V1 in Icelandic: A Multifactorical Visualization of Historical Data

Miriam Butt, Tina Bögel, Kristina Kotcheva, Christin Schätzle,
Christian Rohrdantz, Dominik Sacha, Nicole Dehé, Daniel A. Keim
University of Konstanz
78457 Konstanz, Germany
firstname.lastname@uni-konstanz.de

Abstract

We present an innovative visualization technique for the analysis of historical data. We illustrate our method with respect to a diachronic case study involving V1 word order in Icelandic. A number of interacting factors have been proposed by linguists as being determinative of matrix declarative V1. The significance of these factors in contributing to declarative V1 can be explored interactively via our multifactorial visualization within a given text, but also comparatively over time. We apply the visualization to a corpus study based on the IcePaHC historical corpus of Icelandic and show that new results emerge very clearly out of the visualization component and that the appearance of declarative V1 is not confined to the situations identified so far by linguists. We demonstrate that the multifactorial visualization opens up new avenues for the exploration of alternative explanations. The visualization can be applied to any linguistic problem studying an interaction between several factors across time.

Keywords: V1 declaratives, Icelandic, IcePaHC, language change, corpus, visualization

1. Introduction

We present an innovative method for the visualization of multifactorial linguistic data over time. We present our visualization with respect to a concrete corpus study of V1 matrix declaratives found in the recently released annotated Icelandic parsed historical corpus (IcePaHC) (Wallenberg et al., 2011; Rögnvaldsson et al., 2012). The visualization allows for an "at-a-glance" appraisal of the occurrence of V1 declaratives and we are able to adduce evidence that advances the existing linguistic understanding of V1 in Icelandic.

Each instance of a V1 declarative is coded as to what kind of subject it occurs with and what kind of verbal element is involved. We focus on these factors as they have been proposed by previous literature as being determinants for the occurrence of V1 in declaratives. Our multifactorial visualization also shows whether the occurrence of a given V1 with a certain type of subject and a certain type of verb is higher or lower than expected.

The visualization brings out several unexpected results. For one, contrary to what is suggested in the previous literature, V1 is not overly associated with subject-less sentences, nor does it occur primarily with certain kinds of lexical verbs. Furthermore, although V1 is generally thought to be relatively stable in the history of Icelandic, the visualization shows a marked decrease as of 1900 and an intermediary dip around 1550 CE. The latter can be attributed to genre, as V1 primarily occurs in narratives and the corpus around 1550 mainly comprises religious texts. The reasons for the former effect (decrease of V1 from about 1900) remains subject for further linguistic research.

In the discussion, we emphasize that the majority of our results could not have been obtained via "standard" corpus linguistic methods and that multifactorial visualizations such as the one presented here are poised to play a key role in furthering our understanding of diachronic data.

2. Theoretical Background

2.1. Word Order and Diachronic Change

Questions of word order change in languages have constituted a major point of inquiry within historical linguistics. For example, like most Indo-European languages, the Germanic languages are generally taken to have been SOV languages in their earlier manisfestations (e.g., Eythórsson (1995); Fleischer and Schallert (2011); Fischer et al. (2000)). Today, English is an SVO language and German is what is known as a V2 language. The SOV, SVO or V2 word order generally refers to the word order in basic declarative matrix clauses. As is also well known, word order can vary within a language according to the type of clause expressed. For example, imperatives in German and English are V1 (verb initial), as are yes-no questions. However, German also contains instances of V1 in matrix declaratives (Axel, 2009; Hinterhölzl and Petrova, 2010). Some of the questions that arise with respect to historical change and word order are: 1) what determines changes in word order over time; 2) what determines differences in word order at a synchronic level? Many accounts to date invoke purely syntactic reasons for word order. One major factor that has been implicated in a change from SOV to SVO is the loss of case marking (e.g., see Kiparsky (1997); Kiparsky (2001) for a theory of diachronic change involving an interacting system of case, agreement and word order), factors such as finiteness are also known to play a role. More recently, a perspective that focuses more on the role of pragmatics and information-structure is being articulated in some detail (e.g., see Petrova (2011); Hinterhölzl and Petrova (2010); Hinterhölzl and Petrova (2011)).

In this paper, we look at data from Icelandic. Icelandic is a Germanic language that is rich in case marking and is a V2 language with fairly fixed word order. Moreover, V1 in matrix declaratives has existed throughout the recorded history of Icelandic and continues to exist today. In this paper, we examine data involving V1 matrix declaratives with the aim of identifying factors that facilitate V1 word

order.

