

What can synchronic gradience tell us about reanalysis?
Verb-first conditionals in written German and Swedish

Martin Hilpert
Freiburg Institute for Advanced Studies

Abstract

This paper discusses verb-first conditionals (*Had I known this, I would have stayed at home*) with regard to the hypothesis that the construction developed from a dialogual sequence into a hypotactic structure. While plausible, independent evidence for this scenario has been scarce. An alternative account posits that questions could have been analogized with existing conditional clause types, making it possible for speakers to use questions as conditional protases. The present study assesses these two hypotheses on the basis of present-day corpus data from German and Swedish. Crucial to this approach is the notion that a reanalyzed structure retains aspects of its original source. Put simply, if the protasis of a verb-first conditional developed out of a question, it should retain some question-like characteristics, even after reanalysis. Further, if verb-first conditionals have become grammaticalized more strongly in one language than in another, the less-grammaticalized construction should display more question-like characteristics.

1 Introduction

This paper addresses the third question posed by Traugott and Trousdale in the introduction to this volume.¹ In its most general form, the question asks how we are to understand the relation between synchronic gradience and the process of reanalysis in grammaticalization. The focus of the present discussion will be on a more specific version of that question: Can synchronic gradience, i.e. co-temporal variation among the tokens of a linguistic category (*subsective gradience* in the terminology of Aarts 2007), tell us anything about how hearers reanalyzed a token of one category as a token of a different category?

A brief comment is in order to explain why this question is useful to ask. First and foremost, the question touches on the issue of how synchronic data can be fruitfully investigated in studies of grammaticalization (cf. Traugott, forthcoming). Written records of past language use are often sparse, do not reach far back enough, or are altogether unavailable. Given such a situation, resorting to synchronic data is a common strategy. Heine and Kuteva (2002), for instance, juxtapose many grammaticalized forms with their putative lexical sources, in which correspondences between lexical and grammatical elements are inferred simply on the basis of qualitative similarities and differences in synchronic data.

The present paper argues that synchronic data is useful not only for qualitative comparisons between different forms, but crucially also because of the quantitative information that it holds. Large modern corpora can reveal gradient, quantitative differences between grammatical forms and their lexical counterparts that would elude comparisons of any two isolated examples. To illustrate, Hilpert and Koops (2008) use corpus data to differentiate between a lexical and a grammaticalized variant of the Swedish posture verb *sitta* 'sit'. In the latter, the posture verb is co-ordinated with a second verb and indicates durativity. The two variants differ, amongst other things, with regard to their typical argument structure.

¹ Acknowledgment note: I would like to thank David Denison, Jan Lindström, Amanda Patten, Elizabeth Traugott and Graeme Trousdale for their comments on an earlier version of this paper. Many thanks are also due to Daan Van den Nest, who discussed the topic of this paper with me. Henrik Rosenqvist supplied example (18a). Finally, the audience of NRG-4 (Leuven) provided helpful input. The usual disclaimers apply.

While lexical *sitta* is often elaborated spatially (1a), such elaboration is typically absent with co-ordinated *sitta* (1b). Nevertheless, lexical *sitta* can occur without spatial elaboration (1c), and co-ordinated *sitta* may occur with it (1d) – examples of this kind are perfectly grammatical.

- (1) a. *Jag har suttit vid skrivbordet nästan hela dagen.*
 I have sat at desk.the almost all day
 ‘I’ve sat at the desk almost all day.’
- b. *Jag har suttit och läst hela dagen.*
 I have sat and read all day
 ‘I have been reading all day.’
- c. *Han satt en stund igen.*
 he sat a while again
 ‘He sat (somewhere) for a while again.’
- d. *Jag har suttit vid skrivbordet och läst hela dagen.*
 I have sat at desk.the and read all day
 ‘I’ve sat at the desk and read all day.’

The difference in argument structure between the two variants is thus gradient, not categorical. Hilpert and Koops make the case that grammaticalization can show itself in purely quantitative differences between two forms with a common source (2008: 254).

But what then is the link between gradience, as observed through quantitative corpus data, and reanalysis? In the case of Swedish *sitta*, there is no doubt that the posture verb represents the lexical source that was reanalyzed as an aspectual marker in the context of a co-ordinated structure. A co-ordinated structure with two verb phrases (2a) came to be re-bracketed as a single verb phrase in which the verb *sitta* acquired the status of a grammatical element (2b). Present-day evidence for this reanalysis is that direct objects of the second, lexical verb can undergo extraction (2c), which is not possible in regularly co-ordinated structures (2d).

- (2) a. *Jag [satt] och [läste].*
 I sat and read
 ‘I sat and read.’
- b. *Jag [satt och läste]*
 I sat and read
 ‘I was reading.’
- c. *boken_j jag [satt och läste \emptyset_j]*
 book I sat and read
 ‘the book I was reading’
- d. ** boken_j jag [skrattade och läste \emptyset_j]*
 book I laughed and read
 intended: ‘the book I was reading while laughing’

Hilpert and Koops claim that the reanalysis of (2a) as (2b) was a gradual process rather than a sudden, catastrophic one. Diachronic corpus data indicate that examples such as (2c) only gradually increase in frequency over time (2008: 257).

The main subject of the present study is a case of grammaticalization in which the process of reanalysis is much less clear-cut than with Swedish *sitta* – namely the case of verb-first conditional clauses, a construction that occurs across several Germanic languages. The focus here is on German and Swedish verb-first conditionals, as illustrated in (3a) and (3b).

- (3) a. *Ändern sie das Testament, dann tritt eine neue Rechtslage ein.*
change they the testament then sets a new legal.situation in
'If they change the testament, a new legal situation obtains.'
- b. *Räkna man även barnen, blir siffran avsevärt högre.*
counts one even children becomes number much higher
'If also children are counted, that number increases substantially.'

