Transcription Accuracy in the Icelandic Gigaword Corpus: Evaluating Automatic and Manual Annotation

Johanna Mechler, Lilja Björk Stefánsdóttir, Anton Karl Ingason
University of Iceland
Sæmundargata 2, 102 Reykjavík, Iceland
mechler@hi.is, lbs@hi.is, antoni@hi.is

Abstract

This paper aims to compare automatic and manually corrected annotation data in the Icelandic Gigaword Corpus. We focus on the variable use of Stylistic Fronting (SF) in Icelandic, an optional movement of words or phrases, which indicates a more formal style. Examining SF rates across time, we find that manual coding results in slightly lower SF rates than automatic coding. This difference can be explained by the different sources used in the coding process: For automatic coding, written transcripts compiled by parliament employees are used, and for manual correction, coding relies on audio files of the parliament speeches. Importantly, both types of coding are well suited to trace changing patterns of SF over a span of 16 years, suggesting that the automatic feature extraction reliably reflects the speeches that have been transcribed.

Keywords: evaluation, annotation, quality assurance, sociolinguistics, Icelandic

1. Introduction

In this paper, we evaluate the accuracy of automatic NLP coding of Stylistic Fronting (SF) in the Icelandic Gigaword Corpus (Steingrímsson et al., 2018) and find that despite minor differences between automatic and manual coding, the automatic coding can be reliably used for sociolinguistic research.[1]

Large linguistic corpora have become essential resources for studying language variation and change. As corpus-based research increasingly relies on automatically transcribed and annotated materials, questions of data reliability have gained renewed importance. While much corpus reliability research has focused on annotation accuracy, that is, how consistently part-of-speech tags, morphological features, or syntactic parses are assigned (e.g., Ide and Brew, 2000; Giesbrecht and Evert, 2009), less attention has been paid to transcriber effects, the differences between what was said and what is documented in corpora (but see Stefánsdóttir and Ingason, 2025). A similar concern has emerged in projects such as CHILDES (MacWhinney, 2000, 2014), where transcription fidelity and coder variability have been shown to affect analyses of child speech and developmental patterns. Comparable debates have also emerged in treebank and speech corpus research (e.g., Marcus et al., 1993; Shriberg et al., 2001). These debates highlight that reliability depends not only on the consistency of annotations but also on the interpretive processes involved in transcription.

Our paper addresses this gap by examining how transcription practices influence the analysis of SF in Icelandic, using data from the Icelandic Gigaword Corpus (Steingrímsson et al., 2018) and, more specifically, its subcorpus, the Icelandic Parliament Corpus. SF, an optional movement of words or phrases to clause-initial position where there is a subject gap (see examples (1) and (2), where (2) involves SF), has long been recognized as a marker of formal style in Icelandic (Maling, 1980; Holmberg, 2000, 2006; Thráinsson, 2007; Angantýsson, 2017; Ingason and Wood, 2017).

(1) *Varðandi það* $[_{CP}$ *sem* ***var*** ***sagt*** *hér]*
regarding it that **was** **said** here
'Regarding what was said here'

(2) *Varðandi það* $[_{CP}$ *sem* ***sagt*** ***var*** *hér]*
regarding it that **said** **was** here
'Regarding what was said here'

Ongoing research on SF's use as a stylistic choice across the lifespan shows that its frequency is sensitive to a variety of sociolinguistic forces, such as changes in speakers' social position (Stefánsdóttir and Ingason, 2018), identity negotiation (Stefánsdóttir and Ingason, 2024a), life events (Stefánsdóttir and Ingason, 2024c), institutional roles and/or authority (Stefánsdóttir and Ingason, 2024b; Mechler et al., 2025; Stefánsdóttir et al., Forthcoming). Therefore, its sociolinguistic sensitivity makes SF a necessary test case for exploring how transcription and editorial practices shape quantitative linguistic findings.

