also exist for Danish spellings, and so just as one could reform English spellings, so one could reform the Danish ones. From that point, the idea was set in my mind and I set up a page on the now defunct *Google Wave* (*Apache Wave*, Wikipedia) where one could crowdsource (i.e. work collaboratively on) a reform proposal for Danish.

I was not as naive as to think that no one had had this idea before, so I began looking for literature on the topic. I came across a short book named *Om Moderne Dansk Retskrivning* (Concerning Modern Danish Orthography; Hansen, 1969). It was a review of the problems of Danish orthography and some remarks about how to solve them. The book made a big impression on me and I still like to quote from it when discussing the topic.

While reading other literature on the topic, I continued to work on my proposal for Danish reform for a number of months. Sometimes it was impossible to make a justified decision without doing some statistical studies of Danish, so I conducted a number of such studies (Kirkegaard, 2010a, b, c). My work culminated in the publication of a short book on my proposal, *Lyddansk* (Sound Danish; Kirkegaard, 2010d) after 3-4 months work. In the book, I

---
[1] University of Aarhus, Denmark Department of Linguistics. email: emil@emilkirkegaard.dk

concerned myself almost exclusively with the linguistic details of specific reform proposals and not with the more sociopolitical aspect of which reasons there were for and against reform. Or as I wrote:

> As the work progressed it became clear to me that it was too much work to argue why one should reform, discuss how one could do it in a good way, and suggest a reform proposal in a single essay. This essay is dedicated to the last, namely to give a draft of a proposal for reform. The proposal is not meant as a solution to all problems that befall Danish orthography. It is meant as a reform that solves a number of problems and leaves others for later reforms. [p. 6, my translation]

Originally, I had in mind continuing my work on this subject and following up with a longer essay dealing with the question of whether one should reform or not and how to do so. In fact, I never followed up on the issue in detail. I did however set up a website for the project and wrote articles there sporadically, see: http://www.lyddansk.dk/ (all content in Danish).

The subject has continued to be of interest to me and I have read quite a number of books and papers over the last 4 years relevant to it. It therefore seemed quite natural that I could use the bachelor's thesis to further discuss the subject. The question was then whether to continue to write about it in Danish, as nearly all my previous writings on the topic, or make the switch to English. A great many papers and books have been written about the difficulties of the English orthography, but very little English-language material concerns Danish orthography. A Google Scholar (a free academic search tool comparable to paid services like Web of Science, see de Winter, 2014) for "Danish orthography" returns only 200 results. This is most unfortunate given that it is scientifically fruitful to compare orthography difficulties, as well as sociopolitical issues, across languages. For this reason I decided to write this thesis in English in the hope that it may be of interest to anyone interested in problems with a non-English orthography but who hasn't mastered the language.

The rest of the thesis is structured as follows:

- In section 2 I discuss definitions of key terms and give an overview of language planning in general
- In section 3 I discuss reasons for and against spelling reforms
- In section 4 I discuss the problems of Danish orthography
- In section 5 I sketch my reform proposal as given in Kirkegaard (2014d), so that readers who don't read Danish can understand it

- Finally, in 6 I conclude with some general remarks

2. Key terms and an overview of language planning

Before discussing language planning, one needs to understand a few key terms. The first subsection introduces them and the second gives an overview of language planning.

2.1 Some key terms

Following Coulmas (2003, p. 35ff), I will use the term *writing system* to describe general abstract systems used to represent meaning or sound. These include phonetic, phonemic, syllabic, logographic, ideographic and various in-betweens and mixes. The term *script* refers to the set of symbols used to write with. Some care must be taken with use of this term since it can be used in either narrow or broad senses. In the broad sense, Danish, English, German and all other Germanic languages use the same script, namely the Roman script. In the narrow sense, (almost?) all of them use their own script. The symbol <Å> does not exist in the English script, but exists in Danish, Swedish and Norwegian. Figure 1 below gives an overview of North Germanic languages and the special characters used or not used in them.
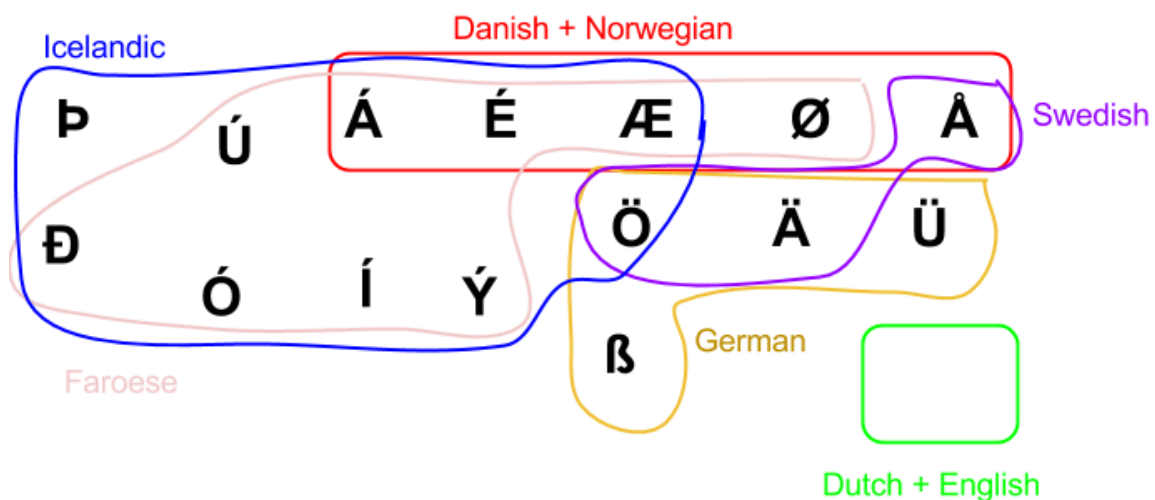

Figure 1 - Special symbols used in selected North Germanic languages. Self-made.

Only Dutch and English share a script, and one may argue that the Dutch use of the ligature <IJ> means that they must have their own script.

Spelling and *orthography* are commonly used with the same meaning, although they can fruitfully be distinguished (Bermel, 2006). I will use the first to refer to the order in which one puts letters, while the second will be a bit more broad and include conventions of punctuation and capitalization. The second word can be problematic because it carries

legalistic overtones in many languages. The reason for this is that the word is derived from two Greek roots meaning 'correct-writing'. Unless he is familiar with Greek, this morpheme analysis of the word will not strike the reader. However, in many other languages the native word is a loan translation (calque) of the Greek which can literally be understood as 'correct-writing', giving it a normative or legalistic tone. This is true for Danish <retskrivning> as well as many other languages such as German <Rechtschreibung>, Russian <правописание> and Czech/Slovac <pravopis> (Bermel, 2006, p. 5).