2.2. V1 in Icelandic

Previous studies on V1 in Icelandic (e.g., Sigurðsson (1990); Franco (2008)) have argued that declarative V1 is mainly confined to narrative inversion and is connected to the introduction of known referents. These studies present a syntactic account by which V1 in essence is actually a V2 (verb second) construction, just with a *pro* (empty) subject. Sigurðsson additionally presents a small corpus study comparing stages of Icelandic. He concludes that V1 is a fairly stable phenomenon in Icelandic, with only two changes, one being that Old Icelandic allowed for referential as well as non-referential *pro* while Modern Icelandic only allows non-referential *pro*. In addition, Modern Icelandic now allows for an expletive in initial position, thus rendering some of the old V1 structures effectively into V2 structures. He also notes that V1 is found mostly in narrative texts, a finding that is confirmed by our larger corpus study. However, we find that V1 is not primarily confined to narrative inversion. The purely syntactic account as promoted by Sigurðsson (1990) and Franco (2008) implies that only unaccusative verbs (including passives, etc.) are possible in V1 declaratives. However, our findings show that V1 is not restricted by lexical class. We also found only 448 instances of subjectless V1 sentences as opposed to 4893 sentences with overt subjects. This result casts serious doubt on the hypothesis that V1 is actually underlyingly a V2 construction, just with an empty subject.

Some typical examples of declarative matrix V1 as identified by our corpus study are shown in (1)–(5).

(1) Vil ég heldur rita þeim hinum fám
will I rather write the.DAT those.DAT few.DAT
sinnum er þarf d og s
times.DAT REL required d and s
'I shall rather write *d* and *s* on the few occasions where it [z] is needed.'
(1150, *First Grammatical Treatise*)

(2) gengur síðan á brott.
goes then to away
'Then (he) goes away.'
(1350, *Finnboga saga ramma*)

(3) Verður Ketill forviða fyrir atsókn þeirra
becomes Ketill.NOM astonished for attack their
'Ketill becomes surprised by their attack.'
(1650, *Illuga saga Tagldarbana*)

(4) Gnæfði gaflinn hátt yfir fjöruna...
towered gable.the.NOM high over beach.the.ACC
'The gable towered hight over the beach...'
(1907, Jón Trausti *Leysing*)

(5) Sýndi drottinn mikla miskunn
showed Lord.NOM great.ACC mercy.ACC
vin sínum sankti Georgíum...
friend.DAT his_own.DAT saint.DAT George.DAT
'The Lord showed his great mercy to his friend St. George...'
(1525, *Georgíus Saga*)

(1) shows an example with a modal verb in V1 position and a pronominal subject, (2) a V1 without an overt subject, (3) illustrates a 'become' verb, (4) a V1 with a definite NP subject. We could so far not identify any particular verb classes that V1 is confined to. We did find V1 with the verb types identified in previous literature, i.e., unaccusatives, verbs of motion, verba dicendi, passives, presentationals, existentials and inchoatives. Significantly, however, we also found agentive transitives as in (5), which are not predicted by the syntactic account.

A significant portion of the V1 structures involved modals or the verbs 'do', 'have', 'be' and 'become'. Previous analyses (e.g., Franco (2008)) have suggested that modals are a fringe phenomena with respect to V1. This could not be confirmed via our corpus study. Furthermore, all types of verbs, auxiliaries and modals are used overwhelmingly with overt subjects. These overt subjects could be definites, indefinites or pronouns.

An alternative approach to V1 is the information structural view of declarative V1 put forward by Hinterhölzl and Petrova (2010), Hinterhölzl and Petrova (2011) and Petrova (2011) for German. Hinterhölzl & Petrova argue that V2 in German is derivative of an original V1 structure in Old High German. They develop a theory of information structure by which V1 occurs in particular contexts. For one, V1 is used systematically for presentational clauses and existential constructions. These clauses lack a topic-comment structure and the entire clause is in the scope of the assertion (focus). For another, V1 is used in declaratives with a range of verbs that includes motion verbs, verbs of saying and transformative/inchoative verbs. Hinterhölzl & Petrova argue that the basic function of V1 clauses is to introduce new referents to the discourse and that the various types of V1 clauses can be subsumed under the characteristics of lacking a topic-comment structure and the entire clause being in the scope of assertion, i.e. the entire clause is in focus.

Our data indicate that an information structural approach to V1 in Icelandic is more promising than a purely syntactic one, though the precise factors identified by Hinterhölzl & Petrova for German are not found in Icelandic. We conclude that more in-depth linguistic analysis is necessary and present the possibilities offered by our multifactorial visualization as a promising avenue for further research.