As is explained in more detail below, there are mutually conflicting accounts of how pre-existing grammatical structures came to be reanalyzed, thus giving rise to the modern construction. The present study makes a methodological contribution, exploring how gradience in synchronic corpus data sheds light on this issue. While synchronic gradience cannot generate a full-fledged account of how a diachronic change happened, the value of synchronic gradience lies in the fact that it can be used to test the predictions of diachronic accounts with respect to synchrony. If different grammaticalization scenarios make conflicting predictions about synchronic states of affairs, large present-day corpora can be used to evaluate the relative plausibility of these accounts. To make matters more concrete, this paper compares the predictions of two grammaticalization accounts of verb-first conditionals against evidence from modern German and Swedish corpora, arguing that synchronic gradience does indeed tell us something about reanalysis.

The remainder of this paper is organized as follows. Section 2 presents the verb-first conditional construction and discusses previous work. The two conflicting grammaticalization accounts of the construction are referred to as *the dialogical account* on the one hand and *the analogical account* on the other. The section spells out the predictions that both accounts make for present-day usage of the construction. Section 3 presents the corpora that are used, explains how data is retrieved, and summarizes the results from the quantitative investigations; section 4 concludes the paper with an argument that the existing evidence provides evidence for the dialogical account.

2 Verb-first conditionals in German and Swedish

Cross-linguistically, conditional constructions that bear strong resemblances to polar interrogative constructions are fairly common (Traugott 1985). Van den Nest (forthcoming) offers a comprehensive list of studies that address the phenomenon across genetically unrelated languages. The focus in this paper is on German and Swedish, which both have a conditional clause construction in which the protasis is formally identical to a polar question, as shown in (4) and (5). For the sake of comparison, the examples in (6) illustrate that a similar construction exists also in English, where it is restricted to the verbs *had*, *should*, and *were*. While it is possible to construct polar questions in correspondence to (6a) and (6b), a question such as *Should you need anything?* appears much less natural than the German and Swedish questions in (4b) and (5b).

- (4) a. *Ändern sie das Testament, dann tritt eine neue Rechtslage ein.*
change they the testament then sets a new legal situation in
'If they change the testament, a new legal situation obtains.'
- b. *Ändern sie das Testament?*
change they the testament
'Do they change the testament?'
- (5) a. *Räknar man även barnen, blir siffran avsevärt högre.*
counts one even children.DEF becomes number.DEF much higher
'If also children are counted, that number increases substantially.'
- b. *Räknar man även barnen?*
counts one even children.DEF
'Are also children counted?'
- (6) a. *Had he known this, he would have cancelled the trip.*
b. *Should you need anything, please call the front desk.*
c. *Were he to fail, he would be ashamed to show his face at home.*

From a functional perspective, the connection between questions and conditionals can receive a straight-forward explanation, as these constructions share the semantic traits of non-factuality or possibility. Given the synchronic evidence of the examples in (4) and (5), buttressed by the typological finding that questions and conditionals resemble each other more often than not, it is hard to resist the idea that historically, polar questions were reanalyzed as the protases of verb-first conditionals. And indeed, this hypothesis has been endorsed in several classic references (Erdmann 1886, Paul 1920, Jespersen 1940, amongst others, cf. the discussion in Van den Nest forthcoming).

Fleshed out, a dialogual scenario that posits the reanalysis of a polar question as the protasis of a conditional clause would have the three stages outlined below (Van den Nest forthcoming). The first stage is dyadic in nature. A speaker asks a polar question (*Is the child younger than three years?*), receives a positive answer, and proceeds with a declarative statement that is contingent on that positive answer (*Then it may ride for free*). As the pattern shown in (7a) routinizes over time, the presence of the positive answer becomes less important. One can imagine a second stage, in which it is sufficient for the hearer to provide only a non-verbal token of acknowledgment, yielding the pseudo-dyadic structure that is given in (7b). As the importance of the answer wanes, the initial question sheds much of its function as an interrogative speech act. In the third and final developmental stage represented in (7c), it merely expresses a condition on the consequence that is expressed in the apodosis. In such a monadic structure, the apodosis may occur without a linking element (*dann* 'then'), since the forefield position is represented by the conditional clause.

- (7) a. Dialogual phase (dyadic)
question – positive answer – consequence
Ist das Kind jünger als drei Jahre? – Ja. – Dann fährt es gratis.
- b. Sequential phase (pseudo-dyadic)
question – (acknowledgement) – consequence
Ist das Kind jünger als drei Jahre? – <nod> – Dann fährt es gratis.
- c. Sentential phase (monadic)

conditional scenario – () – consequence
Ist das Kind jünger als drei Jahre – () – (dann) fährt es gratis.

The dialogical account of verb-first conditionals is intuitively appealing, since it offers a natural, discourse-based explanation of how a complex present-day construction might have come into being. Its integration of two paratactic structures into a single hypotactic structure under a single intonation curve falls squarely into the prototypical developmental scenario of complex clauses (Givón 1979, 2009, Leuschner and Van den Nest forthcoming). Precisely because of this intuitive appeal, we should be wary and consider potential fallacies of this model along with alternative explanations that have been offered in the literature. Fischer (2007: 217) rightly cautions that for many purported developments from parataxis to hypotaxis, there is sparse or no evidence. She points out that any integrated or compressed clausal structure could theoretically be viewed as the end result of clause-combining, if only the analyst is inclined to take this view. If strict unidirectionality from parataxis to hypotaxis is assumed, this would for instance lead analysts to the erroneous conclusion that a recent example of clause elaboration represents older language use than the ‘reduced’ example, which in fact formed the basis for elaboration.

