

A quantitative approach to the development of complex predicates

The case of Swedish Pseudo-Coordination with *sitta* “sit”*

Martin Hilpert and Christian Koops

Freiburg Institute for Advanced Studies / Rice University

This paper traces the historical development of the Swedish Pseudo-Coordination construction with the posture verb *sitta* “sit”. In Swedish a small number of verbs, including posture verbs such as *sitta*, are used in coordination with another verb to convey that the described event has an extended duration or is in progress. Quantitative evidence from Swedish historical corpora suggests that the construction has, even after it established itself as a grammatical construction, undergone a number of gradual changes in the course of the past five centuries. As part of the Pseudo-Coordination construction, the verb *sitta* has changed its argument structure, and the entire construction has increased in syntactic cohesion.

Keywords: complex predicates, pseudo-coordination, Swedish, Germanic, corpus linguistics, grammaticalization, gradual change

1. Introduction

Complex predicates, as defined by Alsina et al. (1997) and Bower (this volume), are not a prominent feature of the Germanic languages. Studies of complex predicates in African, Asian, Australian, and Oceanic languages have addressed the phenomenon in discussions of serial verb constructions (Foley & Olson 1985, Aikhenvald 2006), light verb constructions (Butt 1995, Harris this volume, Hook & Pardeshi 2006), and restructuring predicates (Di Sciullo & Rosen 1990, Alsina

* We would like to thank Claire Bower and the participants of the 11th Biennial Rice Linguistics Symposium for their helpful suggestions. Martin Hilpert thankfully acknowledges funding from the Deutscher Akademischer Austauschdienst.

1997). These studies have provided a rich typology of the possible forms and functions. One might ask, then, what additional insights into complex predicates can be gained by studying the well-researched grammatical systems of, say, English or Swedish.

In this paper we argue that it is specifically the diachronic study of complex predicates that stands to benefit from a focus on Germanic. What makes this possible are the substantial amounts of historical text that are available today for most Germanic languages in electronic form. Together with appropriate corpus linguistic techniques, historical corpora allow the diachronic study of the Germanic languages at a level of detail that is not available for most of the languages treated in the works cited above (see also the comments by Butt & Lahiri 1998: 27). Our diachronic perspective thus complements the existing body of synchronic, fieldwork-based case studies.

Of particular interest in this regard are, of course, typological generalizations about the historical development of complex predicates. For example, synchronic typologies of complex predicates indicate that the diachronic sources of light verbs are strikingly uniform across unrelated languages. This suggests that certain types of light verb constructions develop in quasi-universal ways. A case study of a complex predicate in a Germanic language can serve as a test case to explore this possibility, given the availability of relatively large amounts of textual data spanning several centuries.

Historical corpus data have typically been used to demonstrate qualitative contrasts between examples from different time periods. To take a well-known example, consider the following examples of the English subordinating conjunction *while* in Old English and Early Modern English (Hopper & Traugott 2003: 85).

- (1) *Ðæt lastede þa [xix] winttre wile Stephne was king*
 “That lasted those 19 winters while Stephen was king.”
 (ChronE [Plummer] 1137.36)
- (2) *Whill others aime at greatnes boght with blod,*
Not to bee great thou strives, bot to bee good.
 “While others aim at greatness that is bought with blood,
 You strive not to be great but good.”
 (1617, Sir W. Mure, *Misc. Poems* xxi.23 [OED while 2b])

The examples illustrate that the co-temporal meaning of *while*, which is clearly present in (1), gave rise to the concessive meaning seen in (2). The latter meaning can also be observed today in purely concessive usage, e.g. *While dolphins inhabit the sea, they are actually mammals*. Examples like (1) and (2) show that a change has occurred. However, the comparison of individual examples does not yield a precise picture of when this change was initiated and how it proceeded. As we see

it, an important benefit of large historical corpora is the possibility of precisely quantifying shifting usage patterns of linguistic forms, especially their rate of occurrence in different grammatical contexts. Of course, for historical corpus studies to be feasible, questions about the development of complex predicates have to be phrased in such a way that quantitative evidence yields meaningful answers. In other words, it is not enough to report frequencies as an end in itself. We have to develop with hypotheses that can be either corroborated or falsified through frequency data.

