A diachronic perspective on V1 in Icelandic

Kristina Kotcheva, Tina Bögel, Christin Schätzle, Miriam Butt and
Nicole Dehé

University of Konstanz

`kristina.kotcheva@uni-konstanz.de`

1 Introduction and Background

V1 in declarative clauses is a well-known phenomenon in Germanic. We present a diachronic corpus study of V1 declarative clauses in Icelandic. Our results support Sigurðsson's (1990) finding that V1 in Icelandic is a relatively stable phenomenon. However, the corpus study also yielded results that are surprising in the light of the literature on Icelandic V1: 1) the verb types involved in V1 are not confined to particular classes of verbs (e.g., motion verbs, unaccusatives), but encompass a large range of classes; 2) a significant portion of the V1 structures involved modals or the verbs 'do', 'have', 'be' and 'become'; 3) all types of verbs, auxiliaries and modals are used overwhelmingly with overt subjects (definites, indefinites and pronouns). We suggest that our results instead largely confirm the information structural view of declarative V1 originally formulated by Önnerfors (1997) for German and put forward by Hinterhölzl & Petrova (2010, 2011) and Petrova (2011) for Old High German.

Hinterhölzl & Petrova (2010, 2011) and Petrova (2011) have argued 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 (focus).

2 Diachronic Corpus Study – V1 in Icelandic

2.1 Data

Our data is taken from the Icelandic parsed historical corpus (IcePaHC) (Wallenberg et al. 2011, cf. detailed description in Rögnvaldsson et al. 2012), a treebank consisting of 60 texts dating from the 12th to the 21st century and comprising

ca. 1 million words. The treebank is annotated according to the syntactic annotation scheme of the Penn Treebank (Santorini 2010). Given the syntactic annotation, the corpus lends itself to exploring V1 data, as differences between matrix and embedded clauses are coded, as are null and expletive vs. overt subjects. 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. After excluding V1 questions and imperatives, 3964 matrix declarative V1 sentences remained. These constitute the subject of our study. Typical examples of declarative V1 are given in (1) and (2).

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

> (2) 'become' in V1 (1650, *Illuga saga Tagldarbana*)
> *Varð hann innan stundar sjö manna bani.*
> became he within time six men's death
> 'He became the death (=killed) to six men before long.'

2.2 Previous Findings

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. They present a syntactic account by which V1 in essence is actually a V2 construction, just with a pro 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 Iceland 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.

2.3 Our Findings

However, we find that V1 is not primarily confined to narrative inversion, nor is it conditioned by the referentiality/definiteness of the subject and it certainly is not restricted by lexical class. Our findings go against a purely syntactic account as promoted by Sigurðsson 1990, Axel 2005 and Franco 2008, which implies that only unaccusative verbs (including passives, etc.) are possible in V1 declaratives. We also found only 307 instances of subjectless V1 sentences as opposed to 3964 sentences with overt subjects. This result casts serious doubt on the V1 as V2+pro subject hypothesis.

Instead, our findings are broadly consonant with Hinterhölzl & Petrova's information structural account in that declarative V1 clauses lack a topic-comment structure and

present the entire clause as being in the scope of assertion (all focus). Our data suggest that V1 declaratives allow any verb type to appear in V1 position. They include the verb types identified in previous literature, i.e., unaccusatives, verbs of motion, verba dicendi, passives, presentationals, existentials and inchoatives. Significantly, however, we also find agentive transitives, which are not predicted by an account which focuses on overt subjects only being licensed by structures which contain an underlying object.

The variables involved in understanding the interrelations between the data as suggested in the literature, namely, the occurrence of V1 with certain lexical classes, the presence or absence of an overt subject and the type of the overt subject (definite, indefinite, pronominal) constitute a complex set of interacting factors which we explore with the visual analytic tools provided in work such as Rohrdantz et al. (2011, 2012). For our corpus study, indefinite, definite and pronominal subject constructions were extracted relative to time periods commonly applied in literature on Icelandic (e.g. Haugen 1984; see Table 1). For each of these subject constructions, the frequency of each possible verb tag was determined.

The data was then subjected to a statistical analysis based on the Chi-square method, which measures an observed distribution against an expected distribution. The distribution of the three subject constructions 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. The corpus study yielded several surprising patterns. For one, the discussion in the literature did not lead us to expect a high preponderance of modals (as well as 'do', 'have', 'be', and 'become'). V1 declaratives with modals have been treated as a fringe phenomenon in the literature (or classified as imperatives, cf. Franco 2008). However, they constitute a robust part of V1 declaratives throughout the history of Icelandic. Furthermore, 'become' in the time period before 1350 shows a very significant result with respect to its distribution in that the occurrence of 'become' in conjunction with definite subjects is very high. However, this significance disappears if only the V1 subcorpus is considered. In Figure 1, this relationship is demonstrated. The lines show the expected vs. the observed distribution of the RD Verbtag (="to become") per subject construction over time, with 1 being the earliest time period and 4 the current one.