Harris and Campbell (1995) point out that there are viable alternatives to the hypothesis that complex sentential constructions emerge out of a development from parataxis to hypotaxis; in particular, they suggest that a productive source for the formation of complex constructions is analogy, i.e. the extension of an existing complex structure to a new context:

It is also possible that structural marking that developed in one context was later *extended* to another. While the issue of whether the sources of markers logically imply the sources of structures is an empirical one, we shall refer here to the assumption that they do as the Marker/Structure Fallacy. (Harris and Campbell 1995: 284)

While it is undoubtedly true that the development of hypotaxis out of parataxis is not the only logical possibility for complex constructions to emerge, Harris and Campbell appear to argue a much stronger point. With regard to the relationship between questions and subordinate clauses, they explicitly reject a dialogical explanation, arguing that ‘in explaining complex structures it is not necessary to go beyond the boundary of the sentence (to discourse) or to cite structures in which subordinate clauses have vague relationships to matrix clauses’ (Harris and Campbell 1995: 287).

How then could verb-first conditionals be explained without recourse to a dialogical structure? Harris and Campbell (1995: 298) build their argument on the observation that questions and many types of dependent clauses, among them conditionals, are pragmatically non-assertive. Questions could therefore be analogized to conditional clauses and ‘plugged into’ an existing syntactic frame. If we take the example of verb-first conditionals in German, we could hypothesize that speakers were sensitive enough to the pragmatic similarity between polar questions and the conditional clauses in (8a) and (8b) such as to draw an analogy and extend the use of questions to the expression of a non-assertive condition (8c).

- (8) a. *Wenn das Kind jünger als drei Jahre ist – dann fährt es gratis.*
b. *Falls das Kind jünger als drei Jahre ist – dann fährt es gratis.*
c. *Ist das Kind jünger als drei Jahre – dann fährt es gratis.*

Harris and Campbell argue that the formal similarity between actual dialogue and hypotactic structures is a secondary effect that should not be interpreted as supporting evidence for a dialogical account (1995: 308). If a phenomenon can be explained through well-understood

mechanisms that are known to operate within smaller syntactic units, such explanations are preferable.

The analogical account is, however, less straight-forward than it may initially appear. First, analogy is not in fact as well-understood as Harris and Campbell suggest (cf. Itkonen 2005, chapter 1). Second, the available evidence does not warrant the categorical, across-the-board rejection of dialogual explanations that Harris and Campbell advocate. As they point out themselves in the longer quote above, the relation between source and target is an empirical issue that would have to be settled on a case by case basis. Third, there are structural differences between the conditional protasis in (7c) and the corresponding subordinate clauses in (7a) and (7b) that make the idea of analogy for this particular example quite implausible. Further pertinent criticisms are given in Van den Nest (forthcoming). But regardless of these considerations, the general point made by Harris and Campbell – simpler explanations being preferable – is of course valid.

At this point, it needs to be acknowledged that there are alternatives to both the dialogual account and the analogical account. In both Old High German and Old Swedish, verb-initial structures with declarative and other functions were commonly used (Hopper 1975). Verb-first conditionals may have emerged historically as juxtapositions of two verb-initial declarative clauses. If this were the case, they would in fact not have anything to do with questions, despite their synchronic similarity to polar interrogatives. With this important acknowledgment in mind, this paper nevertheless focuses on the two accounts outlined above, using present-day corpus data to evaluate them.

Different as the dialogual and the analogical account may seem, they do share the assumption that questions and conditional protases were semantically close enough for one to stand in for the other at some point. The major difference between them concerns the mechanism of this change. Whereas the dialogual account posits that verb-first conditionals exemplify the well-known development from parataxis to hypotaxis, which is a standard account of how reanalysis proceeds, the analogical account holds that verb-first conditionals exemplify a construction that developed through parasitic use of an existing hypotactic structure. The general pattern of a protasis followed by an apodosis was already available, affording the development of new conditional constructions by means of an equally well-known mechanism, namely analogy.

While both accounts merit consideration, neither presents a case that would make it inherently and conclusively preferable. At this point, choosing one over the other is a matter of theoretical inclination. An empirical assessment of the question is needed, and it is what this paper aims to provide. But what is the most appropriate source of empirical data? Historical data come to mind as the best possible testing ground, but verb-first conditionals are attested from very early stages of German and Swedish respectively. In German, verb-first conditionals have been in use since Old High German times (Harris and Campbell 1995: 296); in Swedish, examples from the 15th century such as (9) are attested in corpus data.

- (9) *Haffwer hon onth effter barn byrdh tha siwde grabo i watn*
has she pain after child birth then boil mugwort in water
'If she has pain after childbirth, boil mugwort in water.'

Even if historical corpus data can shed light on the constructional development, historical examples cannot yield conclusive evidence about the ultimate origins of verb-first conditionals. Due to the written nature of the historical records, dyadic structures of the kind that would lend credence to the dialogual account are simply not preserved. The analogical account predicts an unmediated onset of verb-first conditionals at some point, but this is what we expect from a patchy written record anyway.

A crucial question to ask is whether the two accounts make different predictions with regard to the structures that are found in current language use. Ideally, different scenarios of how a process of grammaticalization took place would make conflicting predictions that could be tested through the quantitative analysis of corpus data. If we contrast the dialogual account of verb-first conditionals with the analogical account, the following characterizations emerge.

The dialogual account claims that the grammaticalization of verb-first conditionals represents a case of clause union from parataxis to hypotaxis. From this, we can derive a first prediction, namely that the clausal integration of the construction will tighten over time. Even if there is no necessity for protasis and apodosis to fuse completely, they should gradually become more cohesive. Second, the dialogual account claims that the protasis, which originates as an independent polar question, gradually loses its interrogative illocutionary force. From this it follows that the protases of verb-first conditionals should, over time, retain some semantic and pragmatic properties of questions that eventually disappear. For instance, we would expect that the apodoses of early verb-first conditionals would resemble polar questions in expressing real possibilities that may or may not be the case, rather than counterfactual scenarios. Third, the dialogual account would predict that the protasis, as an erstwhile question, generally continues to precede the apodosis. While this is usually the case, there are modern German examples such as (10), in which the order is reversed.