The study of frequency patterns in language use is not universally accepted as a promising avenue of research. Bybee & Hopper (2001:1) explain this with reference to “the widespread acceptance of the premise that language structure is independent of language use”. However, a growing body of literature documents counter-evidence to this premise. This evidence comes in the form of frequency effects, i.e. structural differences between related language forms which can be shown to correlate with a difference in discourse frequency. For example, high token frequency commonly leads to phonetic and phonological reduction. This explains why *memory* and *family* tend to be produced as bi-syllabic words in American English, while the phonetically similar but less frequent items *mammary* and *homily* are tri-syllabic (Hooper 1976). Another frequency effect can be observed in paradigmatic regularization. The irregular past tense forms of verbs such as *weep* and *leap* tend to be regularized to *weeped* and *leaped*, whereas the corresponding forms of the more frequent verbs *keep* and *sleep* are retained as *kept* and *slept* (Bybee 1985). At the level of syntax, recent work by Gahl & Garnsey (2004) shows that frequencies of syntactic patterns correlate with the phonetic reduction of certain elements. The authors demonstrate experimentally that past tense forms of verbs show a higher rate of final /t/ or /d/ deletion if the verb is produced in the context of a complementation pattern that is highly typical for that verb. To illustrate, a reduced production of the past tense form *confirmed* is more likely in (3), where the verb is followed by a direct object, than in (4), where it has a sentential complement.

(3) *The CIA director confirmed [the rumor]_{DO} once it had spread widely.*

(4) *The CIA director confirmed [the rumor should have been stopped sooner.]_{SC}*
Gahl & Garnsey (2004: 762)

The authors show that the bias of *confirm* to occur with direct objects, rather than sentential complements, contributes significantly to this effect.

Our approach to the development of complex predicates in this paper is fully aligned with the idea that language structure is shaped by language use. More specifically, we assume that the diachronic study of frequency patterns can further our understanding of different aspects of complex predicates, such as the integration

of parataxis into hypotaxis (Givón 2006) or the semantic development of lexical verbs into grammatical elements (Bybee et al. 1994, Kuteva 2001). Many of these issues have already been addressed on the basis of crosslinguistic synchronic data (Heine et al. 1991, Bybee et al. 1994, Svorou 1994), *inter alia*, and various semantic regularities across genetically unrelated languages have been established (e.g. Heine & Kuteva 2002). For instance, verbs of movement typically come to express the intentions of human agents and ultimately develop into markers of futurity, as seen for example in French *aller* “go” or English *be going to* (Bybee et al. 1994). Given these regular tendencies, it is possible to derive predictions about parallel diachronic developments in different languages.

2. Swedish Pseudo-Coordination

If something can be learned from the quantitative, diachronic analysis of a Germanic complex predicate, what would be a good candidate for a case study? In this paper we focus on the development of the Swedish Pseudo-Coordination construction (SPC) with the posture verb *sitta* “sit”. Crosslinguistically, posture verbs often grammaticalize into markers of imperfective aspect (Bybee et al. 1994, Austin 1998, Kuteva 2001), so Swedish *sitta* can be hypothesized to have developed along a well-known grammaticalization path. A modern example of this construction is given in (5).¹

- (5) *Vi bara satt och pratade.* (New novels)
 we just sat and talked
 “We were just talking.”

In Swedish, as in Danish and Norwegian, a small number of verbs, including posture verbs such as *sitta* “sit”, can be used in coordination with another verb to impose a particular aspectual contour on the event coded by the second verb. The resulting interpretation is typically that the described event has an extended duration or is in progress, i.e. as durative or progressive. This can be seen in the English translation of example (5). The translation also shows that the lexical meaning of *sitta*, i.e. the idea of being in a sitting posture, is not necessarily a prominent part of the sentence’s meaning (although an interpretation with an emphasis on the postural semantics of *sitta* is not ruled out; see below).

