When is Orthography Optimal?

Höskuldur Thráinsson
University of Iceland
Purpose and organization of the talk

Purpose:
• Try to learn something about orthography in general and its relation to linguistics and language by studying two explicit attempts to design an optimal orthography

Organization:
• Introduction
• The “First Grammarian” and Icelandic orthography in the 12th century
• Faroese orthography in the 19th century
• Concluding remarks
Introduction

Conflicting claims about English spelling

• It has been claimed (some say by George Bernhard Shaw) that it is so irregular that the word for this:

   ![Fish](image)

   could either be spelled *fish* or *ghoti*.

   (cf. *gh* in *enough*, *o* in *women* and *ti* in *nation*)
Introduction, 2

Shaw (Tiough?, cf. nation, ought) with his fish:

(From the cover of the book Language and Literacy: The Sociolinguistics of Reading and Writing by Michael Stubbs.)
Conflicting claims, contd.:

• It has also been claimed that modern English spelling is “near optimal” (emphasis by HTh):

there is ... nothing particularly surprising about the fact that conventional orthography is ... a near optimal system for the lexical representation of English words ... (Chomsky and Halle 1968:49)
The nature of the conflict:

• Some believe that orthography should be **phonetically/phonologically based** (“shallow phonology”) — as close to the principle “one letter ↔ one sound” as possible (Shaw?).

**Question:** Is phonetic transcription easy?

• Others believe that it should be **morphologically/morphophonemically based** — the basic principle being roughly “one morpheme ↔ one orthographic representation” (Chomsky and Halle 1968).

**Question:** Is English orthography really like that?
More on “morpho-phonemic” orthography (emphasis HTh):

- The fundamental principle of orthography is that **phonetic variation is not indicated where predictable by a general rule** ... Orthography is a system designed for **readers who know the language** ... Such readers can produce the correct phonetic forms, given the orthographic representation ... Except for unpredictable variants (e.g. *man* – *men*, *buy* – *bought*), an optimal orthography would have **one representation for each lexical entry** (Chomsky and Halle 1968:49; see also N. Chomsky 1970 for a similar view).

- Simply stated the **conventional spelling of words corresponds more closely to an underlying abstract level of representation** within the sound system of the language than it does to the surface phonetic form that the words assume in the spoken language" (C. Chomsky 1970:28).

- **What the foreigner lacks is just what the child already possesses**, a knowledge of the phonological rules of English that relate underlying representations to sound (C. Chomsky 1970:62)

- Boston University, April 20, 2012

Höskuldur Thráinsson, University of Iceland
Introduction, 6

Some English illustrations of the morphophonemic/morphological principle

- `<a>` for different vowels: nation, national
- `<e>` for different vowels: extreme, extremities
- `<c>` for different consonants: medicate, medicine
- `<g>` for different consonants: sage, sagacity
- `<g>` for a sound and silence: signature, sign
- `<b>` for a sound and silence: bombardment, bomb
- `<s>` for the 3rd person morpheme: likes, plays
- `<ed>` for the past tense morpheme: liked, played

(mostly copied from Cook’s web-site)
Introduction, 7

What follows:

• A report on the design of two radically different orthographic systems, namely one for **Old Icelandic** and for **Faroese**.

Why might this be interesting?

• **Sheds some light on the issues** just outlined (the nature and of different types of orthography and the relation of orthography to linguistic structure).

• Some of it is of **general linguistic interest**, partly also from the point of view of the **history of linguistics** (e.g. the use of **minimal pairs** in linguistic argumentation in the 12th century) and the relationship between language and culture.
Iceland and the Faroes

Iceland (+ the Faroes + Scotland)  The Faroes
Icelandic and Faroese

Closely related North-Germanic (Nordic) languages.

**Icelandic:** Approx. 300,000 speakers
- rich literary tradition, medieval manuscripts (sagas etc.)
- extensive written sources from the 12th century onward
- used throughout in schools, administration, church, written literature ...