Our study compares SF rates derived from two versions of Icelandic parliamentary speech data: one based on automatic annotation of written transcripts produced by parliamentary staff, and an-

[1] In this paper, we use the terms *coding* and *annotation* synonymously.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 4757–4764
11–16 May 2026.

other based on manual correction informed by audio recordings of the same speeches. The comparison reveals systematic differences in SF frequency across these datasets, highlighting the implications of transcription sources for corpus-based linguistic research.

This paper is organized as follows: Section 2 provides details on the background of this paper, more specifically the Icelandic Gigaword Corpus and the parliamentary transcripts, which form the foundation for the two types of annotation and our evaluation. Section 3 outlines the methods, data extraction, and coding processes as well as the data set analyzed in this study. Section 4 offers insight into the sociolinguistics of SF and political speech by examining the SF trajectories of three members of parliament (MP). We follow with the evaluation of the accuracy of the automatic coding in Section 5, comparing it to the manual corrected version based on audio recordings. Finally, Section 6 summarizes the main findings and implications for evaluative practices in dealing with large speech corpora.

2. Background

The Icelandic Gigaword Corpus (Steingrímsson et al., 2018) is an extensive, morphosyntactically annotated collection of 2,429 million running words of text, drawn from a range of written and transcribed sources (Barkarson et al., 2022). Among its subcorpora, the Icelandic Parliament Corpus is particularly notable for its extensive representation of political discourse, comprising transcriptions of parliamentary speeches that span more than a century.

The parliamentary transcripts are produced by employees of the parliament, who transcribe debates from raw audio recordings. These official transcriptions are designed primarily for readability and archival consistency, rather than for a detailed study of linguistic phenomena. As such, they tend to regularize or normalize the spoken language, such as removing repetitions and correcting poorly structured sentences, and occasionally adjusting word order or phrasing to align with standard written Icelandic conventions (Fong et al., 2018). This type of editorial normalization is common in institutional transcription practices and has been shown to introduce subtle yet sometimes systematic distortions in linguistic data (Ochs, 1979; Edwards and Lampert, 1993; Bucholtz, 2007).

This reinforces the longstanding claim that transcriptions are not a neutral or mechanical process; instead, they are a theoretical representation of speech shaped by various methodological and institutional choices (Ochs, 1979; Duranti, 1997; Bucholtz, 2000). Consequently, researchers must treat transcripts not as neutral data but as mediated artifacts that reflect specific theoretical and procedural assumptions. Recognizing and controlling for transcriber effects, the subtle but cumulative influence of transcription practices on speech data, is essential for maintaining the validity and reliability of findings derived from such corpora.

3. Methods and Data

Our work is based on data coded by an automatic part of speech tagger and lemmatizer, using written transcripts in the *Icelandic Gigaword Corpus* as a source (Steingrímsson et al., 2018). The current version of the corpus is tagged with ABLTagger 3.0.0 (Steingrímsson et al., 2019) and lemmatized with Nefnir (Ingólfsdóttir et al., 2019). Examples where SF may apply are then extracted with a custom Python script. The error rate of these tools is very low, so we are confident in the automatic SF coding it provides. Moreover, we are only interested in one specific construction with a fixed type of syntax, always introduced by the relative complementizer *sem* 'that', and we have found no errors in the automatic tagging and lemmatization of this construction. It should also be noted that these current tools significantly improve on the accuracy of previous tools for Icelandic, such as the tagger IceTagger (Loftsson, 2008) and the lemmatizer Lemmald (Ingason et al., 2008). However, it is crucial to evaluate how the written transcripts compare to the actual speech data, which is what we contribute with this study.

