

University of New Hampshire

University of New Hampshire Scholars' Repository

Inquiry Journal 2007

Inquiry Journal

Spring 2007

The Only Differences are the Words and the Sounds: Register Variation in Modern Written Icelandic

Jim Wood
University of New Hampshire

Follow this and additional works at: https://scholars.unh.edu/inquiry_2007



Part of the [Linguistics Commons](#)

Recommended Citation

Wood, Jim, "The Only Differences are the Words and the Sounds: Register Variation in Modern Written Icelandic" (2007). *Inquiry Journal*. 21.

https://scholars.unh.edu/inquiry_2007/21

This Article is brought to you for free and open access by the Inquiry Journal at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Inquiry Journal 2007 by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact Scholarly.Communication@unh.edu.



research ARTICLE

The Only Differences are the Words and the Sounds: Register Variation in Modern Written Icelandic

—**Jim Wood** (Edited by Lee Fetters)

We learn our native language by being exposed to it. Most of it is not explicitly taught. We might remember being taught to say “My friend and I” instead of “Me and my friend,” but no one ever really taught us to say, for example, “the friend” instead of “friend the.” We learn these types of things from the language we are surrounded by as children and also from linguistic properties of our brain that we humans are biologically born with. This second source, our “language instinct,” gives us the flexibility to learn the language of our environment perfectly (its words and sounds), so that we can easily, as children, break down a string of sound (speech) into meaningful parts (such as words).



The author overlooking the main town in Vestmannaeyjar, a small chain of islands off the southern coast of Iceland.

Understanding language is a central aspect of understanding ourselves biologically, socially, and culturally. Biologically, the ability to combine small pieces of sound into an infinite number of abstract representations (such as sounds, words, and phrases) seems to be a communication system belonging solely to human beings, the communication systems of other animals being qualitatively different in many important ways.

As social animals, we use this ability in ways that reveal aspects of our social behavior: When I hear someone say *ain't* and assume them to be uneducated, I am making a *social* judgment based on an arbitrary symbolic distinction; *ain't* is no less logical than *isn't*. Lastly, some uses of language are purely cultural conventions. Saying, “May I use your bathroom?” instead of “Take me to your bathroom” is a useful example. Notice that neither example is intrinsically more polite; each demands the same response on the part of the listener, who in neither case really has the option of saying “No.”

A great deal of linguistic research has been devoted to understanding exactly which parts of language are mechanisms of our language instinct, which parts are reflections of universal social behavior, and which parts are culturally inherited. Thanks to a Summer Undergraduate Research Fellowship I spent the summer of 2005 investigating how similar the patterns of language use in Icelandic are to those in English.

Why Icelandic?

A few years ago I was stationed in Naval Air Station Keflavík, Iceland, with the 56th Rescue Squadron (US Air Force) as a helicopter mechanic. I was curious to see how much of the local language I could learn in a one-year period. At first glance, Icelandic didn't seem to have a lot in common with English. It had letters like *þ*, *æ*, *ð*, *ö*, *á* and could use sentences structures alien to English, such as "The mouse kissed the cat" when they really



The author touching the ground of Vestmannaeyjar, still warm (and steaming) from a volcanic eruption over 30 years ago.

meant that the cat kissed the mouse. It wasn't long, though, before I noticed how strikingly similar the two languages were, and that the differences were mostly superficial. Their way of using negatives was reminiscent of a more Shakespearean English: *I say not*. Their *wh*-words (*who*, *what*, *when*, *where*, *why*) all had an *hv* instead of the *wh*: (in the same order) *hver*, *hvað*, *hvenær*, *hvar*, *hvers vegna*. Like *wh*-words in English, they appeared at the beginnings of information questions. Finally, their prepositions and adjectives came before the nouns they modified, just like in English: ***in the house***, ***red cars***.

What is Register?