2.2 An overview of language planning

Language planning is a broad term that covers any conscious attempt to direct the way language changes or does not change. It covers diverse efforts such as revitalization of dead or nearly so languages or dialects. It may be top-down controlled by the government, or it may be bottom-up controlled in a grassroots fashion. Various scholars have made classification systems of what is included in the term. Wardhaugh and Fuller (2014, p. 367-368) define it as follows:

> Attempts to change languages, in terms of either their form or their function, are usually described as instances of **language planning**. … Language planning is an attempt to interfere deliberately with a language or one of its varieties: it is human intervention into natural processes of language change, diffusion, and erosion. (bold in original)

Their classification includes only two megacategories: *status planning*, which has to do with which languages are official (*de jure* or *de facto*), which are used or allowed in schools or public life in general, and *corpus planning* which concerns efforts to make sure that a given language has enough words to be a useful way of communicating about some topic. It may involve developing an orthography for the language (for many African languages, this meant creating a new one from scratch or fitting a Roman script to the language perhaps by utilizing diacritics, see e.g. Karan 2006), or it may involve inventing new words for an area where previously none or only foreign words existed. This may involve, for instance, technical fields where previously some major language (e.g. Latin, English, or Spanish) was used instead of the local language.

2.2.1 The Turkish reform of 1928

Since I'm mostly concerned with spelling reform, I will only list one non-spelling example.

A famous and interesting case is that of the Turkish reform of 1928. The historical context is that the Republic of Turkey is a relatively new country formed in 1923 from the remnants of the then collapsed Ottoman empire. There were both political and linguistic reasons to reform.

The linguistic reasons were that the country had a very low literacy rate of only around 10%, which the government wanted to improve. To do this, they needed to make the language easier for the general public. They did this in two ways: 1) changing the script and 2) removing foreign elements from the lexicon and grammar (Dogançay-Aktuna, 2004).

The pre-reform written language was written with the Arabic script. Arabic has a relatively simple vowel structure and perhaps for this reason the writing system does not have ways to indicate more complex vowel systems. Often vowel are not written at all, which has led some to classify the script as a special kind of phonemic system variously called *consonantal systems, abjads or consonantaries* (Karan, 2006, p. 44ff). Turkish, on the other hand, has a rich vowel system which cannot adequately be written using the Arabic script (for a description of the system, see e.g. Göksel and Kerslake, 2005, chapter 3).

The purification of foreign elements consisted of collecting material on words from spoken Turkish which could replace those of foreign origin (usually Arabic or Persian) that were traditionally used in the written language. They also researched older Turkish texts to find words that had fallen out of favor but that could be reintroduced to replace a word with foreign origin. Finally, they used already existing roots to derive words that could replace foreign words. The result was a dictionary published in 1934 proposing some 30,000 Turkish equivalents for about 7,000 Ottoman words.

The reform was an effective interplay of good linguistic reasons (easier to learn, expanded domains of expressiveness), bad linguistic reasons ("Sun-Language Theory", the idea that Turkish was the mother of all languages) and political reasons related to nationalism which helped gather public support.

2.2.2 Noah Webster and American spellings

It is well-known that American English and British English have different spellings for a number of words. Briefly put, they concern among other things:

- the writing of <s> or <z> at the ends of words e.g. <reali(z/s)e
- the removal of silent <u> in some words e.g. <colo(u)r>

- the choice of <-re> or <-er> at the ends of words ending with /ə(r)/ e.g. <metre> vs. <meter>
- the removal of final silent <-ue> e.g. <dialogue>
- the removal of silent <a> from some words of Greek origin e.g. <an(a)emia>

As one can see, most of the American ones go towards a better fit with pronunciation, so one might wonder if they were deliberately made or chosen like that. The answer is "yes". It was the American reformist and nationalist Noah Webster (famous for his dictionary), who introduced the spellings in connection with the War of Independence (Algeo, 2000). One might wonder why he didn't go further than the small changes there are between American and British, and the truth is that he wanted to go further, but that the attempts failed. He preferred, for instance, <wimmin> over <women>, and <tung> over <tongue> (Scudder, 1881, chapter 7). His proposals and reasoning are surprisingly modern given that they were written over 200 years ago (Webster, 1789). Clearly, English spelling has had the same problems for many generations.

2.2.3 The German spelling reform of 1996

The German reform of 1998 was, linguistically speaking, a small matter. It covered 6 areas (Johnson, 2005, p. 55ff):

1. Spelling of /ɛ/ as <ä> instead of <e> so as to harmonize spellings among related words e.g. <behende> to <behände> so as to fit with <Hand, Hände> [hand(s)].
2. Doubling of consonants in a few words to fit with related words, e.g. <numerieren> (to number) to >nummerieren> to fit with <nummer> (number).
3. <ss> instead of <ß> after short vowels to harmonize between related words, e.g. <Haß> to <Hass> (hate) so as to fit with <hassen> (to hate).
4. Consistent spellings in compounds to fit with roots, e.g. <Flußsand> to <Flusssand> (river sand), to fit with <Fluss> (river).
5. A small number of words fixed to fit with analogous words, e.g. <rauh> to <rau> (rough) to fit with <grau,schlau> (grey, clever).
6. Germanization of a few foreign words, such as <Delphin> which gained the alternative spelling <Delfin> (dolphin), in line with analogous words e.g. <Megafon> (megaphone).

Given the relatively minor changes, one may have expected the matter to have settled quickly without too much of a fight. This was not to be so. In fact, even though the reform was announced in 1995, by 1998 there were already some 30 court cases related to the

reform. Even the 'supreme court' (constitutional court) had to hear two cases related to the issue (both were decided in favor of the reform). And it didn't end there, in autumn 1998 Schleswig-Holstein, the most northern region (Bundesland) decided in a popular vote to opt out of the reform. However, a year later the regional parliament overturned the decision and they once again had to teach the new orthography in Schleswig-Holstein schools. To this day, there continue to be popular organizations against the reform (Johnson, 2005, p. 1ff).

3. Rational reasons for and against orthographic reforms

There are quite a number of sociolinguistic studies of orthographic reforms (e.g. Czech, Bermel 2007; Turkish, Dogançay-Aktuna 2004; German, Johnson 2005). Generally, they concern mostly the sociopolitical or ideological aspects of reforms, and not the linguistic reasons in detail. They discuss the orthographic system of the language, and what the problems were that the reform sought to fix. They also mention that the goals for fixing them was to make language easier to learn and use. However, they go no further, and do not generally address various non-political counter-arguments made. This section is devoted to discussing the linguistic evidence in more detail for why a simpler, more consistent orthographic system helps to make the language easier to learn and use. I am not concerned with the relationship between socialism, capitalism, nationalism, conservative or any other broad political and language reform.