The complete data set is part of the ERC-funded project *Explaining Individual Lifespan Change* (EILisCh), which investigates linguistic lifespan trajectories of individual MPs and entire political parties based on the written transcripts of political speeches (e.g., Mechler et al., 2025). To assess transcription accuracy for the individual case studies conducted within the project, manual checks of the automatic coding were completed for a subset of data where audio tracks were available (2005–2021)[2]. The final subset includes a total of 20 MPs. In addition to the individuals of the case studies, further MPs were selected to ensure a balanced data set, stratified by MPs' year of birth (range: 1942–1990, mean: 1966.1), gender (10 women, 10 men), and political party affiliation (eight parties across the whole political spectrum). This subset of manual checks (n=25,096) also allows for a bal-

[2]Since 2005, the Icelandic parliament has provided audio recordings of parliamentary speeches. Audio files from earlier periods are text-to-speech synthesized from the transcripts and do not represent the original spoken delivery. For this reason, manual verification of transcription accuracy is only possible for speeches delivered from 2005 onward.

anced assessment of transcription accuracy, since it represents 67.43% of the overall data available for these speakers (*n*=37,219) (see Table 1). Table 1 gives an overview of the data set for all 20 MPs and three individual MPs, whose analyses we will focus on separately in the subsequent sections. For all three MPs, manual checks were completed for more than 71% of each data set (see Table 1).

Data set	Automatic (n)	Manual (n)	Percent. (%)
All 20 MPs	37,219	25,096	67.43
Ásmundur Daðason	1,946	1,478	75.95
Bjarni Benediktsson	4,819	3,659	77.54
Steingrímur Sigfússon	4,339	3,115	71.79

Table 1: Overview of data sets with two annotation types (automatic, manual).

Three independent coders compared the transcripts with the corresponding audio; each token was coded once for presence/absence of SF. Since this classification is very obvious for anyone who listens to the audio, there are virtually no instances of disagreement between annotators. For our dataset, this was validated as two annotators independently coded a subsample (*n*=980), achieving an inter-annotator agreement of 99.29 % (Cohen's Kappa = 0.96, i.e. almost perfect agreement).

Instances of divergence between the audio and the transcripts were annotated and classified according to error type: whether SF was added (*n*=2,627) or omitted (*n*=142) by transcribers in parliament or if it was outside of the envelope of variation (*n*=812). The third category refers to instances where the transcribers in parliament had altered the wording such that the possibility of SF is present in the transcript, even though in the actual speech the speaker did not have a chance to employ SF. This procedure provided a basis for evaluating transcription quality and ensuring reliability in subsequent linguistic analysis. In the following analysis, which was conducted in R (R Core Team, 2024), we excluded all NAs (*n*=18), ambiguous cases (*n*=100) and tokens outside of the envelope of variation (*n*=812; total excluded: *n*=930).

It should be noted that we did consider using Automatic Speech Recognition (ASR) to avoid the issue of transcriber effects, but we decided against this option because previous research has found that the quality of ASR on exactly this corpus is not sufficient for doing precise work without incorporating manual corrections anyway (Fong et al., 2018). We leave it for future work to consider whether this option will become more feasible as new ASR systems for Icelandic become available.

4. The Sociolinguistics of Stylistic Fronting and Political Speech

Although our main focus here is to evaluate the accuracy of the automatic coding that relies on the written transcripts of the parliament speeches, we will contextualize the overall research program by briefly describing the type of findings that are derived from the coding in question. Figure 1 shows the evolution of the use of SF in the speeches of three Icelandic MPs, Ásmundur Daðason, Bjarni Benediktsson, and Steingrímur Sigfússon.

First, Ásmundur Daðason (*n*=1,478) decreases his use of SF between 2012 and 2014 as well as in 2016. The study in Stefánsdóttir and Ingason (2024c) found that these slight decreases correspond to periods where Daðason is undergoing personal crises which he does not have much control over. In the first period, he switched parties and became unpopular, there were claims of his resignation from the parliament, and he faced a negative backlash from the media. The second decrease is during a time when it was reported on the news that he had been intoxicated on a commercial airplane, where he supposedly vomited on another passenger. Thus, during these periods when he lays low and manifests an apparently reactive response to personal crises, this goes hand in hand with decreased SF use.