- (10) *Ich bin im Büro, solltest du mich brauchen.*
I am in.the office should you me need
'I'm in the office, should you need me.'

In example (10), the apodosis follows the matrix clause as a mere afterthought, not as a full-fledged condition. On the dialogual account, the reversed order of condition and consequence should become available only after the original dyadic structure has been completely reanalyzed as a monadic, hypotactic structure that can be rearranged to fit different information-structural needs. It is hence expected that this particular order is only seen at a late stage of grammaticalization, and even then as a relatively rare exception.

These predictions of the dialogual account cannot be derived from the analogical account. First, if we claim that questions came to be used as conditional protases by means of analogy, there is no reason to think that clausal integration between protasis and apodosis should have become tighter over time. Since no such change is expected from the conditional constructions serving as the basis for the analogy, there is no change to be expected from verb-first conditionals. Second, there is no reason to predict that the protases of verb-first conditionals should become less question-like over time. The very precondition for their appearance in a hypotactic conditional frame would be that speakers could take them to be conditional protases, not questions. Third, on the analogical account, postposed verb-first conditionals should appear at the same rate as other postposed conditionals. The postposition of conditional protases is of course quite rare, both cross-linguistically and across different conditional markers (Diessel 2001), but to the extent that it occurs, verb-first conditionals should not behave any different from other members of this grammatical category.

We thus conclude that the two accounts make different predictions, even for language use after the original inception of the construction. But are these predictions falsifiable? Whatever conclusions we hope to draw from a corpus analysis, they will only be as strong as the predictions at the basis of the investigation. In fact, the predictions that can be derived from the analogy account are so weak as to make the theory resistant to almost any empirical criticism. This does not prove it wrong, but in general scientific progress is only possible if theories hold the potential to be falsified. The dialogual account does make a number of falsifiable predictions, and so the present study will examine the evidence that speaks to these

predictions. In the case that the dialogual account can be proven wrong, the analogy account might be a good candidate to retain until the means to disprove it become available.

Questions such as the ones outlined above can be fruitfully investigated through the use of diachronic corpora, which allow the researcher to study a given phenomenon across different historical stages. Van den Nest (forthcoming) takes this route and studies English and German verb-first conditionals on the basis of diachronic corpus data. Three findings emerge that are crucial to the present discussion. First, with respect to German, it is established that verb-first conditionals show less clausal integration in Old High German than is observed in Present-day German. Fully integrated examples, in which there is no linking element and in which the apodosis begins with the inflected verb, do not occur in OHG at all. This finding is fully consistent with the dialogual account. Second, Van den Nest observes a variety of linking elements between protasis and apodosis in OHG, whereas in PDG there are only the elements *so* ‘so’ and *dann* ‘then’. The specialization process of *so* and *dann* suggests that verb-first conditionals gradually lost their question-like properties and narrowed pragmatically into the function of conditional protases. Again, this is consonant with the predictions of the dialogual account. Lastly, Van den Nest shows that in both OHG and Old English, verb-first conditionals predominantly expressed real possibilities, which stands in contrast to the synchronic state of affairs. In Present-Day English, one of the three verb forms found in verb-first conditionals is the form *had*, which exclusively expresses counterfactuals. In PDG, the expression of actual possibilities is more frequent than in English, but still the cognate form *hätte* accounts for substantially more counterfactual examples in PDG as compared to OHG. The rise of counterfactual verb-first conditionals receives a straightforward explanation in the dialogual account as a corollary of weakening interrogative illocutionary force; on the analogical account this development would have to be seen as a mere coincidence. Overall then, the comparison of historical and modern data yields evidence for the dialogual account, even though Van den Nest is careful to point to the scenario that verb-first conditionals may have emerged historically from juxtaposed verb-initial declarative clauses.

The present paper takes a similar approach, but instead of comparing historical stages of the same language against each other, it compares modern verb-first conditionals across two different languages, namely present-day German and Swedish. Like historical comparisons, synchronic comparisons can also be used to investigate language change, if only indirectly so. For instance, it is a common practice in socio-linguistic studies of language change to record subjects of different ages and to establish whether the parameter of age correlates with a given phonological change. Similarly, it is possible to learn something about language change by sampling a grammatical construction from two contemporary languages. If we can determine that the construction is ‘developmentally older’ in one language, we can draw a number of comparisons and check whether the differences between the constructions match our predictions of how these constructions developed in the two languages. If there are conflicting theories, the theory that aligns best with the observed differences should be favoured.

Regarding the question whether verb-first conditionals are more strongly grammaticalized in German or in Swedish, we could take relative text frequency as an index. Provided that there is a difference, the dialogual account makes the following predictions for the cross-linguistic data: First, the construction with higher text frequency will also display greater clausal integration, as measured by the presence of linking elements between protasis and apodosis. Second, the more frequent construction will retain fewer question-like features than its counterpart. These features could be operationalized in a number of ways, but the semantic feature of counterfactuality stands out as a candidate. Third, the more frequent construction will display greater independence from the erstwhile dialogual form and hence show a higher rate of protasis postposition (cf. example 9). If these predictions prove to be

wrong, the dialogical account would appear questionable. If, on the other hand, they hold up, this would serve as a rebuttal to Harris and Campbell (1995: 308), who state that there is no empirical evidence to support a dyadic origin of complex sentential constructions.

3 Data and analysis

The present study draws on several corpora of contemporary written German and Swedish, each with a bias towards newspaper texts. All corpora used are freely accessible over the World Wide Web. The German corpora used in this study were accessed through the IDS Mannheim (<http://www.ids-mannheim.de/cosmas2>). The German LIMAS corpus, which is modelled on the BROWN corpus of American English (Francis and Kucera 1964), holds one million words of running texts from various written genres. The Mannheimer Morgen corpus holds approximately 20 million words of newspaper text. The corresponding Swedish corpora were accessed through the Språkbanken resource (<http://spraakbanken.gu.se/>) at Gothenburg University. The Stockholm-Umeå corpus is a one-million word corpus that was constructed to match the BROWN corpus; the PAROLE corpus consists of 19.4 million words of mainly newspaper texts, but also some narrative prose and some internet-based texts. It is assumed that the German and Swedish resources are broadly comparable.