1. Throughout this paper, examples taken from a corpus are cited using the following abbreviations: *Gothenburg* (*Gothenburg Spoken Language Corpus*), *New novels* (*Bonnier’s romaner II* corpus), *Old novels* (*Äldre svenska romaner* corpus), and *Källtext* (*Källtext* corpus). See §4.1 for a discussion of these corpora. Examples without citation are constructed.

A range of additional syntactic and semantic effects have been pointed out which distinguish the SPC from canonical coordination (Wiklund 1996, Teleman et al. 1999, Lødrup 2002, *inter alia*). We review the evidence in the remainder of this section. Previous analyses differ in certain details, but there is general agreement on a number of basic syntactic and semantic facts. In essence, while examples like (5) are superficially consistent with an analysis as nothing more than two coordinated clauses, a monoclausal analysis is more plausible. Taken together, the evidence warrants a view of the SPC as a complex predicate in the sense of Alsina et al. (1997), i.e. as a grammaticalized monoclausal construction with two verbal heads.

To start out, we need to clarify that the verbs involved in the SPC do not show the features typically associated with auxiliary verbs. Crucially, SPC verbs like *sitta* carry finite morphology, are not phonologically reduced relative to their lexical counterparts, and do not have deficient paradigms. There is also no overt marker of embedding. Thus, an analysis of, for example, *sitta* “sit” as an auxiliary is ruled out.

The standard reference grammar of Swedish (Teleman et al. 1999) characterizes the SPC as a construction in which certain semantically ‘light’ verbs, including *sitta* “sit”, *ta* “take”, and *gå* “go” are coordinated with a lexical verb. Another example is given in (6).

- (6) *Mona satt och sydde i det blå rummet.* (Teleman et al. 1999: 903)
 Mona sat and sewed in the blue room
 “Mona was sewing in the blue room.”

As Teleman et al. point out, the light verb cannot receive primary stress without important changes in interpretation. If *satt* “sat” is stressed in (6) the sentence is likely to receive the interpretation that this is unexpected or in some way unusual. Stress on *satt* leads to the conceptual separation of two actions — here, sitting and sewing — which the construction portrays as integrated by default.

Further evidence for monoclausality comes from the syntactic behavior of the SPC. First, unlike in main clause coordination, the second verb cannot have an overt subject. Example (7) is judged as questionable by Teleman et al. (1999).

- (7) [?]*Mona satt och hon sydde i det blå rummet.* (Teleman et al. 1999: 903)
 Mona sat and she sewed in the blue room
 Intended meaning: “Mona sat and sewed in the blue room.”

Secondly, the order of the two verbs in the SPC is fixed. They cannot be rearranged without disruption of the original meaning.

- (8) **Mona sydde och satt i det blå rummet.* (Teleman et al. 1999: 903)
 Mona sewed and sat in the blue room
 Intended meaning: “Mona sewed and sat in the blue room.”

A third syntactic indicator of monoclausality is that the two-verb sequence cannot be modified with the adverb *både* “both”. Doing so, as in (9), leads to the loss of the original, aspectual interpretation (seen in [6]).

- (9) **Mona både satt och sydde i det blå rummet.* (Teleman et al. 1999: 903)
 Mona both sat and sewed in the blue room
 Intended meaning: “Mona both sat and sewed in the blue room.”

Finally, syntactic movement phenomena suggest that pseudo-coordinated verbs exhibit greater syntactic integration than ordinary coordinated structures. For example, while the object of *läsa* “read” in (10) can occur outside of its default (post-verbal) position, this is not possible in (11).

- (10) *Den där artikeln har jag suttit och läst hela dagen.* (Teleman et al. 1999: 903)
 that there article have I sat and read all day
 “That article I have been reading all day.”
- (11) **Den där artikeln har jag skrattat och läst hela dagen.*
 that there article have I laughed and read all day
 “That article he has laughed and read all day.”

Like sitting and reading, laughing and reading can be co-temporaneous activities, but the SPC cannot be used to express the latter if the object is fronted.