3.1 Rationality and cost-benefit analysis

There are two defining features of rationality: instrumental and epistemic. *Instrumental rationality* concerns choosing the right means towards one or more ends. *Epistemic rationality* concerns the relationship between one's beliefs and the evidence that one has. A famous quote from the philosopher David Hume may be taken as a loose definition of epistemic rationality: "A wise man ... proportions his belief to the evidence" (Hume, 1748, Of Miracles), i.e. that one believes things one has good evidence for, and does not believe things one has poor evidence for (Yudkowsky, 2009). Although it is tempting to reduce either to the other, this appears not to be possible (Kelly, 2003). In any case, the two are very closely related.

Being instrumentally rational about a decision means that one has to consider the costs and benefits with regards to the possible actions and one's preferences. In other words, being rational about decision making is applying cost benefit analysis.

My preferences in this regard are as follows: 1) Reducing economic and time costs to society is good, 2) increasing children's desire to be in school or to learn in general is good, 3) increasing literacy is good, 4) decreasing irrelevant discrimination is good, and 5) learning is good.

3.2 Reasons for reform

The rest of section 3 will be structured as follows. First, I will go through the reasons for believing that an orthographic reform will help towards one or more of the stated goals. I begin with the simple intuitive argument, and then I follow up with reviews of the empirical evidence in two subfields of linguistics that support the intuitive argument. In subsection 3.3 I will go through the reasons usually given against reforms and argue that they are not as big problems as supposed by the opponents.

3.2.1 The intuitive argument

The argument is based on the simple generalization that things that are easier to learn can be learned faster, mastered to a higher degree and learned by more people. Most reformed orthographies go towards simpler and more consistent orthographic systems (for an exception, see Karker 1976). So, it follows that learning to use such a system, i.e. learning to read and write, will be faster, will result in mastery of a higher degree, and will result in mastery by more people compared with the TO. Since quite a lot of time in school is spent teaching students to read and write, the time freed up can either be used for more learning of other stuff (goal #5) or freeing up state resources (goal #1). Furthermore, since it is easier to learn, literacy will increase (goal #3). Finally, the sheer difficulty of learning to read and write may create hostility towards school and learning in general in children, since they get frustrated trying to learn a broken system. As far as this is the case, this hostility can be reduced (goal #2) with a simpler and more consistent orthography.

Sometimes, reform proponents cite correlational studies of literacy to support the above claims (e.g. Cleckler, 2005), but generally, they keep things at the intuitive level (Upward, 1992; Ripman and Archer, 1948).

3.2.2 Evidence from psycholinguistics

It is curious that reform proponents have not cited some of the vast amount of published material from psycholinguistics. Psycholinguistics is the interdisciplinary field that combines cognitive psychology (the study of the mind/brain using computer-like models) with linguistics (Traxler, 2006, 2011). Quite a lot of the research in that field has centered on

figuring out how exactly reading works. Because of space limitations, this review will be very short, but the curious reader can consult my recent review of the literature elsewhere (Kirkegaard, 2014).

While as is common in scientific fields, there is no universal agreement, there are standard models. The dual-route model (and variants thereof) is by now fairly well accepted. The model says that reading works by both a grapheme-phoneme channel (GPC; i.e. by sound) as well as by a visual matching system that matches the image of the word against a database of known words (Traxler, 2011, p. 391). Research provides evidence for both routes. For instance, research shows that people often don't notice when words are 'misspelled'[2] if the misspelling represents the same phonemes. For instance, if the spelling was <meet> instead of <meat>, both pronounced /mi:t/, when a type of food was meant.

Evidence of the direct route (DR) comes from, for instance, studies examining the reading times of words. More common words are read much faster than more uncommon words. This fits with the idea they are directly recognized as effectively one symbol.

The most relevant evidence with regards to spelling reform comes from studies of words that have inconsistent spellings. A large number of studies show that words that have an irregular spelling take more time to read, cause more regressions (when the eyes move backwards when reading because a word was not recognized), and that when people read them, they make more pronunciation errors. This concerns words like <have> of which the predicted pronunciation from GPC is /heɪv/ but the actual is /hav/.

Another line of evidence comes from the reading of written words that have two pronunciations (homographic heterophones). For instance, <tear>, which could be pronounced /tɛə/ (to rend something) or /tiə/ (water from the eyes), is read more slowly than controls, presumably due to inconsistent results from the GPC and DR (see e.g. Folk and Morris, 1995). These words also cause more errors in pronunciation even though they require more time.

Studies have found that when subjects are given a task to quickly decide whether a word belongs to a category or not, and they are given a word that does not belong but is pronounced the same as a word that does, the subjects make more false positives. For instance, if asked to answer whether the word refers to a flower and shown the word <rows>

---

[2] by which I mean only that it did not follow the official or standard spelling, with no implication of the standard being the only correct way.

they tend to say that it belongs, consistent with the idea that the GPC system is used and results in the phonemic word /roʊz/ which is also spelled <rose> (e.g. Van Orden 1987).

Finally, the results come not only from English, but also from e.g. Chinese (Tsai et al, 2005). Generally, the evidence from psycholinguistics shows that people rely on the GPC to read to a large degree and that words that are irregular cause problems in reading.

One word of caution given the psycholinguistic evidence reviewed above is that the results have generally not been meta-analyzed and rely on many small studies with sample sizes of 10-50 persons.

3.2.3 Evidence from phonics

In the teaching of literacy, there are two approaches which have received the most attention. The phonic approach which emphasizes teaching pupils the grapheme-phoneme relationships needed to decode the written word, and the holistic/whole-language approach which emphasizes the need to dive into the literature quickly and for the children to themselves discover the grapheme-phoneme relationships.

By now there are large meta-analysis (combining the results of many studies into one estimate) assessing which method is best to teach. Ehri et al (2001) combined 38 studies and found an overall effect size of systematic phonics of d=.41. The effect was stronger when teaching began earlier (d=.55 before first-grade vs .27 after).

The authors write:

> In sum, systematic phonics instruction proved effective and should be implemented as part of literacy programs to teach beginning reading as well as to prevent and remediate reading difficulties.

Since phonics relies on sound-grapheme relationships in English to work, making these more regular means that it will work better.

3.2.4 Evidence from simplified spelling systems

A direct way of testing whether it is easier to learn a RO is simply to teach such a system to children and compare their literacy abilities to children taught TO. This may sound hard to get ethical approval for, but it has in fact been done.

An old but famous study is Downing (1964) where a large number of students were taught either RO (N=1132) or TO (N=1984). There were clear and marked differences on their reading ability. In their ability to read primer books, the children taught RO *were about twice as far ahead* (their Table 1). Similarly, there were very large differences on reading and spelling tests, even when tested in TO!