When I finished my military service, I decided to major in linguistics at the University of New Hampshire. I learned in several linguistics classes (for the interested non-major, I would recommend Eng 405, Eng 719, Eng 791, or Eng 752) about a type of language variety called *register*. A register is a variety of language defined by the social or cultural situation in which it is used, typically exhibiting linguistic features intrinsic to, or revealing of, that variety. For example, imagine that you are handed two descriptions of the same tornado striking the same house, both 300 words long. One, however, was written for a newspaper, and the other was written by an eye-witness blogger. You would be able to tell, without any special training, which was the newspaper and which was the blog, even though they are both describing the same event in the same language. This is because, along with the language we learn, we also learn different ways of appropriately exploiting that language to give the hearer (or reader) information about the context at hand. Studying inter-language register variation, then, can shed some light on how humans use language: specifically, what kinds of features of language we use to give the reader this situational information.

Before undertaking research of my own, I familiarized myself with some previous research on register. One linguist, Douglas Biber, had designed a particularly interesting way of studying register: by counting various linguistic features in various registers and looking for significant differences. Biber focused on English and used computer software to count the frequency of sixty-seven features of English in various registers. He then looked at which features seemed to “go together” in certain contexts. For a simple example, he found that when prepositions were common, personal pronouns were not, and vice versa. (Think Clark Kent and Superman: when one is there, the other is not.) This is called *complementary distribution*. When Biber found groups of features that bundled together in complementary distribution, he looked at the types of texts in which those features seemed to bundle together. In the above example, prepositions and the features they co-occurred with (*Bundle A*, see list below) were very common in texts whose purpose was to impart information. In contrast, pronouns and the features they co-occurred with (*Bundle B*) were common in texts whose purpose was interactive. Thus, informational features and interactive features seem to have some sort of relationship: they avoid each other. The more informational a text is, the fewer *Bundle B* features it contains. Likewise, the more interactive a text is, the fewer *Bundle A* features it contains (2).

Biber found six such featural relationships in English and called them *dimensions* of register variation. The dimension I just described is called Dimension 1, “Interactive vs. Informational.” Since Biber's original study, similar studies have been done in various other languages such as Korean, Somali, and Nukulaelae Tuvaluan. These studies have shown a great deal of language use to be similar across even quite different types of languages, but no one has really done this sort of research in Icelandic. For my research, I decided that I wanted to see in what sorts of ways Icelandic varied according to register, and I based much of my project on Biber's findings for English. My hypothesis was that when I looked at features of Icelandic that were structurally and functionally similar to English (such as the ones mentioned above), they too would vary along a dimension such as Biber's Dimension 1.

What Could I Hope to Gain from This?

I asked three basic research questions for this study. First, what kinds of linguistic features give a reader situational information, i.e., distinguish two or more registers? Second, to what extent do these features cross linguistic boundaries? Third, since Icelandic exhibits word orders that are not possible in English, does word order ever serve as a feature of register variation?

For the first question, I was able to refer extensively to previous research. I decided that I would operate within Biber's Dimension 1 (Interactive vs. Informational) to decide which features were relevant to the distinction between blogs and newspapers. To answer the second question, I chose a subset of the features relevant to this Dimension: six *Bundle A*, or informational features, and six *Bundle B*, or interactive features. My goal was not only to pick linguistic features that characterized English registers along this dimension, but also to pick those features which seemed to be the least different between English and Icelandic. So a feature such as “passive sentences” would have been less useful because the functional use of English passives corresponds to several different structures in Icelandic, structures which don't necessarily correspond directly back just to English

passives. Prepositions, on the other hand, are structurally similar, e.g., they precede their nouns, and are used essentially the same way in both languages, thereby lending themselves nicely to a cross-linguistic comparison. The other eleven features, as far as I could tell, also conformed to these criteria. For the third question, I counted instances of about a dozen possible word-order types to see if any were more prevalent in one register than the other.

Puzzling the Papers

I compiled two 10,000-word sample texts, or *corpora*, to represent two registers of Icelandic. I used the Internet to do this because it offers immediate access to real-life language use. My registers, chosen to test the interactive vs. informational dimension, were Icelandic blogs and online Icelandic newspapers. I extracted an equal amount of text from each of seven Icelandic bloggers. For the newspaper corpus, I pulled articles from three online newspapers.