Second, Bjarni Benediktsson (*n*=3,659) decreases his use of SF after the economic crash of 2008, a transition that involves him going from a role in the majority of the parliament to one in the minority (Stefánsdóttir et al., Forthcoming). His SF use rises again when he enters the majority once more and becomes minister of finance. This is consistent with an interpretation according to which Benediktsson gains a higher linguistic market value (Sankoff and Laberge, 1978) during this time. This means that the importance of language in his life increases which results in a greater use of formal/standard variants. However, Benediktsson gradually uses less SF in the years following 2015. This may be a response to crises, similar to Daðason. In these years Benediktsson was repeatedly in the news due to high-profile controversies, including a revelation of his Ashley Madison account, "IceHot1" (Fontaine, 2015), ties to prior offshore holdings revealed in the Panama Papers, and a short-lived government led by Benediktsson that collapsed due to a scandal involving his father's actions.

Third, Steingrímur Sigfússon (*n*=3,115) increases his use of SF in the years leading up to the economic crash of 2008 and peaks in his SF use in the years immediately following the crash

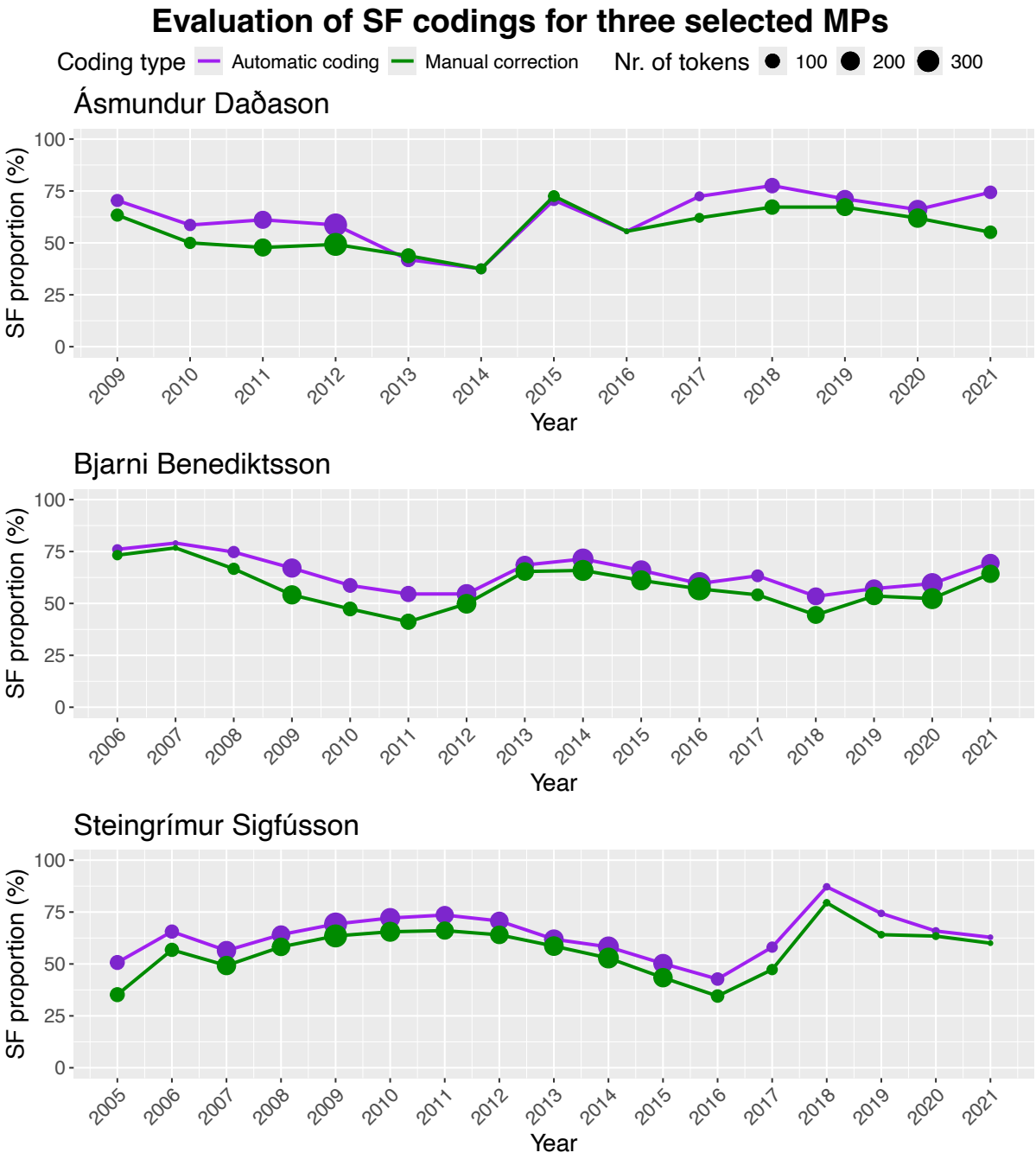

Figure 1: Individual accuracy analysis for Ásmundur Daðason (n=1,478), Bjarni Benediktsson (n=3,659), and Steingrímur Sigfússon (n=3,115), evaluating automatic and manual coding.