Teleman et al. (1999) note that the boundary between pseudo-coordination and regular coordination is fuzzy. Whether a given instance of, say, *sitta* in coordination with another verb is best analyzed as an instance of pseudo-coordination or as an instance of *sitta* used in canonical coordination cannot always be determined simply on the basis of positive evidence. However, as a rule of thumb, the more intervening elements occur between the two verbs, the weaker the conceptual union appears to be. Another way of stating the same generalization would be to say that the more the initial verb (e.g. *sitta*) is individually modified, the more it is understood to function as a lexical verb, whereas a lack of individual modification, as in (6) to (9), is suggestive of its light verb function. As we discuss in more detail below, this characterization is consonant with the changes seen in distributional behavior of the SPC in diachronic corpus data.

It should be noted that while *sitta* is one of the most frequent SPC verbs, the phenomenon of pseudo-coordination goes far beyond what is discussed in this paper. The verbs that occur in the SPC can be thought of as a semi-closed class encompassing several subgroups which correspond to the crosslinguistically

common sources for both light verbs and auxiliaries. Teleman et al. (1999) distinguish five broad classes. First are human posture verbs, especially *sitta* “sit”, *stå* “stand”, and *ligga* “lie”. The second group is formed by motion verbs like *komma* “come” and *gå* “go”. The verbs *börja* “start”, *hålla på* “hold on to, continue”, and *sluta* “stop” are subsumed under the category of “phase verbs” because they highlight a certain phase of an event. The fourth group consists of phrasal verbs with the copula *vara* “be” which encode polite requests. These are typically used in the imperative (e.g. *var snäll* “be nice” or *var vänlig* “be friendly”). The fifth category comprises verbs of communication, such as *ringa* “call” or *skriva* “write”.

3. Predictions and their corpus-linguistic operationalization

This paper aims to test three predictions that follow from the hypothesis that repeated usage events over time shape grammar (Bybee et al. 1994, Bybee & Hopper 2001). This hypothesis contrasts with the view that language change occurs largely in the domain of language acquisition, where children instantaneously reanalyze the input they receive from adults (e.g. Lightfoot 1991). Our view of grammaticalization leads us to expect that linguistic structures, such as the argument structure of a verb which forms part of a grammaticalizing construction, undergo gradual changes. For example, hearing a verb in a novel context, e.g. with an unconventional argument frame, will not lead to an instantaneous and complete reanalysis. Rather, it will merely invite further usage of the verb in that way. As grammaticalization tends to correlate with frequency changes, the developmental course of a grammaticalizing element should be reflected in changing frequency patterns across historical periods. For the present analysis, these assumptions lead to the following predictions regarding the development of the SPC.

3.1 Prediction 1: Change in argument structure

We assume that, as part of the grammaticalizing SPC, the verb *sitta* underwent a change in its argument structure. The reason for this lies in its changing semantics within the SPC, especially the gradual loss, or “bleaching out”, of the verb’s postural and locative meaning. We know from the synchronic data that the construction specifically brings out the aspectual meaning components of the verb, e.g. the sense of duration associated with sitting. At the same time, the postural and locative meaning components are backgrounded. Viewed over time, then, the more *sitta* functions as an aspectual marker, the less relevant its locative meaning becomes. This change in meaning and function should be reflected in the degree to which the verb is used together with prepositional phrases or other adverbials

specifying the location or manner of sitting. In the following, we refer to any such modification of the verb's lexical meaning as *locative elaboration*, and distinguish two broad types: cases where *sitta* is used with some form of locative elaboration and cases where it is not. Note that even in the lexical use of *sitta* locative elaboration is not obligatory. Both (12a) and (12b) are grammatical.

- (12) a. *Jag har suttit vid skrivbordet nästan hela dan.* (New novels)
 I have sat at desk.the almost all day
 "I've sat at the desk almost all day."
 b. *Han satt en stund igen.*
 he sat a while again
 "He sat (somewhere) for a while again."

Nevertheless, as part of the grammaticalizing SPC, the argument structure seen in (12b) should become more frequent. That is, we would expect cases like (12d) to increase in frequency relative to cases like (12c).

- (12) c. *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."
 d. *Jag har suttit och läst hela dagen.*
 I have sat and read all day
 "I have been reading all day."