3. Initial Corpus Study

One of the properties of visual analytic techniques is the ability to present large amounts of data compactly while at the same time allowing for an interactive exploratory analysis of individual data items. The IcePaHC consists of 60 texts dating from the 12th to the 21st century. It comprises about 1 million words. The texts come from different genres and are not representative across the centuries. For example, in some centuries texts from narrative Sagas are more numerous than in others.

IcePaHC is annotated according to the syntactic annotation scheme of the Penn Treebank (Marcus et al., 1993). The annotation scheme is also compatible with the Penn historical corpora (Kroch and Taylor, 2000).

The corpus is in principle ideal for an investigation of V1 matrix declaratives as differences between matrix embedded

clauses are clearly coded, as are null and expletive vs. overt subjects. Among overt subjects, one can further distinguish between pronominal subjects vs. indefinite vs. definite NPs.
We conducted an initial query to the corpus via the CorpusSearch tool.[1] This search tool is specialized for dealing with corpora in the Penn-Treebank format. The initial query conducted with the CorpusSearch tool yielded roughly 4400 matrix V1 sentences with overt subjects (from a total of 73014 sentences) across nine centuries.
An advantage of the IcePaHC annotation is that yes-no questions and imperatives could be excluded automatically via the initial query as these are annotated clearly. Questions and imperatives can be distinguished readily from the matrix V1 declaratives that were the object of our study. However, an examination of the results showed that the results obtained by the CorpusSearch query were not entirely "clean", but contained some false hits of V1 questions and imperatives. These false hits were identified and removed automatically via a script, leaving 3964 matrix declarative V1 sentences.[2]
A further very valuable feature of IcePaHC is that verbs are coded for verb types: BE 'be', HV 'have', DO 'do', RD 'become', MD modals, VB main verbs. Since previous literature had indicated that verb class played a role, we included the type of the verb involved in V1 as one of the factors to be investigated.
A disadvantage we discovered as part of launching the initial query via CorpusSearch lies in the annotation of V1 declaratives that do not contain an overt subject (cf. (2)). IcePaHC conveniently does provide an annotation for empty subjects (NP-SBJ *pro*) so that one can easily identify subjectless sentences.[3] However, IcePaHC places these empty subjects **before** the main verb, thus rendering surface V1 structures as V2 structures in the annotated form. This design decision is consonant with the syntactic approach to V1 outlined above (seeing V1 as being underlyingly V2, just without an overt subject); however, it meant that subjectless V1 constructions are not recognized as such by a query targeting V1 structures.
We thus launched a more sophisticated query and identified a total of 5341 V1 declaratives out of an overall set of 73014 sentences. Table 1 shows the distribution of V1 according to the type of verb involved. The division into time periods has been done according to Haugen (1984).
The results of the corpus query show that declarative V1 occurs with all kinds of verbs and most frequently with main verbs. Thus, rather than being a fringe phenomenon, the occurrence with modals is a regular part of the language (e.g., contra Franco (2008)).
Recall that referentiality of the subject and the presence or absence of an overt subject have also been implicated in the appearance of declarative V1 in previous literature. It would therefore be instructive to know how each of the verbal categories behave with respect to referentiality and overt presence or absence of the subject NP. For each verb type, we therefore calculated the number of times it occurs with the following: a pronominal subject, a definite NP, an indefinite NP, a pro-dropped (overtly absent) subject and a null expletive subject.
We then checked whether the observed correlation between each verb type and each subject type was statistically significant via the χ^2-test, which measures an observed distribution against an expected distribution. The distribution of the five subject types within each time period was taken as the 'expected' factor, while the distribution for each possible verb type in relation to the three subject constructions was taken as the 'observed' factor. That is, for each text, we calculated the overall occurrence of type of subject with type of verb and then checked whether the numbers found for just the V1 declaratives deviated from this expected pattern. We were able to determine several highly significant correlations via this test.
However, some of the highly significant results are based on a very small data set. This is illustrated with respect to the data for the time period from 1900 to the present, shown in Table 2. Here, the patterns found with respect to 'have', 'be', 'do' and 'become' all come out as being highly significant. However, a look at Table 1 shows that the data being calculated with are very small: 10 instances for 'have', 2 for 'do', and 4 for 'become'.
We did find interesting indications from the χ^2-test for some other correlations. However, the overall picture is not clear and the χ^2-test does not identify particular correlations between verb type and subject type — it only points to an overall pattern that deviates from the overall pattern found in the larger corpus. As a next step, we therefore turned to the question of whether methods from visual analytics could help with identifying determining factors for the appearance of V1 throughout the history of Icelandic.