Since the construction under investigation does not have any lexically specified parts, the search procedure had to cast a wide net in order to exhaustively retrieve all tokens from the corpora. The strategy taken here was to search all corpora for sentence-final punctuation followed by a sentence-initial verb. This search procedure yielded a large number of false positives that had to be removed manually from the database. Verb-initial constructions to be removed included questions (11a), exclamatives (11b), topicalized passives (11c), imperatives (11d), truncations (11e), and several other structures. For an overview of Swedish verb-initial constructions, see Lindström and Karlsson (2005).

- (11) a. *Sind unsere Kinder weniger intelligent als früher?*
are our children less intelligent than earlier
'Are our children less intelligent than they used to be?'
- b. *Ist das ärgerlich!*
is that annoying
'How annoying is that!'
- c. *Entlassen wird vorerst niemand.*
laid.off is for.now no-one
'There will be no layoffs for now.'
- d. *Kom in!*
come in
'Come in!'
- e. *Kunde man säga.*
could one say
'One could say that.'

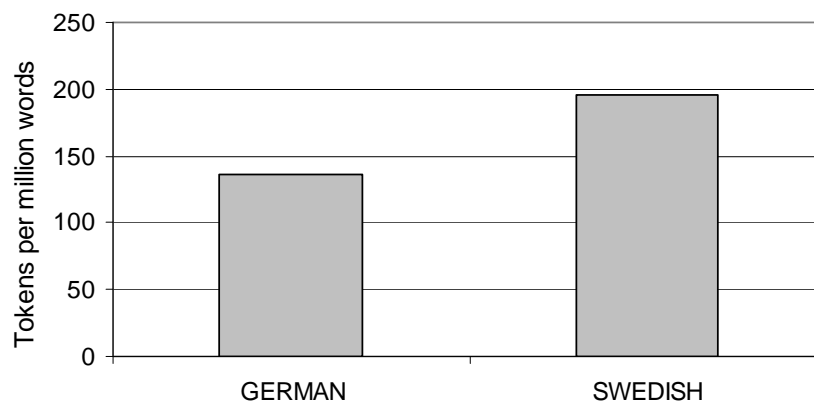
It is necessary to point out that the search procedure can be faulted for disregarding certain examples of verb-initial conditionals. Example (12a) would not be retrieved because the sequence of sentence-final punctuation and the initial verb is interrupted by the presence of a

conjunction. Example (12b) would not be found because a comma is not recognized as sentence-final punctuation. Spot checks in the corpora used indicated that such examples occur in negligible quantities. It is assumed that the omission of such examples in the interest of a manageable retrieval procedure does not invalidate the results.

- (12) a. *Und fragt man drei Experten, erhält man drei Meinungen.*
 and asks one three experts gets one three opinions
 ‘And if you ask three experts you get three different opinions.’
- b. *Also, hätte ich das gewusst, wäre ich nicht hingegangen.*
 well had I that known were I not gone.there
 ‘Well, had I known that, I wouldn’t have gone there.’

After the exclusion of non-target examples, the search yielded 2,859 German examples and 4,001 Swedish examples. Given that the Swedish corpus is somewhat smaller, this discrepancy in raw frequencies equals an even larger, statistically significant discrepancy in expected frequencies ($\chi^2=224.6$, $df=1$, $p<.001$). Figure 1 presents the difference in terms of tokens per million words. The analysis thus proceeds on the working assumption that in written Swedish, verb-first conditionals have grammaticalized to a stronger degree than their German counterparts. Based on this assumption, the dialogical account predicts that Swedish verb-first conditionals display tighter clausal integration, fewer question-like features, and greater syntactic independence from the original dyadic structure. The following paragraphs present several corpus-linguistic operationalizations of these predictions and discuss the findings from synchronic corpus data.

Figure 1: Frequency of verb-first conditionals in German and Swedish



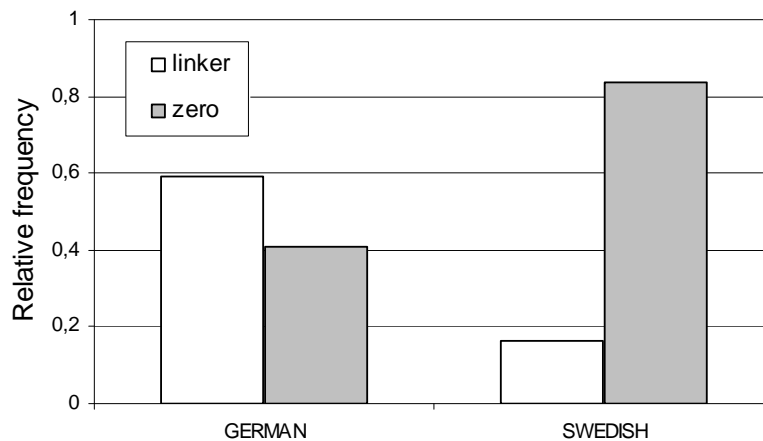
The degree of clausal integration between protasis and apodosis of a verb-first conditional can be measured by assessing the relative frequency of linking elements between the two. In German, verb-first conditionals may occur with *so* or *dann*, the corresponding Swedish elements are *så* and *då*. Alternatively, protasis and apodosis can simply be adjoined without an overt linking element. The examples in (13) illustrate these alternatives.

- (13) a. *Ändern sie das Testament, dann / Ø tritt eine neue Rechtslage ein.*
 change they the testament then sets a new situation in
 ‘If they change the testament, then / Ø a new legal situation obtains.’

- b. *Räknar man även barnen, så / Ø blir siffran avsevärt högre.*
 counts one even children then becomes number substantially higher
 ‘If also children are counted, then / Ø that number increases substantially.’