(Stefánsdóttir and Ingason, 2018). In the years prior to the crash he was seen as the unofficial leader of the opposition in parliament and thus he gained linguistic market value (Sankoff and Laberge, 1978) which may lead to heightened use of formal/standard variants. This trend of his gaining greater responsibility continues after the crash when he enters the majority and becomes minister of finance, and his SF use peaks during this period. After he steps down from this role, his SF use decreases again. So in his case, a national crisis during which he assumes a proactive role, and which increases his linguistic market value, leads to a greater use of SF. In the final years of the graph we see a different pattern where he speaks much less but does so in a very formal style. This is when he becomes speaker of the parliament and his speeches are more about managing the operations of the parliament than policy matters.

In sum, these types of case studies, only briefly described here, demonstrate that the coding that we are analyzing can be used to track the evolution

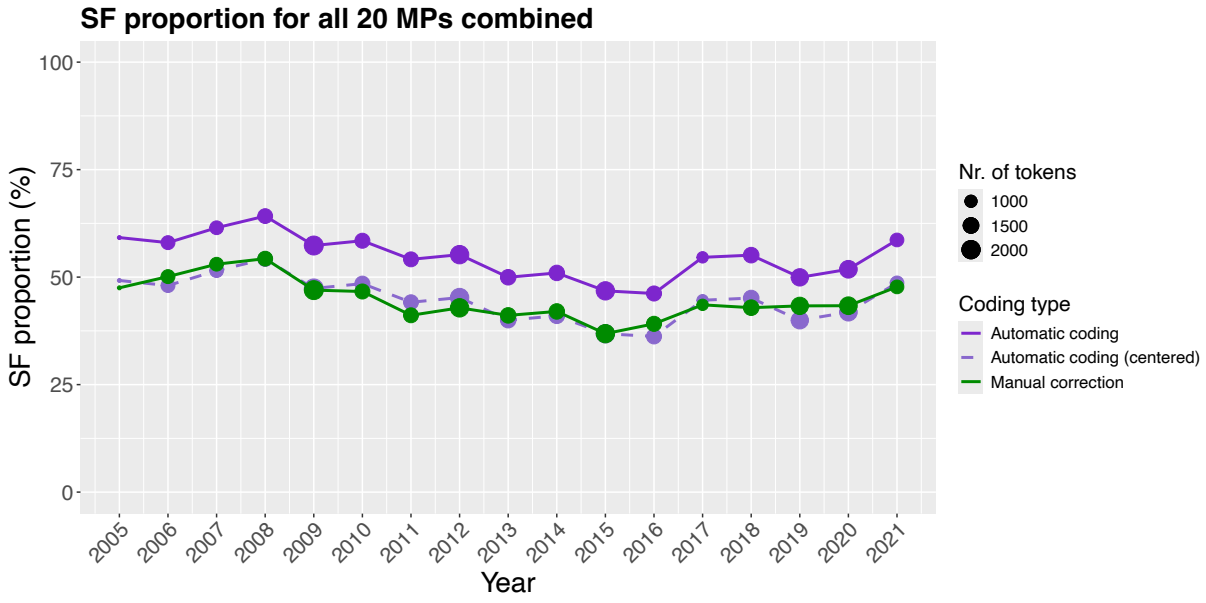

Figure 2: Accuracy analysis for all 20 MPs combined (total *n*=50,192), evaluating automatic and manual coding (2005–2021).

of formal speech for individual politicians, and in turn such changes can be tied to a sociolinguistic interpretation of linguistic lifespan change.