In summary, to test the prediction of a change in the argument structure of *sitta* in the SPC we measure changes in the rate of locative elaboration (as defined above) in both the lexical and the SPC use of *sitta*. Non-elaborated usage should be more likely in the SPC overall and should increase with further grammaticalization of the SPC as a grammatical construction. On the other hand, when used outside of the SPC, as illustrated in (12a,b), the rate of locative elaboration should remain constant, because the original, lexical meaning of *sitta* does not change.

3.2 Prediction 2: Change in adverb placement

Our second prediction rests on the synchronic observation that the SPC is syntactically more cohesive than verbs in canonical coordination. This leads us to expect a measurable increase in syntactic unity over time, as the construction developed from expressing canonical coordination to pseudo-coordination. An indicator of syntactic cohesion is the rate at which the coordinated constituents of the SPC are modified either individually or jointly by temporal adverbials like *hela dagen* "all day" where such modification occurs. There are three logical possibilities for the placement of adverbs like *hela dagen* in the SPC: before, between, or after the two

verbs. When placed on the outside of the two-verb sequence, temporal adverbs are understood as modifying the event as a whole. On the other hand, adverbials placed between the two verbs — that is, following *sitta* — modify the first verb exclusively. We predict that the rate at which *sitta* is individually modified will decrease with increasing grammaticalization as a complex predicate because the view of the two events as integrated favors a form of modification which has both verbs in its scope. Thus, temporal adverbials should increasingly be placed outside the verbal complex. To illustrate, examples like (13a) should gradually give way to examples such as (13b).

- (13) a. *Stock satt en stund tyst och tänkte över vad Marstrand hade sagt.*
 Stock sat a while silent and thought over what Marstrand had said
 “Stock sat silent for a while and thought about what Marstrand had said.” (New novels)
- b. *Vi satt och pratade ett par timmar.* (New novels)
 we sat and talked a few hours
 “We sat and talked for a few hours.”

3.3 Prediction 3: Increase in the rate of object extraction

The third prediction also derives from the SPC’s increase in syntactic unity. In those cases where the second verb occurs with a direct or prepositional object, that object should become available for extraction, which is not generally possible in coordinated structures (cf. §2). As the SPC increases in syntactic cohesion, extraction should become increasingly likely, and hence more frequent over time. To illustrate, examples like (14a) should become more frequent relative to examples like (14b).

- (14) a. *Den där artikeln har jag suttit och läst hela dagen.* (Teleman et al. 1999)
 that there article have I sat and read all day
 “That article I have been reading all day.”
- b. *Jag har suttit och läst den där artikeln hela dagen.*
 I have sat and read that there article all day
 “I have sat and read that article all day.”

4. Methodology

4.1 Corpus data

We analysed four Swedish corpora, all of which are publicly accessible through the homepage of the Linguistics Department at the University of Gothenburg, Sweden (<http://spraakbanken.gu.se>, <http://www.ling.gu.se/projekt/tal> date of access:

03/02/2006). The first corpus is the *Källtext* corpus, a collection of Old Swedish texts of about one million words. These texts comprise different genres, containing for example the religious treatise *Saint Birgitta's revelations*, practical guidance literature such as *Peder Månssons art of farming*, and the five books of Moses. There is some uncertainty as to when some of the texts were composed exactly. It is, however, safe to say that most of them date to the period between 1300 and 1450. The second corpus, *Äldre svenska romaner*, consists of 57 novels from Swedish authors such as C.J.L. Almqvist, Victoria Benedictsson, and Hjalmar Söderberg. Taken together, they contain about 3.7 million words. The novels were originally published between 1839 and 1940. The third corpus, *Bonnier's romaner II*, is a collection of Swedish novels, published in 1980 and 1981. It contains 4 million words. The fourth corpus, the *Gothenburg Spoken Language Corpus*, is a spoken corpus. It contains 1.4 million words of transcribed speech recorded in the 1990s.