In the data retrieved from the German and Swedish corpora, a linking element occurs in 1,690 German verb-first conditionals, but only in 650 Swedish examples ($\chi^2=1063.2$, $df=1$, $p<.001$). Figure 2 illustrates that linkage represents the default case in German, but a fraction of less than 20% in Swedish. This finding is in line with the dialogual account, which would predict tighter clausal integration for Swedish.

Figure 2: Relative frequency of linking elements in German and Swedish



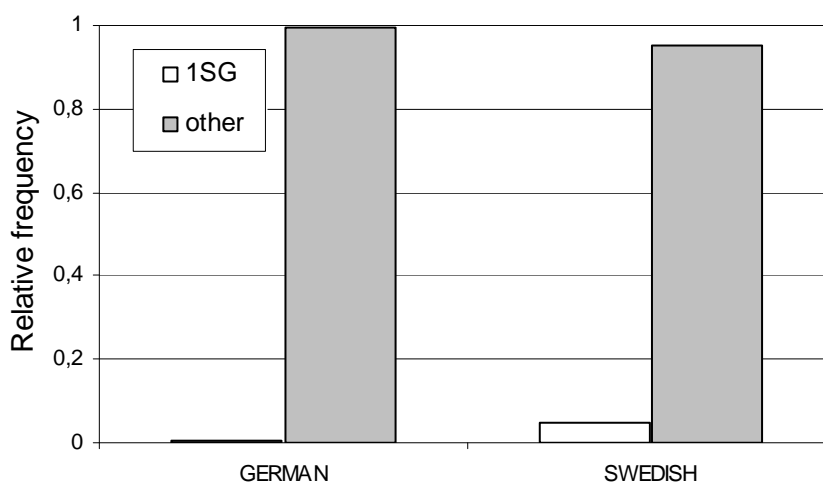
Given the assumption that Swedish verb-first conditionals are more grammaticalized than their German counterparts, the dialogual account would predict that the protasis of German verb-first conditionals should retain more semantic and pragmatic properties of questions.

As a first approximation, let us consider the choice of subject pronouns. Both questions and the protases of verb-first conditionals can occur with any type of pronominal subject. Since one basic function of polar questions is to elicit information that is not introspectively available, questions with first-person singular pronouns, as in (14a) and (14b), should be relatively rare.

- (14) a. *Habe ich mich falsch verhalten?*
 have I myself wrongly conducted
 ‘Did I do something wrong?’
- b. *Är jag så förargelseväckande ?*
 am I so annoying
 ‘Am I so annoying?’
- c. *Vergesse ich es zu reinigen, verstopfen die Späne den Abfluss.*
 forget I it to clean clog the shavings the drain
 ‘If I forget to clean it, the shavings will clog the drain.’
- d. *Skulle jag vara utan kött en hel vecka skulle jag säkert dö.*
 should I be without meat a whole week should I certainly die
 ‘If I were to go without meat for a week, I would surely die.’

We can formulate the hypothesis that if verb-first conditionals evolved out of questions, and Swedish ones are grammaticalized to a relatively stronger degree, German verb-first conditionals will retain more of an aversion against first person singular subjects. Examples such as (14c) should therefore be relatively less frequent than examples such as (14d). The quantitative data shown in Figure 3 support this position. While there are only 10 examples with German *ich* ‘I’ in the database, there are 198 examples with Swedish *jag* ‘I’. The relative frequency of first-person singular examples is thus very low in both languages, but the difference between the languages is significant ($\chi^2=118.3$, $df=1$, $p<.001$).

Figure 3: Relative frequency of 1SG pronouns in German and Swedish



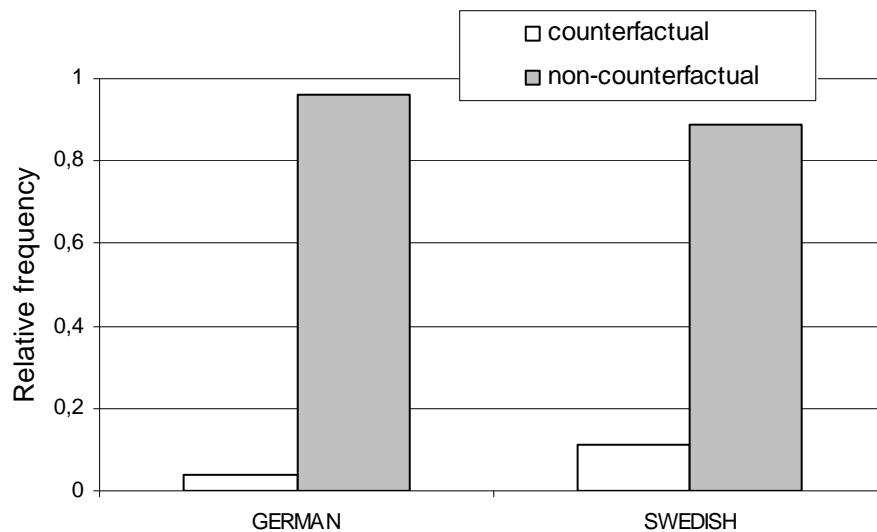
As was already discussed above, another approach to the relative similarity of questions and verb-first conditionals above concerns the distinction of realis and irrealis mood and more specifically the relative frequency of counterfactuals. The dialogual account assumes that verb-first conditionals emerged in a scenario of real possibilities. Asking one’s interlocutor about something known to be not the case would be poor pragmatics. It can therefore be hypothesized that counterfactual verb-first conditionals represent a relatively late developmental stage, which in turn leads us to expect relatively fewer counterfactuals in the German data. In order to measure the relative frequency of counterfactuals, we can rely on the Swedish verb forms *hade* ‘had’ and *vore* ‘were’ and their respective German cognates. The examples in (15) illustrate how these forms are used to express counterfactual conditions.

