



When is Orthography Optimal?

Höskuldur Thráinsson
University of Iceland

Boston Unversity, April 20, 2012



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
- Conclusion



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:



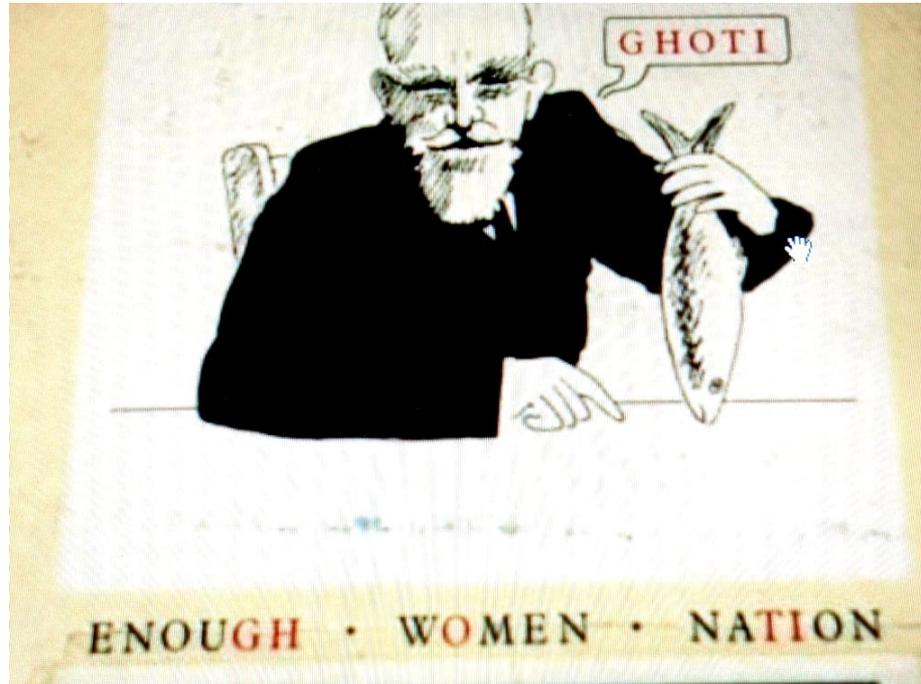
could either be spelled *fish* or *ghoti* .

(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.)



Introduction, 3

Conflicting claims, contd.:

- It has also been claimed that modern English spelling is “**near optimal**” (font changes 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)



Introduction, 4

The nature of the conflict:

- Some believe that orthography should be **phonetically /phonologically based** (“shallow phonology”) — as close to the principle “one letter \leftrightarrow 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 \leftrightarrow one orthographic representation” (Chomsky and Halle 1968).

Question: Is English orthography really like that?



Introduction, 5

More on “morpho-phonemic” orthography (font changes 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)



Introduction, 6

Some 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*
- 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**.

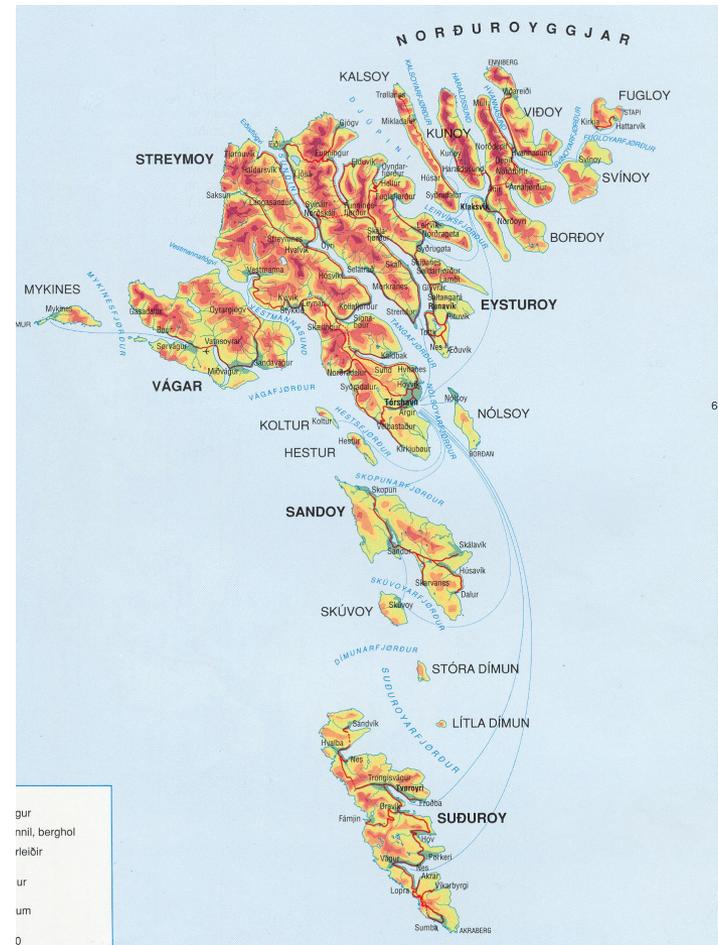
Why might this be interesting?