A newer small study compared Italian children to British children learning either TO or RO (Thorstad, 1991). Italian is much more phonemic and consistent than TO English, so the Italian children were expected to do better. The children learning Italian or RO English did in fact do much better than those learning TO English.

3.2.5 Evidence from cross-country comparative studies

Similar to the small Italian study above, there are a number of studies that have compared children's ability to read and write across countries. These studies present weaker evidence due to the problems of comparing across countries. Countries may differ in other relevant criteria such as average intelligence (Lynn and Vanhanen, 2012), age of school start (World Bank, 2014), length of daily instruction (Center for Public Education, 2011), instruction method, children's motivation to learn to read and write and so on. Due to the large number of such studies I will mention only three.

Landerl et al (1997) compared German and British dyslexics and found that the British ones fared much worse, in line with the expectation based on the relatively superiority of the German orthograph.

Aro and Wimmer (2003) compared British, French, German, Dutch, Spanish, Swedish and Finnish children and found that the British were by far the worst at reading pseudowords (invented words that follow normal spelling patterns).

Juul and Sigurdsson (2005) compared Danish and Icelandic children and found the Icelandic children to perform better in line with their generally superior orthography.

3.3 Reasons against reform

A variety of reasons are offered against reforming orthographies in general. As I see it, there are two main rational but insufficient reasons given against reform. I consider both below. There are also a number of irrational reasons given, but merely listing every such reason offered would take a lot of space. I discuss two of them as an example.

3.3.1 The dangers of homographs

A quite common argument against reforms consists of pointing at homographs created by the reform and claiming that there will be confusion (e.g. Kjertmann, 2006). Generally, they do not provide any statistics as to whether the reform will actually create more homographs than it will remove, so even if homographs were a problem in themselves, one cannot evaluate the strength of the argument.

However, we know that homographs are not generally a problem. Evidence from psycholinguistics (reviewed above) also shows that reading times do not increase when dealing with homographic homophones which is the class of words that will be created most after phoneme-based reforms (e.g. Folk and Morris, 1995). Generally speaking, for homographs to be a problem, they must be *confusable in context*. The context part is important, since context, whether meaning or grammar, almost always clarifies which word was intended. In the design of a reform, one must examine whether confusing homographs will be created and if so, avoid creating them by keeping different spellings for them.

3.3.2 Alienation from existing literature

Sometimes the argument is made that reforming will make it harder to enjoy the existing literature (Hansen, 1969, p. 9). This is true enough, but also a matter of degree. Small reforms, such as the Czech or German, had a very small effect on the mutual intelligibility of historical variants.

A large reform, such as the full LD proposal, would substantially change the look of the written language, but mutual intelligibility is still very high I think. This can be empirically tested by testing how intelligible LD is to persons not trained in the system beforehand.

The problem is also fairly easy to mitigate in that electronic literature can relatively easily be converted between the spelling systems. Printed books cannot be changed, of course, but due to the high mutual intelligibility, this is not that big of a problem.

3.3.3 Language is/must be an organic system

A number of objections consist of various appeals to nature based on the metaphor that language is living, and that one should not disturb or hurt it (Bermel, 2007, p. 274). Generally, when it is stated, it is packed in metaphors that are difficult to make sense of. But perhaps the most plausible interpretation is that one should not change the orthography deliberately because it is natural. However, when stated so clearly, the argument is hopefully

not persuasive to many people. For instance, we regularly deliberately change other natural processes such as cancer. Would they similarly object to applied medicine?

3.3.4 Etymological information will be lost

Sometimes the argument is made that if we change the spelling of foreign words to localized versions, then one will not be so easily able to infer which language the word comes from.

This is true, but of doubtful importance. First, if one really wanted to know the etymology of a given word, one would consult an etymological dictionary. Second, the primary function of the written language is not to provide etymological information about words. *The primary function is to be a means of communication*.

3.3.5 Status quo bias

A number of cognitive biases have been discovered by now (e.g. Kahneman, 2011). A well-studied one is *status quo bias,* which is the tendency for people to favor the current system over alternatives, regardless of what the current system is. They do this by rationalizing their preference for the current system over the alternatives.

The question is how we can rationally figure out when someone, perhaps ourselves!, is affected by status quo bias. The philosophers Nick Bostrom and Toby Ord (2006) proposed the reversal test to solve this. They write:

> Reversal Test: When a proposal to change a certain parameter is thought to have bad overall consequences, consider a change to the same parameter in the opposite direction. If this is also thought to have bad overall consequences, then the onus is on those who reach these conclusions to explain why our position cannot be improved through changes to this parameter. If they are unable to do so, then we have reason to suspect that they suffer from status quo bias. (p. 664-665)

The evidence from the reversal test is defeasible, i.e. if they can supply reasons to think that we are in a local optimum, it can be reasonable to disfavor any change in a parameter.

If we return now to the etymological argument above, we can see that it is plausibly a case of status quo bias. To do this, simply ask proponents of that argument whether they would prefer us to return back to even more etymological spellings. In Danish, we now write <suveræne> (sovereign) instead of the more etymological <souveraine> (as found in the

King Law of 1665[3]). Yet etymology was offered as a reason to prefer the current <mayonnaise> spelling over the Danified <majonæse>. Alternatively, consider the word <løjtnant> (lieutenant) which used to be spelled <lieutenant>. Should we go back?

4. The Danish orthography and its problems

As mentioned in the introduction, Danish orthography does not seem to have received much attention from English speaking researchers. What follows is thus strongly based on my own analyses.

Danish is part of the north Germanic group of languages which also includes Swedish, Norwegian, Faroe Islandic and Icelandic. They are most closely related to the languages in the West Germanic group which includes English, Dutch and German (Stampe Sletten et al, 2005). Denmark has had an extensive trade relationship with Germany through perhaps 800 years, which means that Danish is more German than the other two Scandinavian languages, especially Swedish. To make matters complicated, Denmark and Norway used to be one country (1524–1814), and the most commonly used Norwegian orthography (Bokmål, 'book-language') is based on the old Danish version. As such, Norwegian Bokmål and Danish have very high written mutual-intelligibility while spoken is somewhat less so.

4.1 Phonology

Danish is sometimes called a hard language to learn by Danes (e.g. Koldbye, 2009; Bleses et al, 2011). It is hard to say whether it is a hard language compared with all other languages, but it certainly presents some difficulties for natives and foreigners alike, which I will examine more closely in Section 4.3.