5. Evaluation of the Accuracy of Automatic NLP Coding

When testing for differences between automatic and manually corrected coding, we are mainly interested in how accurately the automatic coding can trace SF rates across time. The evaluation of the three individual MPs, Ásmundur Daðason, Bjarni Benediktsson, and Steingrímur Sigfússon, demonstrates that automatic coding largely mirrors the manually corrected data in capturing fluctuations in SF use, while also revealing some discrepancies that highlight the influence of transcriber effects. Figure 1 depicts the SF distribution for all three politicians as documented by automatic and manual coding. Here again, both annotation types report their variable SF use effectively, unveiling complex specifics of their SF trajectories over time.

Expanding the framework to a larger data set of 20 MPs solidifies the previous finding that the automatic coding traces trends and changes across years well. In Figure 2, we show the analysis of automatic and manual annotation for the selected 20 MPs (total *n*=50,192) between 2005 and 2021. It shows that despite some discrepancy between the automatic and manual coding (X^2(1, 50,192)=491.82, p<.001), the patterns/individual trajectories remain unchanged. These results are consistent with known editorial normalizations practices that are common in institutional transcription, where transcribers adjust spoken language to conform to written standards, e.g. by removing disfluencies or repetitions that might be perceived as informal or stylistically inconsistent with the official record.

In very rare cases, the annotators found an instance of SF in the audio file that was not in the written transcript (see Section 3). However, in the majority of cases when there was a mismatch between the automatic and manually corrected coding, it was caused by a "mistake" (i.e., an added case of SF) in the written transcript. In this way, the MPs come across as more formal, more professional, by transcribers in parliament than they were in reality giving their speech. While the transcripts of the parliament speeches aim to reflect the actual speeches as closely as possible (Skrifstofa Alþingis, 2017), they are not crafted to uphold scientific standards of a close transcript as done for linguistic analysis. It is interesting to note then the stylistic difference between the written transcripts and the audio files.

Despite these differences, our analysis demonstrates that the written transcripts are indeed reliable for sociolinguistic analysis as they trace changes across time accurately, both for individual MPs as well as larger cohorts. We can illustrate this point more clearly by centering the mean proportion of the automatic coding around the mean proportional values of the manual correction. The dashed line in Figure 2 shows the result of this calculation as centered automatic coding. Comparing now the centered automatic coding with the manual coding, we note only minimal

differences between the two trajectories. Thus, we argue that discrepancies between automatic and manual annotation are mainly the result of differences in the quantity of SF, both reflecting important shifts across time.

Crucially, the consistency between the two datasets demonstrates that such transcriber effects have only a limited impact on large-scale quantitative results, as both annotation types reveal parallel diachronic patterns in SF use. The close alignment shows that both types of coding are well-suited to track linguistic malleability over time.

6. Conclusion

The primary aim of this paper was to evaluate the reliability of written parliamentary transcripts as a linguistic data source. By comparing automatically annotated data with manually corrected data based on audio recordings, the study provides evidence on how transcription practices can affect corpus-based syntactic analysis. The present study expands substantially on the dataset examined by Stefánsdóttir and Ingason (2025), incorporating a much larger and more diverse range of parliamentary sessions and speakers. Our findings show that the rates of SF derived from manually corrected coding closely align with those obtained from the automatic coding process, although the manually corrected data consistently yield slightly lower SF proportions.

Importantly, the differences observed should not be interpreted as errors introduced by the automatic coding tool. Rather, they reflect transcriber effects, the subtle, systematic discrepancies between spoken language and its written representation. The written transcripts used in the automatic coding represent a slightly edited and standardized form of speech while the manually corrected dataset draws directly from raw audio, preserving more of the variability and spontaneity of spoken language. These contrasts highlight a critical methodological consideration in corpus-based linguistics: transcripts are not an act of representation, but a selective process shaped by institutional conventions and stylistic norms. Recognizing this allows for a more precise interpretation of corpus-derived results while ensuring confidence in the reliability of Natural Language Processing systems and feature extraction tools, which effectively capture broad syntactic patterns with high accuracy.