- Sheds some light on the issues just outlined (the nature and pros and cons 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

- 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 document (cf. Haugen (ed.) 1950, Benediktsson (ed.) 1972):

- 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 (partially) an alphabet** for Icelandic. Or in the First Grammmarian'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 ...



OI Orthography, 2

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.



OI Orthography, 3

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 ...**



OI Orthography, 4

FG's representation of the OI vowels:

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

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

	front		back	
	unrounded	rounded	unrounded	rounded
high	i	y		u
mid	e	ø		o
low	ɛ		a	ɔ



OI Orthography, 5

The FG's explanation of the symbols chosen (and hence also of their quality):

<ø> 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***

<ę> 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***

<ø> 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 *v* and *u*) ...



OI Orthography, 6

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:

sar : *sqr*,
'wound' (sg:pl)

ser : *sēr*,
'sees' : 'sea'

sor : *sør*,
'swore' : 'fair'

sur : *syR*
'sour' : 'sow' (pig)



OI Orthography, 7

The FG's example sentences for the first minimal pairs:

- A man inflicted one *sar* ('wound') on me, I inflicted many *sor* ('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 distinctively 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*, *sór*, *sór*, *sýr* for the examples above. But this distinction has not been introduced when he 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:

	front		back	
	unrounded	rounded	unrounded	rounded
high	i, í	y, ý		u, ú
mid	e, é	ø, ø		o, ó
low	ɛ, é		a, á	ɔ, ó

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
öl ('ale') is a drink but *ól* ('strap') is a cord

etc.



OI Orthography, 9

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.



Faroese Orthography

The oldest preserved Faroese documents:

- A couple of legal documents from around 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, basically no writing in Faroese until around 1800.

Note: The oldest documents are basically written in Old Norse/Old Icelandic.



Faroese Orthography, 2

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**.

Faroese Orthography, 3

An example of the earliest orthography and Modern Faroese orthography:

Svabo's orthography:

Aarla ve^ar um Morgunin
Se^ulin roär uj Fjødl
Tajr se^uü ajn so miklan Mann
rujä e^av Garsiä Hødl.

Modern Far. orthography:

Árla var um morgunin
sólin roðar í fjøll
teir sóu ein so miklan mann
ríða 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? Might that depend on your language background (e.g. whether you know a Scandinavian language or not)?



Faroese Orthography, 4

Some problems with the first published books:

- *St. Matthew* was sent to all households in the Faroes but was not well received. Some reasons:
 1. People were not used to Faroese in the church.
 2. People didn't know how 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)



Faroese Orthography, 6

One of the issues: Long and short vowels



Faroese Orthography, 6



Faroese Orthography, 6



Faroese Orthography, 6



Faroese Orthography, 6



Faroese Orthography, 6







References

- Árnason, Kistján. 2011. *The Phonology of Icelandic and Faroese*. Oxford University Press, Oxford.
- Barnes, Michael P., and Eiwind Weyhe. 1994. Faroese. In Ekkehard König and Johan van der Auwera (eds.): *The Germanic Languages*, pp. 190–218. Routledge, London
- Benediktsson, Hreinn (ed.). 1972. *The First Grammatical Treatise*. Introduction. Text. Notes. Translation. Vocabulary. Facsimiles. Institute of Nordic Linguistics, Reykjavík.
- Chomsky, Carol. 1970. Reading, Writing and Phonology. *Harvard Educational Review* 40, 2:287–309.
- Chomsky, Noam. 1970. Phonology and Reading. In H. Levin and J.P. Williams (eds.): *Basic Studies in Reading*, pp. 3–18. Basic Books, New York. [Also published by Harper & Row, New York, 1971.]
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. Harper and Row, New York.



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>
- Haugen, Einar (ed.). 1950. *First Grammatical Treatise*. The Earliest Germanic Phonology. An Edition, Translation and Commentary. Language Monograph 25. Baltimore, Maryland.
- Stubbs, Michael. 1980. *Language and Literacy: The Sociolinguistics of Reading and Writing*. Routledge, London.
- Thráinsson, Höskuldur. 1994. Icelandic. In Ekkehard König and Johan van der Auwera (eds.): *The Germanic Languages*, pp. 142–189. Routledge, London.
- Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge University Press, Cambridge.
- Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen and Zakaris Svabo Hansen. 2012. *Faroese*. An Overview and Reference Grammar. 2nd edition. Fróðskapur, Tórshavn, and Linguistic Institute, University of Iceland, Reykjavík.

Spelling can be difficult ...

(copied from Cooks web-site)

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

According to a research at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and last letter be at the right place. The rest can be a total mess and you can still read it without a problem. This is because the human mind does not read every letter by itself, but the word as a whole. Amazing huh?

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