As mentioned above, I picked out the features from Biber's English Dimension 1 which I thought were the most similar in Icelandic. The informational and interactive features I chose were:

Bundle A – Informational Features

- Regular nouns (as opposed to pronouns)
- Word length—more long words (counted in letters)
- Prepositions such as *in, around, by, under, for*, etc.
- Type-token ratio (the number of different words per total number of words)
- Attributive adjectives (adjective precedes noun, for example: *the red car* as opposed to *the car is red*.)
- Location adverbials (for example, *upstream, abroad, on board*, etc.)

Bundle B – Interactive Features

- “Be” as a finite verb (for example, *am, is, was*, etc.)
- Analytic negation “not”: *I do not have an answer* vs. *I have no answer*.
- Wh-questions (*who, what, when, where, why, and how*)
- Demonstrative pronouns (for example, “that” in *That is what I said*)
- Indefinite pronouns (for example, *someone, no one, something*, etc.)
- First- and second-person personal pronouns (*I, you, and we*)

I counted the presence of each of these features in both texts, predicting that the interactive features would be significantly higher in blogs than in newspapers and the informational features would be significantly higher in newspapers.

Icelandic allows some word orders that English does not allow; English has a more rigid word order than many other languages. This presented a good opportunity to see how word order might be exploited in a register-specific way. After spending considerable time with the texts, looking at and counting various word orders, I noticed that newspapers, much more than blogs, seemed to exhibit a particular word-order pattern that is not used in English. I decided to count this type of word order, known as *Stylistic Fronting (SF)*, in two very common types of clauses (3).

The first kind is called a (passive) impersonal clause, which is a clause without an agent (or subject, more generally) in the sentence. In Icelandic, impersonal clauses can take one of two forms—*Stylistic Fronting* or *Expletive Insertion*—whereas in English only *Expletive Insertion* is a possible word order. Both mean “It was said that...”

Impersonal Clauses

- **Stylistic Fronting** (Icelandic Only)
Sagt var að...
said was that...
- **Expletive Insertion** (English and Icelandic)
Það var sagt að...
It was said that...

Það or “it” (in Icelandic and English respectively) is an *expletive*, a word which doesn't really refer to anything but is there just to fill in the subject space. *Stylistic Fronting* is now relatively unique to Icelandic and its relative Faroese, although it used to exist in Old French, Old Spanish, Old English, Old Danish, and Old Catalan (4), (5).

The second type of clause I examined is a relative clause, or a clause that modifies a noun. Again, with (subject) relative clauses, both *Stylistic Fronting* and *No Stylistic Fronting* are possible in Icelandic, but only *No Stylistic Fronting* is possible in English.

Relative Clauses:

- **Stylistic Fronting** (Icelandic Only)
Kötturinn sem tekinn var
the cat which taken was
- **No Stylistic Fronting** (English and Icelandic)
Kötturinn sem var tekinn
the cat which was taken

Icelandic Behaves like English

For most features, I divided the number of times each feature appeared by the number of times it could have appeared to get some easily comparable numbers. Figure 1 (see Appendix) shows the presence of the interactive features in blogs and newspapers. As the graph shows, all features acted as predicted: they all behaved the same way in Icelandic as in English.

However, only four of the informational features could be graphed in the same way as the interactive features. Two features, mean word length and type-token ratio, could not be counted in terms of instances per opportunities and are given, instead, in raw numbers. Figures 2, 3, and 4 show these results (see Appendix). Again, all features which Biber had shown to be informational in English were more prevalent in Icelandic newspapers than in blogs. These features, then, carry much of their situational function across language boundaries, or at least across the boundary between English and Icelandic.