**Faroese:** Approx. 50,000 speakers
- virtually no written sources between 1400 and 1800
- not used in schools, administration, church nor in (written) literature until the 19th and the 20th century (Danish was the official language until the middle of the 20th century)
- ballads preserved in oral tradition (typically connected to folk dances)

(See e.g. Thráinsson 1994, Barnes and Weyhe 1994, Thráinsson 2007, Árnason 2011, Thráinsson et al. 2012.)
Designing Old Icelandic (OI) Orthography


• The First Grammatical Treatise (FGT) from approx. 1175.
• Preserved in a vellum manuscript (*Codex Wormianus*) from around 1350 (contains three other grammatical treatises).
• The explicit purpose was to design an orthography and (to some extent) an alphabet for Old Icelandic. Or in the First Grammarian’s (FG’s) own words (cf. Hreinn Benediktsson (ed.) 1972:207ff. — his translation (emphasis HTh)):

  because languages differ from each other, which previously parted or branched off from one and the same tongue, different letters are needed in each, and not the same in all, just as the Greeks do not write Greek with Latin letters ...
FG’s own words (contd.):

Whatever language one intends to write with the letters of another language, **some letters will be lacking** [because each language has sounds that are not to be found in the other language; and likewise, **some letters are superfluous**] because the sound of the surplus letters does not exist in the language. Thus, Englishmen write English with all those Latin letters that can be rightly pronounced in English, but where these do not suffice, they apply other letters, as many and of such a kind as needed, but they put aside those that cannot be rightly pronounced in their language.
FG’s own words (contd.):
Now, following their example, since we are of one tongue (with them), although one of the two (tongues) has changed greatly, or both somewhat, in order that it may become easier to write and read, as is now customary in this country ... [the FG then mentions laws, genealogies, translations of religious texts, the book of settlement ...] I have composed an alphabet for us Icelanders as well, both of all those Latin letters that seemed to me to fit our language well, in such a way that they could retain their proper pronunciation, and of those others that seemed to me to be needed in (the alphabet), but those were left out that do not suit the sounds of our language. A few consonants are left out of the Latin alphabet, and some put in; no vowels are left out, but a good many put in, because our language has almost all sonants or vowels ...
FG’s representation of the OI vowels:

- Uses the **five standard Latin** ones: < i, e, a, u, o >
- **Adds four**: < q, ɛ, ọ, y >

Common presentation of the OI short vowel system (the “added” vowel symbols in red):

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unrounded</td>
<td>rounded</td>
</tr>
<tr>
<td>high</td>
<td>i</td>
<td>y</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>ọ</td>
</tr>
<tr>
<td>low</td>
<td>ɛ</td>
<td></td>
</tr>
</tbody>
</table>
The FG’s explanation of the symbols chosen (and hence also of their quality):

<œ> has the loop from a and the circle from o because it is a blending of the sounds of these two, **pronounced with the mouth less open than a but more than o**

<œ> is written with the loop of a but with the full shape of e, just as it is composed of the two, **with the mouth less open than a but more than e**

<œ> is composed of the sounds e and o, **pronounced with the mouth less open than e but more than o** and therefore, in fact, written with the cross-bar of e and the circle of o

<y> is made into a single sound from the sounds of i and u, **pronounced with the mouth less open than i and more than u** and therefore ... [describes a combination of j and v, as j and i were not always distinguished in OI spelling nor were v and u]
The FG’s use of minimal pairs to show that the vowel symbols he argues for represent distinct phonemes (speech sounds):

Now I shall place these ... letters ... between the same two consonants, each in its turn, and show and give examples how each of them, with the support of the same letters (and) placed in the same position ... makes a discourse of its own, and in this way give examples, throughout this booklet, of the most delicate distinctions that are made between the letters:

\[
\begin{align*}
    \text{sar} : \text{sor}, & \quad \text{ser} : \text{sør}, & \quad \text{sor} : \text{sør}, & \quad \text{sur} : \text{syr} \\
    \text{‘wound’ (sg:pl)} & , \quad \text{‘sees’ : ‘sea’} & , \quad \text{‘swore’ : ‘fair’} & , \quad \text{‘sour’ : ‘sow’ (pig)}
\end{align*}
\]
The FG’s example sentences for the first minimal pairs:

- A man inflicted one *sar* (‘wound’) on me, I inflicted many *sør* (‘wounds’) on him.
- The priest *sor* (‘swore’) the *sør* (‘fair’) oaths only.
- The eyes of the *syr* (‘sow’) are *sur* (‘sour’) ...