4. Visual Analytics

4.1. Motivation

Statistical analyses provide just a very rough picture of the data and are suited to either confirm or reject previously expected or at least anticipated knowledge, i.e. they are limited to the investigation of more or less evident hypotheses on the data characteristics.
A visual analysis of a given data set on the other hand allows one to push beyond what is possible with traditional corpus analysis. Instead of merely (dis)confirming the expected, it can lead to unexpected insights by granting an interactive explorative access to the data. The structured statistical analysis process becomes open-ended, in that visualizations of the data may enable researchers to spot unexpected patterns or outliers that they had not been aware of before and that might be of high relevance to them. In particular, visualizations may reveal feature patterns that disappear when using conventional statistics, especially when the statistics require parameters to be fixed, e.g., fixed time epochs that are the basis for some calculations.
In order to support such a visual analytics process in a beneficial manner, carefully designed visual mappings are required (cf. Mayer et al. (2010), Rohrdantz et al. (2012)). Usually, the very first design idea will not be maintained, but

[1] http://corpussearch.sourceforge.net/CS.html

[2] The results also revealed some errors in the annotation in that the corpus provides the infinitive (citation) form of verbs for each finite verb encountered. However, some of these infinitive forms were incorrect. We manually corrected these.

[3] In fact, it distinguishes between two different kinds of empty subjects: expletives and pro-dropped subjects.

Time	have	do	be	become	main verb	modal	Total
upto 1350	67	23	336	50	1045	127	1648
upto 1550	47	10	185	42	944	81	1309
upto 1750	58	6	225	47	569	64	969
upto 1900	64	6	217	62	818	75	1242
upto present	10	2	23	4	121	13	173
Total	246	47	986	205	3497	360	5341

Table 1: Instances of V1 according to type of verbal element

		Type of Subject in %					
Time	Verb Type	Pronoun	Indef. NP	Def. NP	*pro*	expletive	χ^2
1900–present	have	90	10	0	0	0	88.20***
1900–present	be	43	22	9	9	17	61.65***
1900–present	do	100	0	0	0	0	118.39***
1900–present	modal	54	8	0	31	8	7.63
1900–present	become	50	25	0	25	0	42.65***
1900–present	main verb	41	6	5	44	4	4.52
expected	all types	46	9	5	35	6	

Table 2: Significant Correlations according to the χ^2-test. Highly significant correlations are indicated with "***".

it enables insights that feed back into the design process. Our initial visualization showed each V1 sentence in the data individually, but then we decided to aggregate the sentence characteristics to the document level for visualization. Similarly, further details were changed iteratively during the design process. This included the integration of derivations of the original data: sometimes it is not the original numbers contained in the data that provide the best insights, but further calculations based on the initial raw numbers. We rejected the idea of visually representing the absolute number of appearances of certain phenomena in a text because the text lengths are quite heterogeneous and a proper comparison would have been difficult. Instead, the visual representation was designed to show whether such a number of appearances was higher or lower than expected.

4.2. Design and Development Process

Even though the visualization does not depend on previous assumptions, it requires some crucial initial design decisions which subsequently frame the kinds of insights that will be enabled. The challenge is to come up with solutions to plot the concrete data for generic explorations of historical data. Of course, the design can be improved iteratively, but it is quite helpful to reason first about which parts of the data might in principle be relevant and interesting to convey visually. Visualization designers and subject matter experts need to work tightly together and try to understand as much about the background and perspective of their counterparts as possible. In this project we pursued such a strategy. As a result of this, we suggest a novel visualization in order to provide both an overview on the described language corpus as well as insights at different levels of detail.

From an information visualization perspective the data consists of several texts, each of them including meta data and a set of sentences. Additional meta data about the texts, include, for example, genre, year and number of sentences. For each individual sentence, linguistic features can be derived as well. As a next step these features can be mapped to so called visual variables. We use a glyph for the visualization of a text within the corpus. Glyph representations of documents are known from the Information Retrieval field, e.g. the TileBars technique, which provides a compact and informative iconic representation of the documents' contents with respect to certain query terms (Hearst, 1995).