To make sure that these distributions weren't random accidents, Professor Naomi Nagy showed me how to test for statistical significance. We conducted what is known as a chi-square test, which tests the likelihood of a distribution (of any sort) being due to chance. In these cases, all distributions had less than a 5% chance of being random, and most were as low as .1%. Thus, these features are being exploited by speakers/writers at a statistically meaningful level not by accident but without speakers intending it.

To compare the presence of Stylistic Fronting word order in Icelandic with that of Expletive Insertion (in impersonal clauses) or No Stylistic Fronting (in relative clauses), I found every instance where both were grammatically possible. I noted which structure was chosen in each case and calculated the percentage of each within its register. The pie charts in Figures 5 and 6 (see Appendix) show these results.

In both cases, the word order which is not possible in English (Stylistic Fronting) was far more common in newspapers than in blogs, whereas the more English-like structures were more common in blogs than in newspapers. In Icelandic, then, Stylistic Fronting is a feature of register variation, something speakers use to give situational information. In English, a similar contrast might be made between *The cup I put change in* vs. *The cup in which I put change*, the latter being akin to Icelandic's Stylistic Fronting in the sense that it is likely to be used in more formal registers such as newspapers.

Nature and Nurture

Some features of register seem to give the same situational information in Icelandic as in English. Others, such as Stylistic Fronting, are language-specific. So how do language users know how to use and recognize these markers? To some extent, they could be taught. For example, journalists undergo significant training concerning what is and is not appropriate language in their profession. But bloggers are never taught to use more indefinite pronouns and fewer attributive adjectives. Registers vary in ways their speakers could not possibly intend. Many of the ways in which this variation occurs are similar across linguistic boundaries. These

two facts are evidence that language users do not learn these varieties explicitly, but tacitly. That is, given enough exposure to a register, a language user will acquire the properties of that register without necessarily meaning to, as part of his/her acquisition of the language.

Competency in a register does not have to extend to production either. Just as many people can understand a foreign language better than they can produce it, some people have enough exposure to a register, such as newspapers, to be able to recognize it without being able to produce it, that is, write a newspaper article. The more a language user practices producing in a register, the more he or she will become competent in exploiting the features considered appropriate for that register. Most of us have had enough exposure to newspapers to be familiar with their properties, and we would recognize a newspaper article even if we could not explain the actual linguistic features which facilitated that recognition. We have, therefore, tacit, unconscious knowledge of language varieties, including registers. Eventually the extensive study of cross-linguistic register variation will reveal a great deal about what kinds of situational variation are language-dependent, what kinds are culture-dependent, and what kinds are universal. The more we examine our biological, social, and cultural behavior, the more we will understand about ourselves, the talking primates.

I would like to thank my mentor, Dr. Naomi Nagy, for assisting my research every step of the way, for editing numerous drafts at various stages of the project, for offering suggestions and references for further consideration, and for showing me how to test for statistical significance. I would also like to thank Gunnar Hrafn Hrafnbjargarson, Lund University, for his correspondence concerning several aspects of Icelandic linguistics; Miriam Meyerhoff, University of Edinburgh, for directing me to some interesting and relevant papers; Peter Akerman and Donna Brown, along with the staff of the University Research Opportunities Program (UROP), for their support; and Dana Hamel, whose donation to UROP made my research possible through a Summer Undergraduate Research Fellowship. All mistakes and shortcomings are my own responsibility.

References

1. S. Pinker. *The Language Instinct*. (Harper Perennial Modern Classics, 2000).
2. D. Biber. *Dimensions of Register Variation*. (Cambridge University Press, New York, 1995), pp. 141-235.
3. J. Maling. Inversion in Embedded Clauses in Modern Icelandic. In *Syntax and Semantics: Volume 24, Modern Icelandic Syntax*. J. Maling and A. Zaenen, Eds. (Academic Press, New York, 1990).
4. E. Mathieu. Stylistic Fronting in Old French. *Probus*. 18, 219-266 (2006).
5. G. Hrafnbjargarson. Stylistic Fronting. *Studia Linguistica*. 58(2), 88-134 (2004).