Now note:
All the vowels in these examples were distinctly long in OI. Later in the treatise the FG suggests that long vowels should be distinguished from short ones by an accent, i.e. as *sár, sór, sýr, sør, sýr, súr* for the examples above. But this distinction has not been introduced in the FGT when the FG presents these minimal pairs. But see the next slide!
OI Orthography, 8

The quantity distinction in OI and the FGT: Long vowels indicated by an acute accent:

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>i, í</td>
<td>y, ý</td>
</tr>
<tr>
<td>Mid</td>
<td>e, é</td>
<td>õ, õ</td>
</tr>
<tr>
<td>Low</td>
<td>ë, ë</td>
<td>a, á</td>
</tr>
</tbody>
</table>

Some minimal pairs in example sentences used by the FG to argue for this distinction:

- *far* (‘vessel’) is a ship and *fár* (‘harm’) is a kind of distress
- *ôl* (‘ale’) is a drink but *ól* (‘strap’) is a cord

---

Boston University, April 20, 2012

Höskuldur Thráinsson, University of Iceland
The nature and linguistic interest of the FGT:

• Obviously phonetic and (structuralist) phonological rather than morphemic/morphological (cf. the types discussed above).

• Some of the orthographic distinctions that the FG suggests were never consistently made in Icelandic medieval manuscripts (e.g. to use dots over nasal(ized) vowels), but the writing tradition from the 12th century is unbroken and extensive manuscripts preserved from all centuries.

• The main interest of the FGT is the information it contains on the OI sound system (especially the vowel system) and the linguistic (structuralist) argumentation it uses (minimal pairs) more than 700 years before the rise of structuralism in Europe and the US.
The oldest preserved Faroese documents:

- A couple of legal documents from around or shortly after 1300 (‘The Sheep Document’ and ‘The Dog Document’).
- Four letters from around 1400 (‘The Húsavík Letters’).
- A transcription of ‘The Sheep Document’ from around 1600.
- Other than this, there are virtually no written sources in Faroese until around or after 1800.

Note: The oldest documents are basically written in Old Norse/Old Icelandic — hardly any specific Faroese traits.
The (re-)emergence of writing in Faroese after 1800 (cf. Thráinsson et al. 2012):

**Svabo’s manuscripts:**
- Transcription of traditional ballads began around 1800 (not published until the 20th century).
- A manuscript of a Faroese-Danish-Latin dictionary around 1800 (also not published until the 20th century).

**The first books:**
- A collection of ballads published in 1822 (first book in Far.)
- *The Gospel according to St. Matthew* published 1823.
- *The Faroe Islanders’ Saga* published 1832.

The orthography used in the earliest published books varied somewhat — **no standardization** and some **dialectal differences**.
An example of the earliest orthography and Modern Faroese orthography:

**Svabo’s orthography:**
Aarla vēar um Morgunin
Sēulin roār uj Fjōdl
Tajr sēuē ajn so miklan Mann
rujā ēav Garsiā Hōdl.

**Modern Far. orthography:**
Árla var um morgunin
sólin roðar í fjøll
teir sóu ein so miklan mann
rīða av Garsia høll.

(Rough translation: ‘It was early in the morning, (when) the sun was coloring the mountains, (that) they saw a great man ride from Garsia’s palace.’)

**Questions:**
• How and why did they get from Svabo’s orthography to the modern one and what is the difference between the two?
• Are they equally easy to read? What if you know (another) Scandinavian language?
Some problems with the first published books:

• *St. Matthew* was sent to all households in the Faroes (1200 in all!) but was not well received. Some reasons:
  1. People were not used to Faroese in the church.
  2. People had never learned to read in Faroese.
  3. Some complained about dialectal traits.