age*	HV	DO	BE	RD	VB	MD	SentNum
beginning-1350	67	20	291	46	617	133	1174
1350-1550	46	10	160	39	544	82	881
1550-1900	117	12	408	102	1056	143	1838

Table 1

*Time intervals chosen accroding to Haugen (1984); HV: *hafa* 'have'; DO: gera 'do'; BE: *vera* 'be'; RD *verður* 'become'; VB: main V; MD: modal V; PRON: pronoun; DEF: definite noun phrase; INDEF: Indefinite noun phrase; SentNum: total number of sentences.
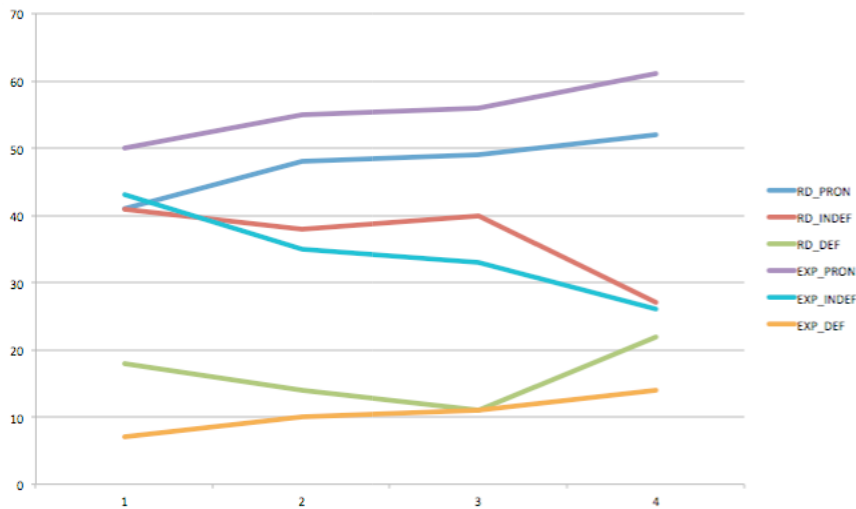
This means that while 'become' occurs more often than expected with definite subjects in the overall language, the V1 declaratives differ from the overall pattern in the language in that they do not favor definite subjects. This finding is in line with Hinterhölzl & Petrova's idea that V1 is associated with the presentation of new material, rather than discourse old referents.

3 Conclusion

In conclusion, while our diachronic corpus study of matrix V1 declarative clauses in Icelandic support Sigurðsson's (1990) finding that V1 in Icelandic is a relatively stable phenomenon, we were able to uncover significant patterns in the data that speak against a purely syntactic account in which V1 is really underlyingly an instance of V2, just with a non- overt subject. The features yielded up by our data set are instead more in line with the 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 and in that V1 is not connected with discourse old referents, but is rather tied to constructions without a topic comment structure. Additionally, 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.

References

Axel, K. (2005) *Studies on Old High German Syntax*. Left sentence periphery, verb placement and verb-second. Benjamins, Amsterdam.

Eiríkur R. et al (2012) The Icelandic Parsed Historical Corpus_ (IcePaHC). *LREC* 2012, pp.1977–1984.

Franco, I. (2008) V1, V2 and criterial movement in Icelandic. Studies in Linguistics (STiL), *CISCL Working Papers*.

Halldór Á. Sigurðsson (1990) V1 declaratives and verb raising in Icelandic, *Syntax and Semantics* 24, 41–69.

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

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

Hinterhölzl, R. & S. Petrova (2010) From V1 to V2 in Older Germanic. *Lingua,* 120(2): 315–328.

Önnerfors, O. (1997) *Verb-erst-Deklarativsätze. Grammatik und Pragmatik.* Almqvist & Wiksell, Stockholm.

Petrova, S. (2011) Modeling word order variation in discourse: On the pragmatic properties of VS order in Old High German. In E. Welo ed. *Proceedings from the Workshop on Indo-European Syntax and Pragmatics*, Athens, GA, May 13–15, 2009, Oslo: Studies in Linguistics.

Rohrdantz, C., A. Niekler, A. Hautli, M. Butt & D.A. Keim. (2012) Lexical Semantics and Distribution of Suffixes – A Visual Analysis. *EACL 2012 Joint Workshop of LINGVIS*, Avignon.

Rohrdantz, C., A. Hautli, Th. Mayer, M. Butt, D.A. Keim and F. Plank. (2011) Towards Tracking Semantic Change via Visual Analytics. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (ACL-HLT). Portland, Oregon.

Santorini, B. (2010). *Annotation Manual for the Penn historical corpora and the PCEEC* (http://www.ling.upenn.edu/hist-corpora/annotation).

Wallenberg et al (2011) *Icelandic Parced Historical Corpus* (IcePaHC), (http://www.linguist.is/icelandic_treebank).