Our glyph contains several visual variables, namely *position*, *shape* and *color* in order to encode sentence features aggregated on the document level. Each glyph holds some shapes that provide information about the occurrence of (V1-)sentence features for the corresponding text. We therefore transform all elements (texts/documents, sentences and related features) into our internal data model and offer aggregated features on different concept levels (e.g., sentence features are aggregated to text features based on the number of sentences). We chose the glyph representation because the number of features/dimensions are suitable for a glyph visualization and in addition we are able to integrate shapes into those glyphs. That is why we designed a novel glyph instead of using standard histograms, line/bar/pie-charts or glyphs (e.g., star glyphs). The glyphs are arranged in space for exploration. We encode the text age and the genre to x- and y-position as they are the most interesting variables. In general, our intent is to use redundant encodings for relevant features.

Our visualization offers several interaction techniques in order to drill down in to the data if desired. First, we added zooming and panning interactions in order to navigate within the visualization's viewport. Our visualization also enables analysts to get details on demand through tooltip operations.

Our system is built using the *Java Piccolo 2D*[4] framework as it matches our requirements and has basic visualizations (e.g., basic shapes) and interactions built in. Language processing scripts that are the basis for feature extraction are integrated into the visualization system.

4.3. Our Multifactorial Visualization

Figures 3 and 4 show the overall visualization we arrived at. The texts contained in the IcePaHC are arranged from oldest to newest with the oldest represented at the very top, the newest at the very bottom. Each text is visualized as one composed glyph. A horizontal line extends from the middle of each glyph towards the right. This horizontal line is bisected by a vertical line. The horizontal line indicates the time span covered by the corpus. The vertical line shows where in the time span the text represented by the glyph is to be situated, see Figure 2 for an example.

Each glyph consists of three main parts. A horizontal bar on the top represents the length of the text in comparison to the longest text in the corpus. The bar of the longest text covers the whole width of the glyph. At the positions of the sentences containing V1, vertical light gray stripes are drawn in this horizontal bar. The user can thus determine the position of the V1 occurrence in the text and can determine where in the narrative flow of the text the V1 declarative is situated.

The main part of the glyph is a matrix containing colored items. The matrix represents the interaction of verb types with subject types. The columns encode verb types while the rows encode the subject types. The different verb and subject types can be identified by their position alone. However, for better visibility, verb types have redundantly been encoded by color and subject types have been encoded redundantly by shape. If a certain matrix cell is empty, the corresponding combination did not occur in the given text.

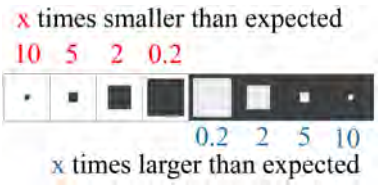

Figure 1: Different visual encodings depending on whether an observed frequency was higher or lower than expected. If the observed frequency is lower than expected, the outer part is white and the inner part is dyed in the color of the verb. If the observed frequency is higher than expected, the inner part is light gray and the outer part is dyed in the color of the verb. The ratio along one axis between both rectangle parts corresponds to the ratio between the observed and expected frequency.

The colored cell furthermore shows whether the combination of subject and verb type depicted by the cell occurred more or less frequently within the given text than expected. The deviation from the expected value is quantified via the scale shown in Figure 1.

[4]http://www.piccolo2d.org/

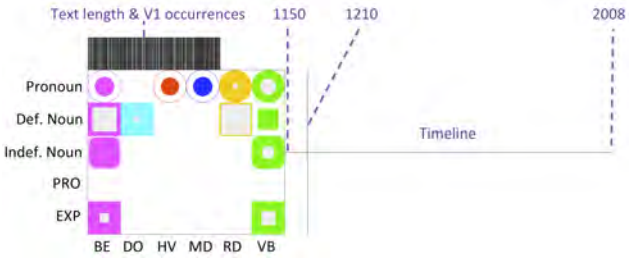

Figure 2: Glyph representation for a text consisting of three main parts: (1) A horizontal bar representing the text length and V1-occurrences, (2) A matrix containing colored shapes that belong to the occurrence of a particular verb and (3) a timeline on the right.

The visualization opens up new avenues of research for linguists. For one, it allows for a single, comprehensive view of the entire 60 texts of the corpus, as shown in Figures 3 and 4. Many details and possible influencing factors are displayed. For another, the visualization is interactive. After having identified potentially interesting patterns within the overall picture, the user can zoom into particular glyphs and examine the distribution of V1 declaratives within a particular text. The horizontal bar shows how long a text is and where in the text the particular V1 declarative is situated, allowing for an immediate assessment of the significance of the data. Furthermore, the visualization incorporates meta data on genre type. These can be accessed via mouse-over for each glyph. In Figures 3 and 4 all the Sagas have been arranged to appear on the very left of the visualization. The clump of texts in the middle (they appear around 1550) that are further to the right represent mainly religious texts (in the Figures, they appear at the bottom of Figure 3 and the top of Figure 4).