As will be clear from this description, the four corpora do not form a temporally continuous and internally homogeneous database. The represented time periods are not evenly spaced, and there are differences with regard to both genre and register. We cannot rule out the possibility that our results may be affected by these differences. The reason why we nonetheless base our analysis on these data is that, to our knowledge, no more suitable database is available at present. And rather than refraining from the use of these data altogether, we submit that the patterns we observe in the data have to be interpreted with the appropriate caution.

4.2 Corpus analysis

Our quantitative methodology requires that the use of *sitta* be analyzed in its entirety, rather than on the basis of selected examples. Therefore, the first step of our analysis was an exhaustive extraction of all forms of *sitta* from each corpus. This was done through searches with several wildcards in order to accommodate the orthographical variety that is typically found in older corpus data. For example, for the paradigm of *sitta* our data include forms such as *sato*, *sithiande*, *sithiä*, and many others. All instances were retrieved, manually inspected, and entered into a database together with the sentence it occurred in. This procedure yielded four concordances of *sitta*, one for each corpus.

Next, all instances of the verb *sitta* in each concordance were coded for several parameters. The first and most basic distinction was whether a given example potentially instantiates the SPC or not. Our operational criterion was whether *sitta* was used in some form of VP coordination. For example, cases analyzed as SPC exclude examples like (15a), but include presentative constructions with subject *det* “there”, as in (15b).

- (15) a. *Snälla Blodwen, berätta var dom sitter och hur dom ser ut.*
 dear Blodwen, tell where they sit and how they see out
 “Dear Blodwen, tell me where they sit and what they look like.”
- b. *Det satt en liten fågel vid min fot och kvittrade så vackert.*
 there sat a little bird at my foot and chirped so beautifully
 “A little bird sat at my foot and chirped so beautifully.”

Second, all instances of *sitta* were coded for the presence of modifiers falling into the category of locative elaboration (cf. §3.1). We counted as forms of locative elaboration any modification of *sitta* by means of an adverb or prepositional phrase specifying the location or manner of sitting. For instance, the prepositional phrase *vid min fot* “at my foot” in example (15b) elaborates the locative sense of the *sitta*.

Third, we took account of all instances in which SPC *sitta* is modified by temporal adverbials. Besides temporal adverbials such as *hela kvällen* “all evening long” we also included a small number of other adverbs, for example adverbs indicating psychological states such as *fundersam* “pensive” which are not clearly spatial in nature. We coded the placement of these adverbials in terms of two mutually exclusive environments: external to or between the two verbs, as in (16a) and (16b), respectively.

- (16) a. *Hon har suttit och sett på dig hela kvällen.* (Old novels)
 she has sat and looked at you all evening
 “She’s been looking at you all evening.”
- b. *Magistern satt fundersam och tittade på vägen.* (Old novels)
 magister sat pensive and looked at path.the
 “The teacher pensively looked along the path.”

Finally, we coded all instances of *sitta* in SPC contexts with regard to whether the second verb can take a direct or prepositional object, and if so, whether the object of the verb is in-situ or not. Example (17a) shows an extracted prepositional object, while the object of (17b) is in-situ.

- (17) a. *Men vet ni vad jag sitter och tänker på?* (Old novels)
 but know you what I sit and think about
 “But do you know what I am thinking about?”
- b. *Jag satt och tänkte på Brita.* (New novels)
 I sat and thought about Brita
 “I was thinking of Brita.”

5. Results and Discussion

Table 1 presents the absolute and relative frequencies of *sitta* inside and outside of potential SPC contexts for each of the four periods. The frequencies show that the use of *sitta* in these contexts has increased over time, from 18.4% of all instances in the Källtext corpus to about 40% in the later corpora. This frequency increase in itself suggests an underlying qualitative change.

Table 1. Token figures and relative frequencies of the SPC in four corpora of Swedish

	Källtext		Old Novels		New Novels		Gothenburg	
Size	1 M		3.7 M		4 M		1.4 M	
	n	%	n	%	n	%	n	%
SPC	96	18.4	2,053	38.6	2,138	41.5	278	40.1
non-SPC	426	81.5	3,269	61.4	3,012	58.5	416	59.9
Totals	522	100	5,322	100	5,150	100	694	100

A closer inspection of the attested use of *sitta* in the Källtext data suggests that an SPC-like construction already existed in the 14th century. An example is given in (18).