In sum, this study illustrates the effective application of automatic annotation to large-scale spoken data and how it compares to manually corrected coding. It also underscores the importance of contextualizing data by acknowledging the nature of the underlying data: whether it is written, transcribed, or spoken, and the interpretive practices that shape it. Integrating both automatic and manual approaches enables a more comprehensive account of language use in institutional settings and enhances the methodological accuracy of corpus-based linguistic study. While manual correction offers greater precision when feasible, automatic annotation provides an efficient and sufficiently reliable basis for large-scale linguistic analysis.

7. Ethics Statement

For this analysis, we use only public data, none of which are sensitive, and we do not believe that there are any ethical issues that need to be addressed.

8. Limitations

While we do think that our paper makes certain points that are valuable to the wider field that we are working within, we do not want to overstate the generalizability of our findings. Here, we are focusing on one variable, Stylistic Fronting, in one language, Icelandic, and while our findings are discussed within the broader context of transcription practices in general, it would certainly be useful to compare our findings with similar studies that would be carried out on more variables and in other languages. We leave such comparisons for future work.

9. Acknowledgments

This project is supported by a grant from the European Research Council (ERC), project ID 101117824. We would like to thank the transcribers without whom this paper would not be possible. We would also like to thank the reviewers for helpful comments which improved this paper.

10. Bibliographical References

Ásgrímur Angantýsson. 2017. Stylistic Fronting and related constructions in the Insular Scandinavian languages. In Höskuldur Þráinsson, Caroline Heycock, and Zakaris Svabo, editors, *Syntactic Variation in Scandinavian. Studies in Germanic Linguistics*, pages 277–306. John Benjamins Publishing Company, Netherlands.

Starkaður Barkarson, Steinþór Steingrímsson, and Hildur Hafsteinsdóttir. 2022. Evolving large text corpora: Four versions of the Icelandic gigaword corpus. In *Proceedings of the Thirteenth*

Language Resources and Evaluation Conference, pages 2371–2381.

Mary Bucholtz. 2000. The politics of transcription. *Journal of Pragmatics*, 32(10):1439–1465.

Mary Bucholtz. 2007. Variation in transcription. *Discourse Studies*, 9(6):784–808.

Alessandro Duranti. 1997. *Linguistic Anthropology*. Cambridge University Press, Cambridge.

Janet A. Edwards and Martin D. Lampert. 1993. *Talking data: Transcription and coding in discourse research*. Lawrence Erlbaum, Hillsdale, New Jersey.

Judy Y Fong, Michal Borsky, Inga R Helgadóttir, and Jon Gudnason. 2018. Manual post-editing of automatically transcribed speeches from the Icelandic parliament - Althingi. *arXiv preprint arXiv:1807.11893*.

Andie Sophia Fontaine. 2015. Iceland's finance minister amongst Ashley Madison users. Reykjavík Grapevine. https://grapevine.is/news/2015/08/31/icelands-finance-minister-amongst-ashley-madison-users/.

Eugenie Giesbrecht and Stephanie Evert. 2009. Is part-of-speech tagging a solved task? An evaluation of POS taggers for the German web as corpus. In *Proceedings of the Fifth Web as Corpus Workshop*, pages 27–35.

Anders Holmberg. 2000. Scandinavian Stylistic Fronting: How any category can become an expletive. *Linguistic Inquiry*, 31(3):445–483.

Anders Holmberg. 2006. Stylistic Fronting. *The Blackwell Companion to Syntax*, pages 532–565.

Nancy Ide and Chris Brew. 2000. Requirements, tools, and architectures for annotated corpora. In *Proceedings of data architectures and software support for large corpora*, pages 1–5.