5. Insights and Discussion

The visualization shown in Figures 3 and 4 represents a powerful new tool for inquiries into diachronic change. Our findings to date show that proposals put forward to explain the occurrence of V1 declaratives in Icelandic so far do not match up with the actual patterns observed via the IcePaHC diachronic corpus, though we can confirm some of the findings of the previous literature.

Recall that Sigurðsson (1990) concluded that V1 is a relatively stable phenomenon in Icelandic, with the caveat that Modern Icelandic now allows for an expletive in initial position, thus rendering some of the old V1 structures effectively into V2 structures. Overall, this conclusion appears to be valid, as is his note that V1 is found mostly in narrative texts. As Figures 3 and 4 show, V1 is predominantly found in the Sagas (entirely narrative in nature). Our data also show that V1 shows a marked decrease as of 1900. This result is not entirely in keeping with Sigurðsson's observation. The innovation of an overt expletive by itself should not make such a huge difference as expletives overall did not constitute a large part of the V1 declarative occurrences (total of 239 of 5341 instances; 10 expletive V1 instances were found in the modern Icelandic texts).

Overall we note that a syntactic account which seeks to

understand declarative V1 primarily in terms of non-overt subjects (i.e., as underlying V2s) is not supported by our data. As shown in Figures 3 and 4, empty subjects (*pro*) and non-overt expletives are not generally found more frequently than expected. In absolute numbers, they make up a total of 448 of 5341 V1 declarative instances found in IcePaHC.

Figures 3 and 4 also show very clearly that all verb types occurred with V1, especially modals and lexical (main) verbs. V1 does not seem to have been restricted to only a subclass of verbal types. In particular, our data show that modals are not fringe phenomena with respect to V1, as suggested by the existing literature, but constitute a central part of the phenomenon.

Finally, the visualization of genre type in Figures 3 and 4 makes a set of data stand out. Although Icelandic is generally held to be stable with respect to V1 declaratives up until modern times, the texts used to represent the Icelandic of around 1550 show a comparative absence of V1. Since the texts are all not Sagas, but tend to be drawn from religious literature, we suspect that IcePaHC is characterized by a genre effect that was not originally intended by the developers of IcePaHC. We would assume that the Icelandic of around 1550 as according to IcePaHC will exhibit differences to the other parts of the corpus in other areas as well — this is a factor that could seriously impede diachronic studies. However, it is also a factor that is brought to the forefront of one's attention immediately by the visualization and can thus be factored into any analysis that is developed.

An issue that we have not addressed is whether the Icelandic data provide support for the alternative information structural approach to V1 articulated by Hinterhölzl & Petrova (2010, 2011) and Petrova (2011) for German.

The characteristics of our data set match their results in that the verb types involved in V1 are not confined to particular classes of verbs. However, V1 seems to appear with discourse old referents (pronouns, definite NPs) to a large degree, so if there is an information structural effect, it is not immediately obvious that it is the same one that was identified for German.

6. Conclusion

We present a powerful new visualization tool for the study of diachronic change. We have developed this tool with respect to a case study of V1 declaratives in Icelandic, however, the tool should be applicable to any diachronic study that seeks to understand a multifactorial interaction of the type exemplified by Icelandic declarative V1.

7. Acknowledgements

This work was partially funded by the German Research Foundation (DFG) under grant BU 1806/7-1 "Visual Analysis of Language Change and Use Patterns" and the Research Initiative *LingVisAnn* at the University of Konstanz, Germany.

8. References

Axel, K. (2009). The verb-second property in Old High German: Different ways of filling the prefield. In Hinterhölzl, R. and Petrova, S., editors, *Information Structure and Language Change: New Approaches to Word Order Variation in Germanic*, pages 17 – 43. Berlin: de Gruyter.

Eythórsson, T. (1995). *Verbal syntax in the Early Germanic languages*. Ph.D. thesis, Cornell University.

Fischer, O., van Kemenade, A., Koopman, W., and van der Wurff, W. (2000). *The Syntax of Early English*. Cambridge University Press, Cambridge.

Fleischer, J. and Schallert, O. (2011). *Historische Syntax des Deutschen: Eine Einführung*. Narr, Tübingen.

Franco, I. (2008). V1, V2 and criterial movement in Icelandic. *Studies in Linguistics*, 2:141 – 164.

Haugen, E. (1984). *Die skandinavischen Sprachen*. Hamburg: Buske.