The blogs I used to assemble the blog corpus were as follows:

- http://spaces.msn.com/eliasblondal/PersonalSpace.aspx?_c=
- <http://www.larsson.com/blog>
- <http://arnor.bloggar.is/>
- <http://kkkson.blog.com/2005/12/>
- <http://sigrundogg.blogspot.com>
- <http://almasigurdar.blogspot.com>
- http://www.hi.is/~krs/stina/2004_02_01_stina_archive.html#107580609665205000

The newspapers I used to assemble the newspaper corpus were as follows:

- <http://www.mbl.is>
- <http://www.visir.is>
- <http://www.sudurland.is>

Appendix

Figure 1

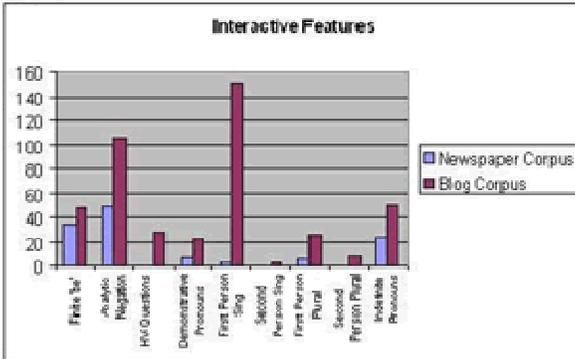


Figure 3

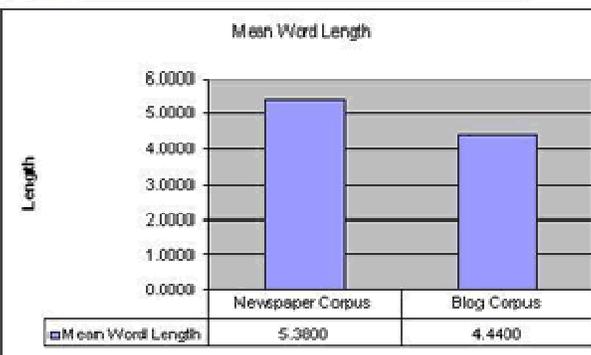


Figure 2

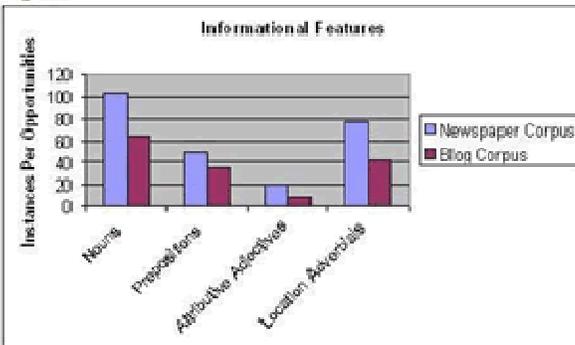


Figure 4

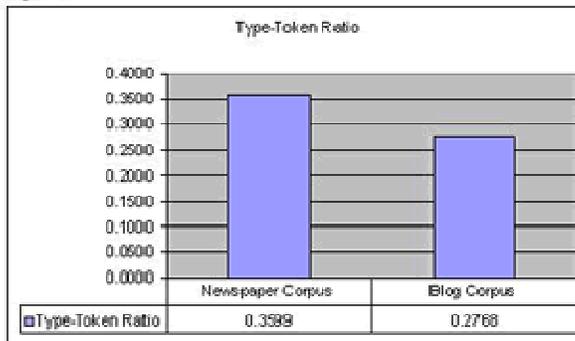


Figure 1 compares the presence of Interactive features in blogs and newspapers. In all cases, these features appeared more frequently in blogs than newspapers. Contrarily, **Figure 2** shows that newspapers had more Informational features than blogs. **Figures 3 and 4** also show specific Informational features, and, as expected, these features appeared more frequently in newspapers.