4.1.1 Vowels

There is no universal agreement about the phonological system, but in a phonemic analysis, it has approximately 14 different vowels which can be both short and long (which is why it is sometimes said it has over 20). Grønnum (1998) thinks there are 10, not including short/long versions, diphthongs related to glides, or the schwa /ə/ sound. My analysis differs from hers in that I include schwa, another <a>-vowel, another <e>-vowel, and another <å>-vowel. The perhaps most authoritative source (Basbøll, 2005, p. 50, Table 2.3) lists 12 vowel phonemes "outside /r/-contexts" and not including length variations. The difference again being schwa

---

[3]Lex Regia. https://www.retsinformation.dk/Forms/r0710.aspx?id=20950

and the second <e>-vowel (which only occurs in /r/-contexts). I have my reasons for my analysis, but the space is too limited here to discuss them.

Table 1 gives an overview. DANIA is a special system developed for writing the sounds found in Danish (Jespersen, 1890). It is similar to the letters used in the writing system, so it is easier to use when solely concerned with Danish or if one is not linguistically trained. I list both IPA (International Phonetic Alphabet) and simplified DANIA in the table, but will use DANIA-S in the text.

IPA	DANIA-simp.	Example words
a	a	tal, sal, hval, mal, sand, tand, sagde, tilbage
ɑ	á	sang, lang, tang, far, kaffe, fakkel, lap, gaffel
ɛ	æ	fæ, læ, stær, sæt
æ	ä	er, frisk, bjerg, jern
e	e	fe, le, se, be, te, Peter, til, vil, spil, mere, flere
i	i	ti, si, bi, ni, sige, pige, lige, hi, lig, gide, pine
y	y	sy, Thy, by, ly, ny, fly, dy, fy, lyd, myg, styg
ø	ø	ø, bøg, sø, nødig, kø, tø, bøtte, flytte, nytte
œ	ö	høvl, brøk, rød, brød, tømmer, tør, søn, køn
ɔ	å	å, ål, sål, bål, mål, tåle
o	o	ko, flod, so, to, bro
ʌ	ó	ånd, rådne, åre, såre, lår, over, tov, ovn, snot
ə	é	nazisme, forbyde, gibbe, hændelse, høne,
u	u	sur, pur, lur, nu, sgu, kunne, mulig, skur, bur

Aside from the large number of vowel phonemes, there is an even larger number of allophones. One of the reasons for the different number of vowel phonemes in different analyses concerns the question of how to analyze the vowels into the phonemic system. One could, and some do (e.g. Basbøl, 2005), spend quite a lot of time analyzing these vowels, but since writing systems are usually phonemic-based not phonetic, the issue does not concern us here. Native speakers do not generally notice these differences, so they could not distinguish between them in writing without special training.

4.1.2 Consonants

In contrast with the vowels, Danish consonants are quite simple. Table 2 gives an overview similar to the one before.

IPA	DANIA-simp.	Example words
b	b	bi, be, bid, busk, bas, bat, bit, byt, båt
d	d	de, det, den, dem, da, du, dø, dyt, dus
f	f	fe, far, Fie, Fur, fabel, elefant, fantastisk
g	g	gø, gå, gøre, Gasolin, gamma, ganske
h	h	hø, ho, ha, har, hej, hus, hidsig, hent
j	j	ja, jo, *j*eg, jep, Jeppe, javel, jantelov, *j*age
k	k	ko, kø, kan, kendt, kunne, kilde, kulde, kunstig
l	l	le, lo, ly, læ, lå, lad, led, Lis, los, langsom
m	m	må, mø, man, men, min, mon, mink, mene
n	n	ni, nu, ny, nå, næ, nas, nok, nemlig, nok

ŋ	ŋ	penge, konge, ting, ding, mink, flink, pink
p	p	på, pi, pæn, pas, pus, pose, pibe, penis
ʁ	r	rå, ro, ry, ru, ret, rar, ros, rum, rus, rask, race
s	s	se, si, so, så, sy, sad, set, sod, sok, syde
t	t	te, ti, to, ty, tø, tå, tag, tid, tog, tus, tøs, tale
ʊ	v	ve, vi, vag, ved, vis, vor, vug, vås, hvad

Table 2 - 16 consonant phonemes of Danish.

Of major note concerning consonants is that there is a morphophonemic rule that unvoiced phonemes gain voice when in final position, e.g. /knæk/ becomes [knæg] in speech. Similarly /d/ becomes [ð] and /v/ becomes [ʊ].

4.1.3 Stød

In the linguistic world, Danish is perhaps most known for its stød (written as /ˀ/ in this text), a kind of glottal stop that occurs at the ends of some words. It is the difference between <mor> (mom) and <mord> (murder).

4.2 The Danish writing system

As with all the other Germanic languages, Danish uses a Roman-based alphabetic system probably best understood as an imperfect morphophonemic system. There are 29 monographs (i.e. not di- or trigraphs), which consist of the normal 26 from the Roman alphabet along with the three additional <æ>, <ø> and <å>. In total this gives 9 graphs for vowels, and 20 for consonants. If we think back to the phonemic system, we can perhaps guess what the problems will be: There are more vowel phonemes than vowel graphs (14>9), and more consonant graphs than consonants (20>16).

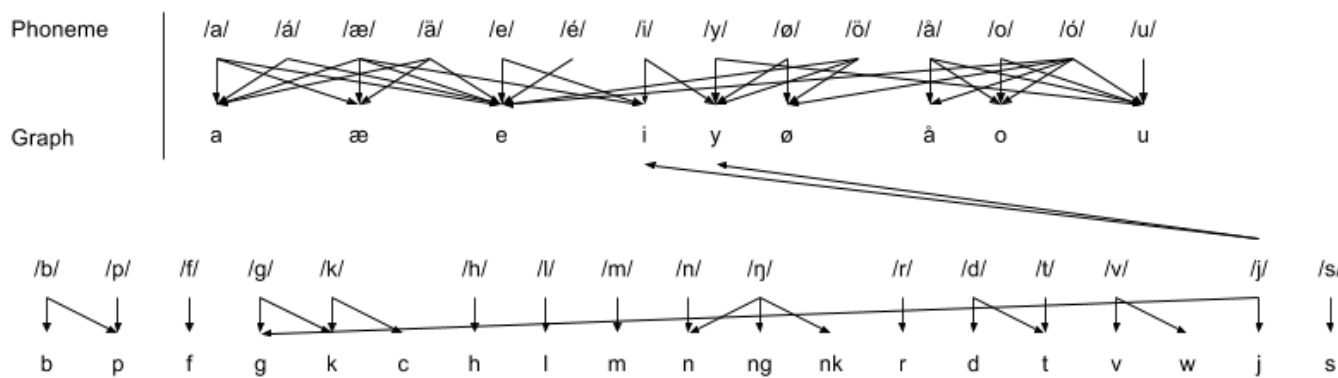Figure 2 gives an overview of the connections between phonemes and graphs made using the public Bogstavlyd ('lettersound') database at http://bogstavlyd.ku.dk/.