Hearst, M. A. (1995). Tilebars: Visualization of term distribution information in full text information access. In *CHI*, pages 59–66.

Hinterhölzl, R. and Petrova, S. (2010). From V1 to V2 in West Germanic. *Lingua*, 120.2:315–328.

Hinterhölzl, R. and Petrova, S. (2011). Rhetorical Relations and Verb Placement in Old High German tatian. In *Salience. Multidisciplinary perspectives on its function in discourse*, pages 173–201. Berlin: de Gruyter.

Kiparsky, P. (1997). The rise of positional licensing. In van Kemenade, A. and Vincent, N., editors, *Parameters of Morphosyntactic Change*, pages 460–494. Cambridge University Press, Cambridge.

Kiparsky, P. (2001). Structural case in Finnish. *Lingua*, 111:315–376.

Kroch, A. and Taylor, A. (2000). Penn-Helsinki parsed corpus of Middle English. second edition.

Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993). Building a large annotated corpus of english: the penn treebank. *Computational Linguistics*, 19(2):313–330.

Mayer, T., Rohrdantz, C., Butt, M., Plank, F., and Keim, D. A. (2010). Visualizing Vowel Harmony. *Linguistic Issues in Language Technology*, 4(Issue 2):1–33, December.

Petrova, S. (2011). Modeling word order variation in discourse: On the pragmatic properties of VS order in Old High German. *Oslo Studies in Language*, 3(3):209–228.

Rögnvaldsson, E., Ingason, A. K., Sigurðsson, E. F., and Wallenberg, J. (2012). The Icelandic Parsed Historical Corpus (IcePaHC). In *Proceedings of LREC 2012*, pages 1978–1984.

Rohrdantz, C., Hund, M., Mayer, T., Wälchli, B., and Keim, D. A. (2012). The world's languages explorer: Visual analysis of language features in genealogical and areal contexts. *Comput. Graph. Forum*, 31(3):935–944.

Wallenberg, J. C., Ingason, A. K., Sigurðsson, E. F., and Rögnvaldsson, E. (2011). *Icelandic Parced Historical Corpus (IcePaHC)*.