• *The Faroe Islanders’ Saga* was much better received. Some reasons:
  1. One of the main characters fights against foreign (Norwegian) rule of the islands.
  2. It had parallel texts in three languages: Faroese, Old Norse and Danish (*St. Matthew* had Danish and Faroese).
  But there were still some dialect problems (spelling not equally natural for all dialects).
Faroese Orthography, 5

Some issues raised in the discussion of Faroese orthography between 1830 and 1850:

- **which letters to use** to represent speech sounds where there was no dialect variation (cf. the FG on OI)
- **which variant to choose** where there was dialect variation (cf. comments above, not mentioned by the FG)
- **which principle** to follow:
  1. *phonetic/(surface) phonological* (cf. the FG for Old Icelandic — and Svabo etc. for 19th cent. Far.)
  2. *morphophonemic/morphological* (cf. Chomsky and Halle; often referred to as “historical” in the Faroese discussion)
  3. *etymological* (also referred to as “historical”; not discussed above)
One of the issues: Long and short vowels

- In Old Icelandic vowel length was **distinctive** and the FG suggested indicating it with an acute accent (cf. the table on slide 19)
- In Modern Faroese (and Modern Icelandic) vowel quantity is **positionally determined**: stressed vowels are long in open syllables — vowels are short elsewhere, including diphthongs.
- In Modern Faroese there is **considerable phonetic difference between long and short variants of vowels**, the long monophthongs frequently being higher or diphthongized and some short diphthongs can also be monophthongized (cf. below).

**Question:**

Should long and short vowels be represented differently in Modern Faroese orthography?
Orthographic representation of long and short vowels in Old Icelandic and Faroese:

<table>
<thead>
<tr>
<th>Old Icel. (FG)</th>
<th>Svabo:</th>
<th>Mod. Far. Orth.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a [a] (far ‘vessel, ship’)</td>
<td>ea [ɛaː] (fɛar)</td>
<td>a [ɛaː] (far)</td>
</tr>
<tr>
<td>a [a] (fars ‘ship (Gen.)’)</td>
<td>a [a] (fars)</td>
<td>a [a] (fars)</td>
</tr>
<tr>
<td>i [i] (bil ‘moment’)</td>
<td>ii [iː] (biil)</td>
<td>i [iː] (bil)</td>
</tr>
<tr>
<td>i [i] (bils ‘moment (Gen.)’)</td>
<td>i [i] (bils)</td>
<td>i [i] (bils)</td>
</tr>
<tr>
<td>í [iː] (vín ‘wine’)</td>
<td>uj [uiː] (vujn)</td>
<td>í [uiː] (vín)</td>
</tr>
<tr>
<td>í [iː] (vins ‘wine(Gen.)’)</td>
<td>uj [ui] (vujns)</td>
<td>í [ui] (víns)</td>
</tr>
</tbody>
</table>

Which principles are being followed here?
Different principles discussed in the Faroes (19th cent.), e.g. the following:

1. Faroese orthography should “imitate pronunciation” to the extent possible (cf. Svabo; this was also the opinion of the Danish linguist Rasmus Chr. Rask at the time).

2. Faroese orthography should be as close to Old Norse (and Modern Icelandic) orthography as possible (Nordic scholars and philologists).

Question:

• What is the purpose? What should the orthography be good for? Who should make use of it (e.g. read Faroese)?
Different motives:

1. To **preserve records of a dying language**. Chief proponents: Svabo and the original compilers/publishers of the traditional Faroese ballads.

2. To create a **written form** of the language which would be easy for the Faroese themselves to read — and make it possible to provide them with something useful/interesting to read. Chief proponents: publishers of the biblical texts and literature (St. Matthew, The Faroe Islanders’ Saga, folk tales ...).

3. To make Faroese **accessible to scholars knowing Old Norse** (and/or Modern Icelandic). Chief proponents: Scandinavian philologists around the middle of the 19th century.
Different types of orthography suitable depending on the motive:

• The **phonetic principle** (cf. Shaw; Svabo, the ballads, St. Matthew, The Faroe Islanders’ Saga) was appropriate for the *preservation motive* — and to some extent the *native reading motive*, except for the dialect problems that arose and problems inherent in trying to stick to this principle.

• The **morphophonemic principle** (cf. Chomsky and Halle) was (is) appropriate for *native reading motive* — also to some extent for the *accessibility motive*, partly depending on the choice of letters (cf. below).

• An **etymological** (or **historical**) principle was appropriate for the *accessibility motive*. 
What does the **etymological/historical principle** involve?

- **Not changing the orthographic symbols** (letters) although the relevant sounds have changed — or even disappeared from the language.

cf. the table on slide 27 and the examples on the next slide.