Figure 2 - The connections between phonemes and graphs. Only connections that occur at least 1% of the time are shown. Doubled consonant-letters are not shown. Vowel length and stød are ignored.

Going from phonemes to writing is of course what the writer does (unless the word form is memorized itself as is normal for common words). The reader is faced with a different scenario, namely going from graphs to phonemes. The reverse connections are shown in Figure 3.
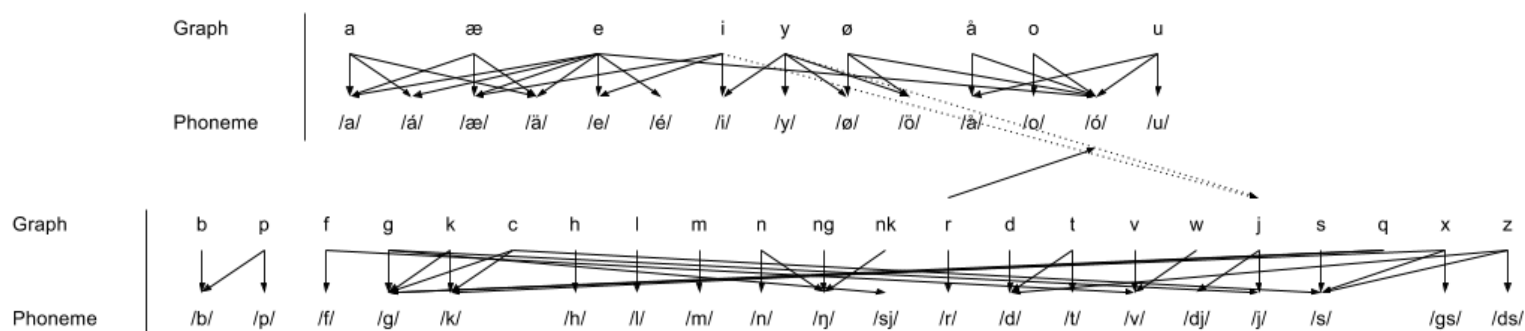

Figure 3 - The connections between graphs and phonemes. Only connections that occur at least 1% of the time are shown. Doubled consonant-letters are not shown. Vowel length and stød are ignored. The stippled lines have no special significance, the stippling is merely because they overlap with the text.

These figures hide considerable detail. They do not show the relative strengths of the connections. For instance, <c> is connected with /s/ 44% of the time, /k/ 40%, and /g/ 15%. On first appearances it seems very bad for the graph <c>, but it is actually relatively predictable if one knows the next graph too. When <c> precedes <e,i,y,æ> then it is more than 90% of the time pronounced <s>. In the other cases, it is more than 90% of the time pronounced <k,g>. A relatively simple figure such as the above cannot show this more complicated relationship.

Another problem with the figures is that the morphophonemic rule of unvoiced consonants becoming voiced in the coda is not taken into account in their analysis. This means that graphs like <k> are actually more predictable than their analysis and the figures show.

Still, if one were to consider a perfect phonemic system, then any graph would be connected to exactly one phoneme, and vice versa. It is immediately apparent that Danish is very far from a perfect phonemic system.

4.2.1 Silent graphs

Not shown in the figures either is the large amount of silent or 'silent' graphs. For instance, many common words, including all the question-words, have a silent <h> in front of them. Aside from knowing whether the word is a question-word, these are not predictable from the sound or the meaning. As an example, <hval> /va:l/ (whale), but <valg> /val'/ (election, choice). In the current version of the official dictionary (*Retskrivningsordbogen*, the correct-writing-word-book), there are 152 words beginning with <hv>, so the problem is quite limited.[4]

A larger problem concerns the use of <ll> and <nn> vs. <ld> and <nd>. For <ll> and <ld>, there are 19 word minimal pairs with respect to spelling. For instance, <balle> and <balde> pair, and are both pronounced /balé/ (compressed ball of grass or corn; buttocks; DAT). Another example is <bilde> and <bille>, both /bilé/ (to let someone on; beetle). Some pairs however are not pronounced the same, such as <spille> /spelé/ and <spilde> /spilé/. Eye-balling the results for all words with <ild> in them (253, DAT), it seems that <i> is always pronounced /i/, but in words with <ill> the pronunciation varies between <i,e>. This is another example of the above where there is a 'deeper pattern' that one can learn. It is for this reason that orthographies such as the Danish, English and French are said to be "deep orthographies".

If we turn now to <nn> and <nd>, there are 10 minimal spelling pairs (DAT). For instance, <finne> and <finde> are both pronounced /fenə/ (Finn, to find). As before, there are also pairs where pronunciation is different such as <kunne> /kuné/ (could) and <kunde> /kåné/ (customer). There are likely deep patterns concerning the pronunciation here as well, but often the spelling is not predictable from the sound and conversely.

---

[4] The number comes from using my Dictionary Analysis Tool (DAT), which can be found here: http://emilkirkegaard.dk/dat/

The use of quasi-doubled consonant graphs is associated with a precedent short vowel as in <gylde> /gylé/ (dung), but sometimes it is long in recent imports from English such as <allround> /ó:lrávnd/. However, words with no (quasi)doubled consonant graph also are often short as in <tal> which is both /ta:l/ (speak, imp.) and <tal'> (number[s]).

4.2.2 Vowels

As one could guess from the mere number of vowels (14) and the number of available graphs (9), there are many problems with these. If one looks up <i> in Bogstavlyd, one sees that it is /i/ 55% of the time, /e/ 34%, /æ/ 7% and /j/ 2%. There is thus a considerable task before the reader, he either has to find some deeper pattern which he can use to probabilistically infer the vowel, or he must take a guess.

In fact, most of the vowel graphs are like <i> above. Table 3 gives the percentages for each.

Graph	Sound (%)
a	á (43), a (39), ä (16)
æ	æ (76%), a (12), ä (12)
e	é (62), æ (14), e (11), á (5), ó (4), ä (2), a (2)
i	i (55), e (34), æ (7), j (2)
y	y (69), ø (16), ö (9), k (3), i (2)
ø	ø (51), ö (35), ó (13)
å	å (63), ó (37)
o	ó (53), o (39), å (7)
u	u (61), å (32), v (3), ó (2)

Table 3 - Table of vowel graph to sound connections in percent. Length and stød is ignored. Made from data from the Bogstavlyd database. http://bogstavlyd.ku.dk/forside/

The table shows the data from which the top part of Figure 3 is made. Similarly, one could show a table for the bottom part of Figure 3 as well as both parts of Figure 2. But it is already clear that the system is messy, at least at the superficial level.