Anton Karl Ingason, Sigrún Helgadóttir, Hrafn Loftsson, and Eiríkur Rögnvaldsson. 2008. A mixed method lemmatization algorithm using a Hierarchy of Linguistic Identities (HOLI). In *Advances in Natural Language Processing*, pages 205–216, Berlin, Heidelberg. Springer.

Anton Karl Ingason and Jim Wood. 2017. Clause bounded movement: Stylistic Fronting and phase theory. *Linguistic Inquiry*, 48(3):513–527.

Svanhvít Lilja Ingólfsdóttir, Hrafn Loftsson, Jón Friðrik Daðason, and Kristín Bjarnadóttir. 2019. Nefnir: A high accuracy lemmatizer for Icelandic. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 310–315.

Hrafn Loftsson. 2008. Tagging Icelandic text: A linguistic rule-based approach. *Nordic Journal of Linguistics*, 31(1):47–72.

Brian MacWhinney. 2000. *The CHILDES Project: The Database*, volume 2. Psychology Press, New York.

Brian MacWhinney. 2014. *The CHILDES Project: Tools for Analyzing Talk, Volume I: Transcription Format and Programs*. Psychology Press, New York.

Joan Maling. 1980. Inversion in embedded clauses in Modern Icelandic. *Íslenskt mál*, 2:175–193.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Johanna Mechler, Lilja Björk Stefánsdóttir, and Anton Ingason. 2025. Language use of political parties over time: Stylistic fronting in the icelandic gigaword corpus. In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, pages 313–318.

Elinor Ochs. 1979. Transcription as theory. In E. Ochs and B. B. Schieffelin, editors, *Developmental pragmatics*, pages 43–72. Academic Press, New York.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

David Sankoff and Suzanne Laberge. 1978. The linguistic market and the statistical explanation of variability. In David Sankoff, editor, *Linguistic Variation: Models and Methods*, pages 239–250. Academic Press, New York.

Elizabeth Shriberg, Andreas Stolcke, and Don Baron. 2001. Observations on overlap: findings and implications for automatic processing of multi-party conversation. In *Interspeech*, pages 1359–1362.

Skrifstofa Alþingis. 2017. *Háttvirtur þingmaður – handbók um þingstörfin*. Published on the website of the Icelandic parliament.

Lilja Björk Stefánsdóttir and Anton Karl Ingason. 2018. A high definition study of syntactic lifespan change. *U. Penn Working Papers in Linguistics*, 24(1):1–10.

Lilja Björk Stefánsdóttir and Anton Karl Ingason. 2024a. Reinventing an identity for a more liberal audience. *New Ways of Analyzing Variation*, 52.

Lilja Björk Stefánsdóttir and Anton Karl Ingason. 2024b. Using the Icelandic Gigaword Corpus to explain lifespan change. In *Proceedings of CLARIN Annual Conference 2024*, pages 6–9.

Lilja Björk Stefánsdóttir and Anton Karl Ingason. 2024c. Wiggly lifespan change in a crisis. contrasting reactive and proactive identity construction. *U. Penn Working Papers in Linguistics*, 30(2):119–125.

Lilja Björk Stefánsdóttir, Johanna Mechler, and Anton Karl Ingason. Forthcoming. The stylistic aftermath of the financial crisis in the Icelandic Parliament: Reinventing identity at the level of political parties. *U. Penn Working Papers in Linguistics*.

Lilja Björk Stefánsdóttir and Anton Karl Ingason. 2025. Transcriber effects in the Icelandic parliament corpus. *Digital Humanities in the Nordic and Baltic Countries Publications*, 7(1):106–110.

Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson, and Jón Guðnason. 2018. Risamálheild: A very large Icelandic text corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, LREC 2018, pages 4361–4366, Miyazaki, Japan.

Steinþór Steingrímsson, Örvar Kárason, and Hrafn Loftsson. 2019. Augmenting a BiLSTM tagger with a morphological lexicon and a lexical category identification step. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1161–1168, Varna, Bulgaria.

Höskuldur Thráinsson. 2007. *The Syntax of Icelandic*. Cambridge University Press, Cambridge.