- (15) a. *Hätte ich das gewusst, wäre ich nicht hingegangen.*
had I that known were I not gone.there
‘Had I known that, I wouldn’t have gone there.’
- b. *Hade de frågat mig hade jag säkert svarat ja*
had they asked me had I surely answered yes
‘If they had asked me, I would have said yes.’
- c. *Men vore det nu jag, så sökte jag nåd hos Gud.*
but were that now I so searched I mercy with god
‘But if it were me, I would be looking for God’s mercy.’

- d. *Wäre der Wagen nicht gepanzert gewesen, hätte es ein Unglück gegeben.*
 were the car not armored been had it a disaster given
 ‘If the car had not been armored, there would have been a disaster.’

The dialogical account predicts that such examples will account for a relatively smaller proportion of the German data, as compared to the Swedish data. Figure 4 shows that this prediction is borne out. There are 116 German counterfactual examples as opposed to 445 Swedish counterfactuals ($\chi^2=109.1$, $df=1$, $p<.001$).

Figure 4: Relative frequency of counterfactuals in German and Swedish



As a third assessment of the relative similarity of questions and verb-first conditionals, we now turn to collocational overlap. This step is motivated by the reasoning that if verb-first conditionals develop in such a way that they become less and less question-like over time, then German verb-first conditionals and polar questions should occur with a roughly similar set of verbs, while Swedish verb-first conditionals and polar questions should show less collocational overlap. Again, this measure turns to the inflected verb of the conditional protasis. The examples in (16) show that the same verb may occur both in a question and a conditional.

- (16) a. *Gibt es keine Suppe mehr?*
 gives it no soup anymore
 ‘Is there no soup left?’
- b. *Gibt es keine Suppe mehr, essen wir eben belegte Brote.*
 gives it no soup anymore eat we just sandwiches
 ‘If there is no soup left, we’ll just have sandwiches.’

There are different ways to measure collocational overlap. For the present purposes, both the German and the Swedish corpora were searched for polar questions, which are of course much more frequent than verb-first conditionals. An exhaustive collocational analysis of polar question was therefore not attempted. Instead, random samples of 1,000 questions were generated from the results, and these were subsequently compared to random samples of 1,000 verb-first conditionals from German and Swedish. Collocational overlap was measured

by counting the raw frequencies of matches. To illustrate, the German sample of questions included 23 examples with the verb form *gibt* 'gives'. The corresponding sample of verb-first conditionals holds seven examples with *gibt*. This results in a count of seven matches. Counts are tallied for all matching forms. The results of these counts stand in contradiction to the prediction of the dialogal account. Swedish verb-first conditionals actually show greater collocational overlap with polar questions than their German counterparts. Whereas there are 414 matching examples for Swedish, there are only 368 for German. The proportional difference is not significant ($\chi^2=2.59$, $df=1$, n.s.). This particular measure of collocational overlap does therefore not distinguish between the two accounts.

The last measure to be discussed here concerns the relative positions of protasis and apodosis in verb-first conditionals. In German and English, the protasis may appear postposed, as an afterthought to the apodosis. Alternatively, it may be parenthetically inserted into the apodosis.

- (17) a. *Ich bin im Büro, solltest du mich brauchen.*
 I am in.the office should you me need
 'I'm in the office, if you should need me.'
- b. *Er wird, sollte es nötig sein, eine zweite Chance erhalten.*
 he will should it necessary be a second chance receive
 'If it will become necessary, he will get a second chance.'
- c. *Nurse, I shall leave instructions for Anna's treatment, should it be necessary.*
- d. *She would, were it to be successful, be more in need of his favours than Harvey's.*

If we assume a dyadic structure as the ultimate source of verb-first conditionals, the examples in (17) require some explanation, since they violate the normal sequence of question and answer. On the dialogal account, it would have to be reasoned that these examples represent a late developmental stage in which verb-first conditionals have become completely emancipated from their erstwhile source. Displaced protases should only become an option with an advanced degree of grammaticalization, after which they may gradually increase in frequency. With regard to the contrast of German and Swedish, the dialogal account predicts that displaced protases are relatively more frequent in Swedish.

In order to test this prediction, the corpora were searched for German *sollte/sollten* 'should' and its Swedish cognate *skulle*, each preceded by a comma. The resulting concordances were inspected manually to identify postposed and parenthetically inserted protases. The inspection reveals that the prediction of the dialogal account is not borne out. In the German data, there are 40 displaced protases with *sollte* and *sollten*; in the Swedish data there are none at all. This is consistent with the standard reference grammar of Swedish (Teleman et al. 1999: 647), which describes the initial position of the protasis as near-obligatory. Still, postposed protases are not altogether ungrammatical, as the following examples suggest.

- (18) a. *Jag gör det gärna, får jag bara tid.* (Henrik Rosenkvist, p.c.)
 I do that gladly get I only time
 'I'll do that gladly, if I can find the time.'
- b. *Vi har mer mandelmassa, skulle det behövas.* (internet forum)
 we have more almond paste should it be.needed

‘There is more almond paste, should we need it.’

- c. *Det klarar han inte, håller han så på i tio år.* (Teleman et al. 1999: 467)
that manages he not keeps he so on for ten years
‘He won’t manage, if he keeps going like that for ten years.’

Example (18b) is gathered from an internet forum in which the grammaticality of the sentence is debated in a folk-linguistic way. The author of the example discusses a situation in which she uttered the example at the kitchen table, only to be criticized by her father for using what he perceived as normatively incorrect grammar. While this reaction lends support to the observations of Teleman et al., showing that some speakers of Swedish do not consider the sentence to be part of their grammar, the three examples evidence that this pattern does surface in usage from time to time. At any rate, the low frequency of displaced protases in Swedish verb-first conditionals runs counter to predictions of the dialogal account and hence does not corroborate it.