5. *Lyddansk*: a reform proposal for Danish

As mentioned in the introduction, I worked out this proposal in 2009-2010. Back then, I was unaware of the Bogstavlyd database, but I used a dictionary with pronunciation data to work out tables like Table 3 above. For each vowel graph, I looked up 100 randomly chosen words in the dictionary and noted their pronunciation (I used Politiken, 2005). Then I calculated the percentages (Kirkegaard, 2010a, also in 2010c p. 17-21). To get the tables with sound to graph relationships, I found the total number of words with each graph, and calculated the percentage of words with a given sound which were spelled in each way (Kirkegaard, 2010b).

Generally, my results are broadly consistent with those from the *Bogstavlyd* database. Aside from the sampling error arising from using only 100 words, the two sources disagree in their transcription in that *Bogstavlyd* is has more narrow transcriptions and uses different symbols. They also sometimes disagree about the actual pronunciation of words. As a comparison, in my study, <a> was found to be related to /a/ 60% of the time, and /á/ 40%. However, in the case of <å>, the numbers are exactly the same using both methods: 63% /å/ and 37% /ó/.

5.1 Moderation as a guiding principle

I was aware of the troubles concerning previous very minor changes in *Retskrivningsordbogen* in Denmark. Often, one would see articles in the mainstream newspapers talking about the decay of the Danish language or "grammar" because this or that foreign word had a change in spelling, or some rule was changed regarding when words are joined or not (e.g. Brammer et al, 2010 "Language Board guilty of language decay", Pøhler, 2012 "Danish grammar is heading straight into the abyss"). Because of nonsense like this, for political reasons, one should change as few things as possible. No radical reform could possibly succeed in such a hostile environment.

Aside from the tactical reason above, another reason to be as moderate as possible is because any changes make the transition harder for a person already proficient in Danish. We don't want to convert a good speller into a bad one in a split second.

Because of this moderation, I set myself certain limits:

- No new letters or diacritics
- No indication of stød or pressure
- No attempted solution of the use of <ld,ll>, <nn,nd> or vowel length indication
- No attempted solution of <r,er,re,ere> in the ends of words

- No change in spelling with regards to the morphophonemic rule that changes unvoiced to voiced consonants in the coda (Section 2 in Kirkegaard, 2010c)

The goal was not to create a perfect phonemic system for writing Danish, merely to make things easier for those who have to use the language: from everyday users, to children and foreigners learning it.

5.2 How does the system work?

The system is based on that from *Cut Spelling* which was discussed in the introduction (Upward, 1992). The reform is *not* a new radical system through which one can work out the correct spelling of every word purely from its pronunciation. Instead, the system is a number of rules that changes the existing spellings under certain conditions. As such, it is made to be easy to learn for those who know the existing version, as well as being generally simpler which helps first time learners.

The system is almost completely modular in that each of the rules can generally be put into effect independently of each other without causing problems. There are some exceptions to this. The very common words <de> /di/ (these) and <det> /de/ (this, that) are changed to <di> and <de>. It would be foolish to make the second change without making the first since this would cause a very confusing homograph. Probably, one would need to make the first change and then wait some years before making the second to avoid causing confusion.

The general modularity of the proposal means that one can pick and choose between the rules, perhaps let people vote on them so as to introduce public opinion into the reform. Previous cases have shown that for reforms to succeed in democratic countries, it is very important to consult the public before introducing reforms (see Bermel, 2007).

5.2.1 Examples of Lyddansk rules

In the latest published version of the system (version 1.2, dated 2013 February), there are 43 rules which are grouped up by which letter they concern.

For instance, there are five rules concerning <c> which together removed the letter entirely from the spellings. For every rule, a few examples are given in TO and *Lyddansk* (LD).

Rule C1 concerns words spelled with <c> pronounced as /s/. These are now to be spelled with <s>. Examples: TO *cykle, komplicere,* LD *sykle, komplisere*

C2: words with <c> and /k/ to <k>. TO *cafe/café, Cypern, cancer* LD *kafe/kafé, Kypern.* Curiously, in TO, it is inconsistently spelled <Kario> (Cairo) but <Cypern> instead of <Kypern>. The spelling of foreign names is clearly not consistent.

C3: words spelled with <sc> which are pronounced /s/. These are now <s>. To *fascinerende, disciplin, discipel* LD *fasinerende, disiplin, disipel.*

C4: words with <ch> and /tj/ (tʃ) which are now to be spelled <tj>. TO *charter, chat, chili* LD *tjarter, tjat, tjili.*

C5: words with <ch> and /sj/ (ɕ) changes to <sj> (like Norwegian). TO charme*, chef, chartek* LD *sjarme*, *sjef*, *sjartek*.

5.3 Comparisons of Lyddansk and TO

Using the full set of rules, I translated a piece of text so as to examine how large the change is in total. The chosen text is chapter 5 from *Alice in Wonderland*. It is 2238 words long.

Using the *Dictionary Analysis Tool*, one can compare two pieces of text for a number of properties. The number of changed words was 754, which is 34% of the total words. These are the words that have received at least one change. The average length of words fell about 2%, which perhaps has to do with the cutting away of silent letters, especially in words from French. For instance, rule O3 changes <our> to <ur> when the pronunciation is /ur/ (e.g. TO <amour> LD <amur>).

The total number of changes is 929. By letter, the changes are shown in Table 4.

Letter	Change (to LD)
a	+27%
c	-15%[5]
e	-12%
f	-2%
g	-38%
h	-17%

---

[5] One might wonder why this number is not -100% given that <c> is completely removed from the spellings. The reason is that the protagonist is named <Alice> and her name has not been changed. It could be changed to <Alis> if one wanted, but people are even more conservative with names.

i	-31%
j	+93%
o	-4%
s	+1%
t	-2%
u	-10%
v	+22%
y	-59%
æ	+232%
ø	+40%
å	+25%

Table 4 - Change in percentages of letters used in a text from applying all the LD rules. Only letters with changes are shown.

We see that the Dano-Norwegian letters <æ,ø,å> become more much common, especially the first. This is due to the fact that these are currently written often with other letters due to the general sound change that has occurred in Danish which included a lowering of many vowels. Similarly, we see that <j> increases sharply which has to do with the fact that it is commonly written with <g,i> right now. The increase in <a> has to do with one very common word that is spelled with <e> (<jeg> /jáj, I).

Together, the overall change seems to make Danish more distinct from both the other Scandinavian languages and the other Germanic languages in general.

6. Conclusion

Regardless of the desirability of reforming orthographics, it is a tough battle in any democracy. I have no expectation that the Danish orthography will be reformed in the near future, perhaps not even in my lifetime. Because of this, the relationship between spelling and sound will become worse and children and foreigners will be forced to learn the increasingly inconsistent and nonsensical spellings wasting an unknown but probably very large amount of tax-money. My interest in designing a reform proposal has mostly been rational-scientific, not ideological-political.