Figure 5 - Stylistic Fronting vs. Expletive Insertion in Impersonal Clauses

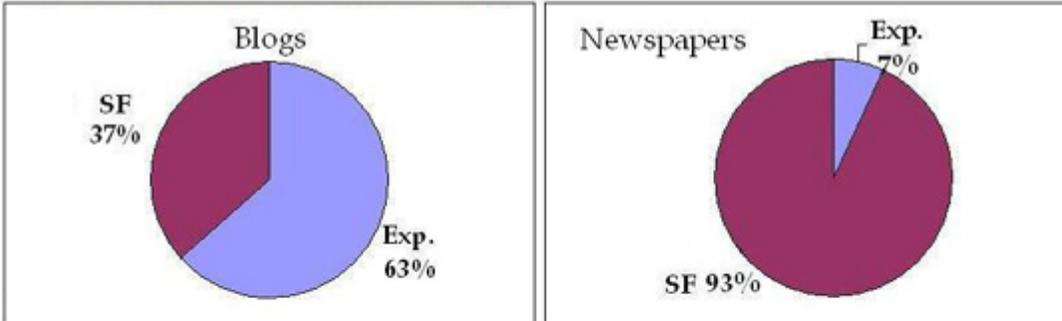
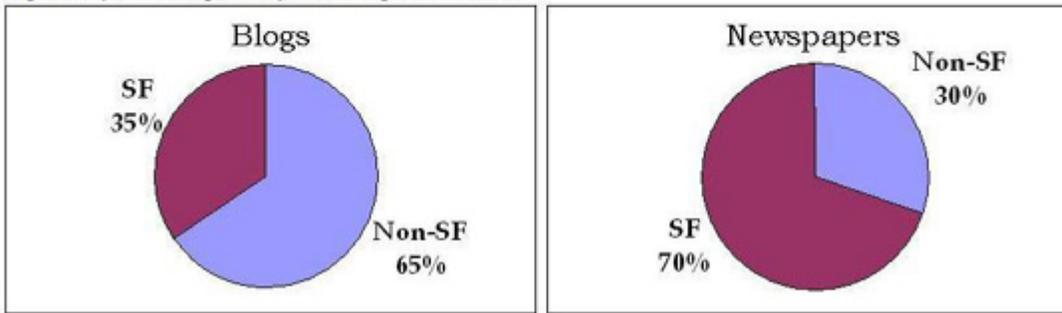


Figure 6 - Stylistic Fronting vs. No Stylistic Fronting in Relative Clauses



Figures 5 and 6: Stylistic Fronting was observed only a little bit more than 33% of the time when it would be possible in both impersonal clauses and relative clauses. In newspapers, however, it was used 93% of the time in impersonal clauses and 70% of the time in relative clauses. Stylistic Fronting seems to co-occur with the presence of Informational features.

Copyright 2007 Jim Wood

Author Bio

Jim Wood, a linguistics major from Nashua, New Hampshire, graduated with a bachelor of arts in December 2006. He became interested in linguistics and the Icelandic language after being stationed in Iceland with the Air Force. Wood performed much of his research over the summer of 2006 with a Summer Undergraduate Research Fellowship (SURF). His work allowed him to learn the different ways that people study register and gave him the chance to examine a wide range of linguistic theory. As he points out, "Understanding these concepts has allowed me to really see how Icelandic is not very different from English in many important respects." Next fall Jim will begin graduate school at New York University in pursuit of a Ph.D. in linguistics. Someday he hopes to own a summer home in Iceland.

Mentor Bio

Dr. Naomi Nagy is an associate professor of English and is the coordinator of the linguistics program at the University of New Hampshire, where she has taught for eleven years. She specializes in sociolinguistics, studying variation in how people speak and how that is used to represent identity. Dr. Nagy has worked with six Summer Undergraduate Research Fellowship (SURF) awardees, as well as with several students on honors projects, theses, and independent studies. She was particularly excited working with Jim Wood: "He's making really excellent new connections between different parts of the field of linguistics, and it's been invigorating to follow his exploration," she commented. Dr. Nagy appreciates Jim's use of innovative methods and data in researching a source that isn't commonly looked at—online Icelandic.