4 Conclusions

In the introduction of this paper, it was asked whether synchronic gradience could tell us anything about the reanalysis of one grammatical structure as a different one. In view of the analyses presented in the previous section, the short answer to that would be a tentative yes. If there are conflicting accounts of how a given process of reanalysis happened, we can turn towards gradience in present-day corpus data and investigate whether the predictions of these accounts are consistent with quantitative tendencies that we observe.

In the ideal case, the results of different measures will harmonically align and thus distinguish reliably between the alternative scenarios. Many corpus linguists will agree that this rarely happens in practice. Messy data are the rule, rather than the exception. In the present study, six measures were applied to the cross-linguistic analysis of verb-first conditionals: besides the text frequency of the constructions these measures probed the relative frequency of linking elements, first-person singular subject pronouns, and counterfactuals, the relative degree of collocational overlap, and the presence of displaced protases. Out of these six, the first four align in a way that is consistent with the classic dialogal account of verb-first conditionals. Higher text frequency points towards a relatively higher degree of grammaticalization of the Swedish construction, fewer linking elements, more first-person subjects, and more counterfactuals corroborate this assumption. In combination with the diachronic evidence that is available (Van den Nest forthcoming), this means that the dialogal account is actually doing quite well – several strong predictions turn out to be correct. That said, greater collocational overlap with questions and the complete absence of displaced protases in Swedish remain in need of an adequate explanation. Also, these findings do not falsify the analogy account by Harris and Campbell (1995), much less the alternative account of two juxtaposed declarative clauses (Van den Nest forthcoming). What an analysis of synchronic gradience primarily provides is an assessment of relative plausibility, given two alternatives. In addition, the results reported in this paper challenge alternative analyses of the same topic to provide explanations why we see this particular constellation of structures and frequencies.

Another remaining question with regard to verb-first conditionals is the question of genre and modality differences. Verb-first conditionals remain a widely applicable construction in written and spoken Swedish (Auer and Lindström 2008), whereas in German the construction tends to occur in formal written contexts of stating regularities or law-like procedures, as illustrated in (19) below.

- (19) *Ist ein Betrieb zahlungsunfähig, stehen drei Möglichkeiten zur Verfügung.*
is a company bankrupt stand three options at disposal
'If a company is bankrupt, then there are three options.'

Genre-dependency would account for the higher frequency of displaced protases in German, since this phenomenon is quite rare except in the genre of elevated journalese, which permits the use of markedly complex structures more freely than other genres do.

Another important question concerns the issue how English fits into the overall picture. Whereas verb-first conditionals are productive in Swedish and German, the English construction is restricted to *had*, *should*, and *were*. Do the three languages represent different developmental stages of similar processes, or are we in fact looking at different phenomena?

On a more general level, the results of this study suggest that the process of reanalysis is gradual; a reanalyzed structure retains aspects of its original source, and it does so for considerable time. This means that synchronic, cross-linguistic comparisons of cognate constructions can be used to evaluate theories about the development of these constructions. Given a proposed development from A to B, it can be investigated whether a construction in modern usage of one language is, in several respects, closer to the hypothesized source than a corresponding construction in another language. One lesson to be learned from the present study is that constructions may not behave uniformly across the chosen measures. Still, quantitative information from modern corpora is useful, as it can provide fine-grained comparisons of constructions which can ultimately be used to decide between theories.

References

- Aarts, Bas. 2007. *Syntactic Gradience. The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.
- Auer, Peter and Jan K. Lindström. 2008. Verb-first conditionals in German and Swedish. Paper presented at FRIAS workshop on *Emergent Constructions*, May 2008.
- Diessel, Holger. 2001. The ordering distribution of main and adverbial clauses: A typological study. *Language* 77, 343–65.
- Erdmann, Oskar. 1886. *Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung*. Stuttgart: J.G. Cotta.
- Fischer, Olga. 2007. *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: Oxford University Press.
- Francis, W. Nelson and Henry Kucera. 1964. *Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English, for Use with Digital Computers*. Providence: Brown University.
- Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
- . 2009. Multiple routes to clause union: The diachrony of complex verb phrases. In *Syntactic Complexity*, T. Givón and M. Shibatani (eds), 81-118. Amsterdam: John Benjamins.
- Harris, Alice and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hilpert, Martin and Christian Koops. 2008. A quantitative approach to the development of complex predicates: The case of Swedish Pseudo-Coordination with *sitta* 'sit'. *Diachronica* 25/2, 242-261.
- Hopper, Paul J. 1975. *The syntax of the simple sentence in proto-Germanic*. The Hague: Mouton.
- Hopper, Paul J. and Elizabeth C. Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.
- Itkonen, Esa. 2005. *Analogy as Structure and Process. Approaches in linguistics, cognitive psychology and philosophy of science*. Amsterdam: John Benjamins.
- Jespersen, Otto. 1940. *A Modern English Grammar on Historical Principles. Part V: Syntax*. Copenhagen: Munksgaard.
- Leuschner, Torsten and Daan Van den Nest. Forthcoming. The Diachronic Emergence of Hypotaxis as Prototype: Processes and Consequences in German and Dutch. *Language Sciences*.
- Lindström, Jan K. and Susanna Karlsson. 2005. Verb-first constructions as a syntactic and functional resource in Swedish. *Nordic Journal of Linguistics* 28, 97-131.
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. Halle: Niemeyer
- Traugott, Elizabeth C. 1985. Conditional Markers. In *Iconicity in Syntax*, J. Haiman (ed.), 289-307. Amsterdam & Philadelphia: John Benjamins.
- Traugott, Elizabeth C. Forthcoming. On the intersection of synchronic variation with diachronic grammaticalization.
- Van den Nest, Daan. Forthcoming. Should conditionals be emergent ... Asyndetic Subordination in German and English as a Challenge to Grammaticalisation Research. In *Grammaticalization and grammar*, K. Davidse, H. Cuyckens, and J.-C. Verstraete (eds). Amsterdam: John Benjamins.