References

Algeo, J. (2000). "The Effects of the Revolution on Language", in A Companion to the American Revolution. John Wiley & Sons,

Apache Wave. (2014, November 1). In *Wikipedia, The Free Encyclopedia*. Retrieved 17:40, November 9, 2014, from

http://en.wikipedia.org/w/index.php?title=Apache_Wave&oldid=632013547

Basbøll, H. (2005). *The Phonology of Danish*. Oxford University Press.

Bermel, N. (2007). *Linguistic authority, language ideology, and metaphor: the Czech orthography wars* (Vol. 17). Walter de Gruyter.

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive processes*, *26*(8), 1193-1231.

Bostrom, N. and Ord, T. (2006). The Reversal Test: Eliminating Status QUo Bias in Applied Ethics. *Ethics*, 116, July.

Brammer, J., Mejer, L., Grevil, F. S. (2010). Sprognævnet medskyldig i sprogligt forfald. *Berlingske* 9. November.

http://www.b.dk/kommentarer/sprognaevnet-medskyldig-i-sprogligt-forfald

Center for Public Education. (2011). Time in school: How does the U.S. compare?. http://www.centerforpubliceducation.org/Main-Menu/Organizing-a-school/Time-in-school-How-does-the-US-compare

Cleckler, B. C. (2005). *Lets End Our Literacy Crisis: The Desperately Needed Idea Whose Time Has Come*. American University & Colleges Press

Coulmas, F. (2003). *Writing systems: An introduction to their linguistic analysis*. Cambridge University Press.

Dogançay-Aktuna, S. (2004). Language planning in Turkey: Yesterday and today. *International journal of the sociology of language*, (165), 5-32.

Downing, J. (1964). The ita (Initial Teaching Alphabet) Reading Experiment. *The Reading Teacher*, *18*(2), 105-110.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panel's meta-analysis. *Review of Educational Research*, 71, 393–447.

Folk, J. R., & Morris, R. K. (1995). Multiple lexical codes in reading: Evidence from eye movements, naming time, and oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(6), 1412.

Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association*, 28, pp 99-105. doi:10.1017/S0025100300006290.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. Psychology Press.

Hansen, A. (1969). *Om Moderne Dansk Retskrivning*. Gads forlag, København.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*.

Jespersen, O. (1890). *Danias Lydskrift*. http://www.hum.au.dk/jysk/publikationer/dania.pdf

Johnson, S. A. (2005). *Spelling trouble?: language, ideology and the reform of German orthography*. Multilingual Matters.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Karan, E. (2006). *Writing system development and reform: A process.* Doctoral dissertation, University of North Dakota.

Karker, A. (1976). Nordisk retskrivning – den ideale fordring. *Sprog i Norden*. s. 39-84. http://ojs.statsbiblioteket.dk/index.php/sin/article/download/17861/15627

Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, *66*(3), 612-640.

Kirkegaard, E. O. W. (2010a). *Kvantitativ undersøgelse af sammenhængen mellem stavemåde og vokal i dansk retskrivning*. http://www.lyddansk.dk/litteratur

Kirkegaard, E. O. W. (2010b). *Kvantitativ undersøgelse af sammenhængen mellem vokal og stavemåde i dansk retskrivning*. http://www.lyddansk.dk/litteratur

Kirkegaard, E. O. W. (2010c). *Kvantitativ undersøgelse af sammenhængen mellem initialt J og udtale i Dansk retskrivning*. http://www.lyddansk.dk/litteratur

Kirkegaard, E. O. W. (2010c). *Lyddansk*

Kirkegaard, E. O. W. (2014). *Orthographic reform and psycholinguistics: a selective synthesizing review*. http://emilkirkegaard.dk/en/?p=4483

Kjertmann, K. (2006). *Læsning, stavning og retskrivningsreform*. Information. 6. April. http://www.information.dk/122274

Koldbye, C. (2009). *Derfor er det svært at lære dansk*. Videnskab.dk. http://videnskab.dk/kultur-samfund/derfor-er-det-svaert-laere-dansk

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*(3), 315-334.

Lynn, R., & Vanhanen, T. (2012). *Intelligence: A unifying construct for the social sciences*. London: Ulster Institute for Social Research.

Politiken. (2005). Politikens Nudansk Ordbog med etymologi. 3rd edition. Electronic version.

Pøhler, J. (2012). Dansk grammatik er på vej ud over afgrunden. *Politiken*. 30. October.

Ripman, Walter; Archer, William; Jones, Daniel (ed.) and Orton, Harold (ed.). (1948). *New Spelling. Being proposals for simplifying the spelling of English without the introduction of new letters*. 6th edition. Sir Isaac Pitman & Sons, London. http://emilkirkegaard.dk/en/wp-content/uploads/New-Spelling-book.pdf

Scudder, H. E. (1881). *Noah Webster*. Cambridge, Mass.: The Riverside Press. http://www.gutenberg.org/files/31238/31238-h/31238-h.htm

Stampe Sletten, I., Torp, A., Häkkinen, K., Svonni, M., & Christian Olsen, C. (2005). Nordens sprog-med rødder og fødder. Nordisk Råd.

Thorstad, G. (1991). The effect of orthography on the acquisition of literacy skills. *British Journal of Psychology*, *82*(4), 527-537.

Traxler, M. J. (2006). *Handbook of Psycholinguistics*. Academic Press.

Traxler, M. J. (2011). Introduction to psycholinguistics: Understanding language science. John Wiley & Sons.

Tsai, J. L., Su, E. C. I., Tzeng, O. J., & Hung, D. L. (2005). Consistency, regularity, and frequency effects in naming Chinese characters. *LANGUAGE AND LINGUISTICS* 6.1:75-107

Upward, C. (1992). *A handbook to the simplification of written english by omission of redundant letters*. Simplified Spelling Society

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(3), 371.

Wardhaugh, R & Fuller, J. M. (2014). An Introduction to Sociolinguistics. Wiley-Blackwell, 7th edition.

Webster, N. (1789). "An Essay on the Necessity, Advantages, and Practicality of Reforming the Mode of Spelling and of Rendering the Orthography of Words Correspondent to Pronunciation", in *Dissertations on the English Language: With Notes, Historical and Critical, to Which is Added, by Way of Appendix, an Essay on a Reformed Mode of Spelling, with Dr. Franklin's Arguments on That Subject.* Boston.
http://edweb.sdsu.edu/people/DKitchen/new_655/webster_language.htm

de Winter, J. C., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics, 98*(2), 1547-1565.

World Bank. (2014). Primary school starting age (years) [database].
http://data.worldbank.org/indicator/SE.PRM.AGES

Yudkowsky, E. (2009). What Do We Mean By "Rationality"?. LessWrong.
http://lesswrong.com/lw/31/what_do_we_mean_by_rationality/