**Question:**
Are any aspects of Modern English orthography arguably etymological rather than simply morphophonemic (morphological)?
Keeping OI (or ON) letters in Mod. Far. orthography:

In addition:

<ð> was introduced into the OI alphabet in the 12th century to denote a dental fricative (cf. OI (and Modern Icel.) bróðir, English brother). It is used in Modern Faroese orthography (e.g. bróðir) although there is no dental fricative in Faroese.
Comparison of Svabo and Modern Far. orthography:

<table>
<thead>
<tr>
<th></th>
<th>Svabo (etc.)</th>
<th>Mod. Far.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Disregards the merger of ON /i, í/ and /y, ý/ using &lt;y, ý, i, í&gt;</td>
<td>no</td>
</tr>
<tr>
<td>b.</td>
<td>Disregards the merger of initial hv- and kv- using &lt;hv-&gt; for old /hv/ and &lt;kv-&gt; for old /kv/</td>
<td>no</td>
</tr>
<tr>
<td>c.</td>
<td>Disregards the change ll &gt; dl and nn &gt; dn</td>
<td>no</td>
</tr>
<tr>
<td>d.</td>
<td>Uses &lt;ð&gt; where ON had /ð/ (and Mod.Icel. does)</td>
<td>no</td>
</tr>
</tbody>
</table>

- **a, b** are **general phonological mergers** and there is no synchronic way of retrieving the distinction.
- The change in **c** is **historical** and not a productive phonological rule governing paradigmatic alternations.
- In most of the cases where <ð> is used in (modern) Faroese orthography (cf. d), there is no synchronic evidence for any (“underlying”) dental consonant (but see ráða ‘decide’, past tense ráddi, where -dd- goes back to /ð+ð/).
Further comparison of Svabo and Mod. Far. orth.

<table>
<thead>
<tr>
<th>Svabo (etc.)</th>
<th>Mod. Far.</th>
</tr>
</thead>
<tbody>
<tr>
<td>e. Disregards the quite general change -ang- ( \rightarrow ) -eng-</td>
<td>no \hspace{1cm} yes</td>
</tr>
<tr>
<td>f. Disregards the quality differences between long and short variants of vowels</td>
<td>no \hspace{1cm} yes</td>
</tr>
<tr>
<td>g. Disregards regular palatalization of velars</td>
<td>no \hspace{1cm} yes</td>
</tr>
</tbody>
</table>

- The change -ang- \( \rightarrow \) -eng- (cf. e) didn’t happen in all Faroese dialects.
- The alternations in f, g are completely regular phonological (morphophonemic) alternations — something which is obscured in a “phonetically based” orthography.
Which aspects of Modern Far. orthogr. are difficult?

a. When to use i or y, e.g. inna ‘do’ : ynna ‘love’ and í or ý, e.g. síl ‘awl’ : síl ‘trout’
b. When (not) to write ð, e.g. 
   *borðið* ‘the table’ : *borið* ‘carried’ : *bori* ‘drill(Dsg.)
c. When to write hv- or kv-, e.g. *hvalir* ‘whales’ : *kvalir* ‘pains’

Concluding remarks:
So when is orthography optimal?


References, 2

Cook, Vivian. (No year.) Writing Systems. In Cook’s own words: “a web-site that provides information, amusement and advice about writing systems in general and English spelling and the English writing system in particular”. [http://homepage.ntlworld.com/vivian.c/index.htm](http://homepage.ntlworld.com/vivian.c/index.htm)


An Ode to the Spelling Chequer
Janet E. Byfor

Prays the Lord for the spelling chequer
That came with our pea sea!
Mecca mistake and it puts you rite
Its so easy to ewes, you sea.

I never used to no, was it e before eye?
(Four sometimes its eye before e.)
But now I've discovered the quay to success
It's as simple as won, too, free!

Sew watt if you lose a letter or two,
The whirled won't come two an end!
Can't you sea? It's as plane as the knows on yore face
S. Chequer's my very best friend

I've always had trubble with letters that double
"Is it one or to S's?" I'd wine
But now, as I've tolled you this chequer is grate
And its hi thyme you got won, like mine.
... but it doesn’t really matter

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a total mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe. Amzanig huh?

(This has been floating around on the Internet, but it is attributed to G. Rawlinson at Nottingham University in the UK.)