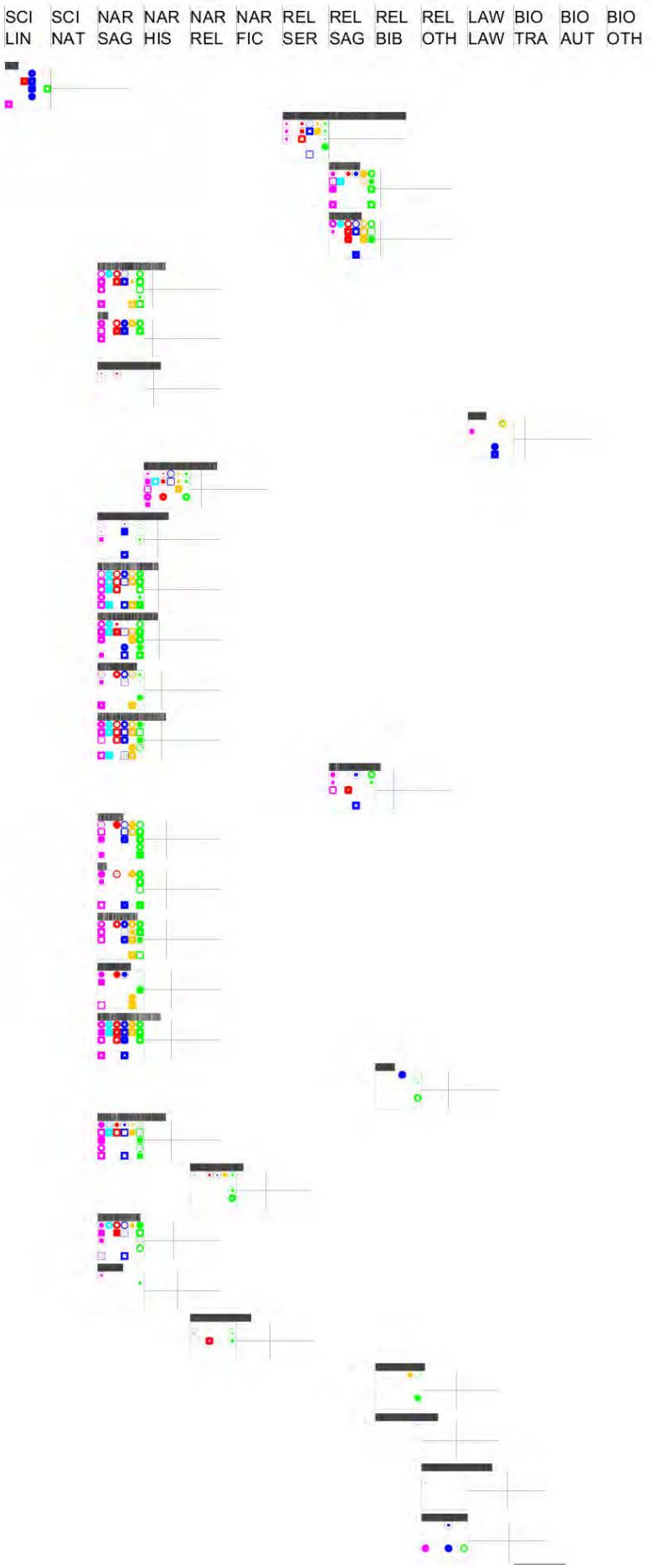

Figure 3: Top half of Visualization of V1 Occurrences in IcePaHC

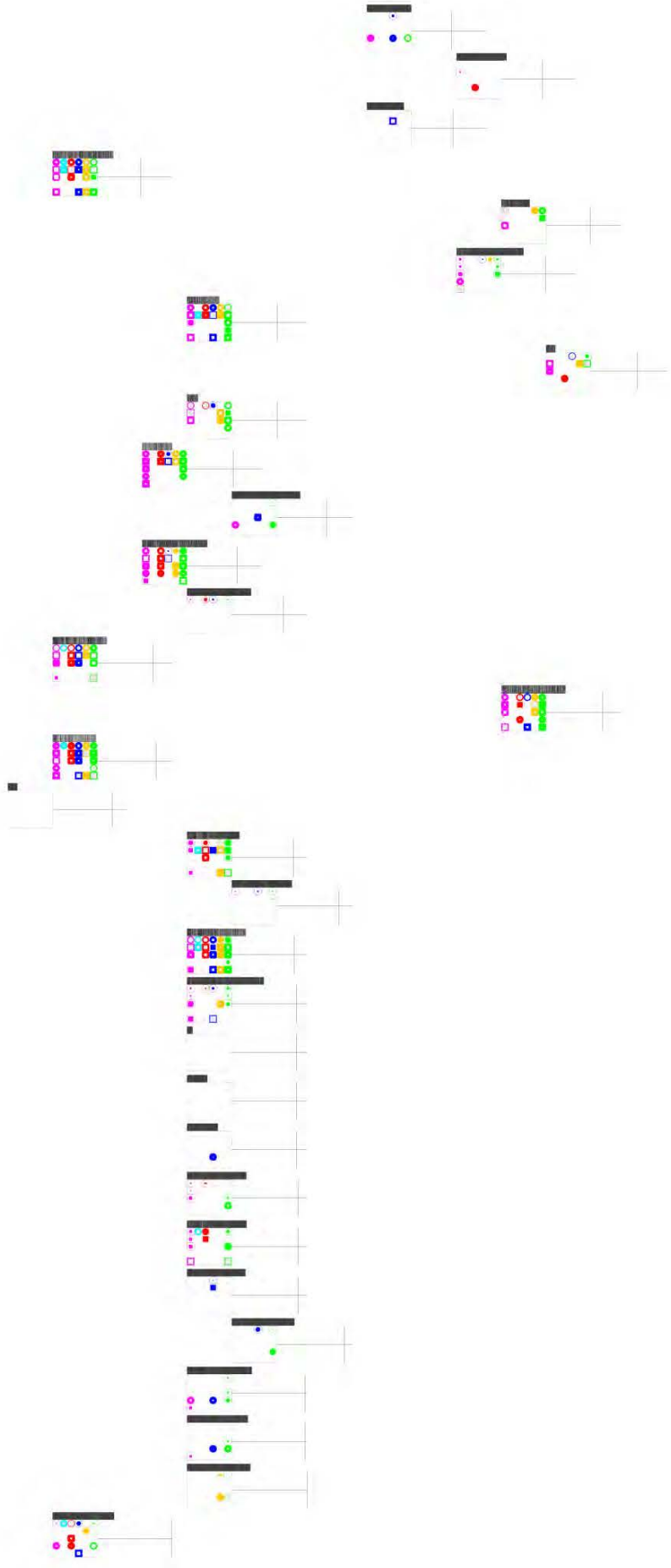

Figure 4: Bottom half of Visualization of V1 Occurrences in IcePaHC