- (18) *Ther sato nokre kompana oc drukko oc lifdho i ofwerflødhlikheth.* (Källtext)
 there sat some friends and drank and lived in abundance
 “There sat some friends and drank and lived in abundance.”

Given the problems with identifying pseudo-coordination strictly on the basis of positive evidence, one might attempt to explain cases like (18) simply as canonical coordination. Indeed, all of the examples in the Källtext corpus express actions that are either conventionally carried out in a sitting posture, such as eating or drinking, or are at least fully compatible with a sitting posture, such as talking. However, there is qualitative evidence which suggests that the construction had already acquired a grammatical function at this stage. Consider (19), which describes two temporally intersecting events.

- (19) *Tha han kom til qwinnan oc sat och taladhe med henne, tha frestade diaefwillin honom ...* (Källtext)
 when he came to woman and sat and talked with her then tempted devil him
 “When he came to the woman and sat and talked to her, the devil tempted him...”

In this example, the first event (talking) provides the temporal background for the second event (the devil’s temptation). This use of *sitta* in (19) is characteristic of the discourse function of progressive and other imperfective aspect constructions (Comrie 1976: 33). While the backgrounded event is given an imperfective

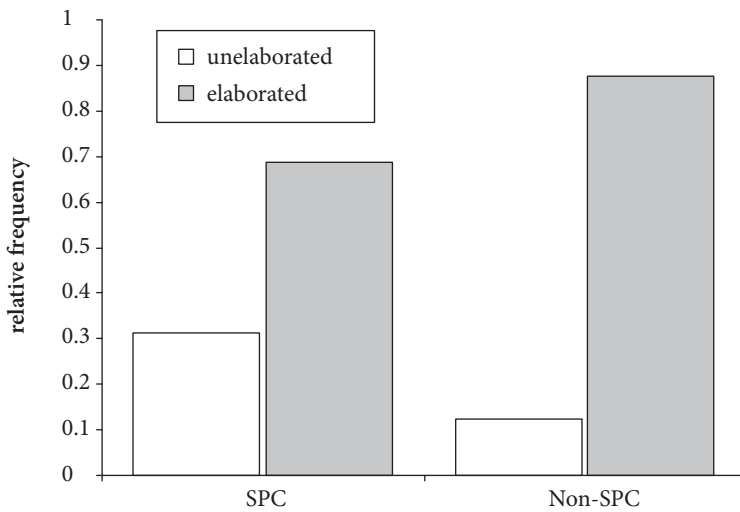


Figure 1. SPC/non-SPC and elaboration/non-elaboration in the *Källtext* corpus

construal by means of *sitta*, the foregrounded event takes perfective marking. Examples like (19), then, suggest that the grammaticalization of the SPC was already underway at the earliest period.

Besides these qualitative observations, there is also quantitative evidence showing that the SPC already existed as a grammaticalized construction at this state. Figure 1 shows the overall frequency of locative elaboration of *sitta* inside or outside of SPC contexts (for the absolute frequencies, please see the Appendix).

There is a statistically significant preference for zero-elaboration when *sitta* occurs in an SPC context ($df = 1$, $\chi^2 = 20.73$, $p < .001$, absolute frequencies are given in the Appendix). In other words, there is a measurable preference for *sitta* to be used in its “bare” form when in coordination with another verb. This asymmetry in the argument structure can only be captured by saying that *sitta* was used in its unelaborated form *specifically* when used together with another verb. And this is tantamount to saying that a *sitta*-based grammatical construction existed.

If *sitta* already formed part of a complex predicate at this early stage, the next question is whether there is evidence of further grammaticalization. Any additional development in the same direction should be reflected in an increase in the asymmetry in Figure 1. The difference between the ratio of non-elaborated and elaborated instances of *sitta* in the SPC and the same ratio outside of the SPC should become more dramatic. Figure 2 presents these relative frequencies (for the absolute frequencies, please see the Appendix).

The shaded bars show that when *sitta* is used as a lexical verb it occurs only about 10% of the time without some form of locative elaboration. This rate is

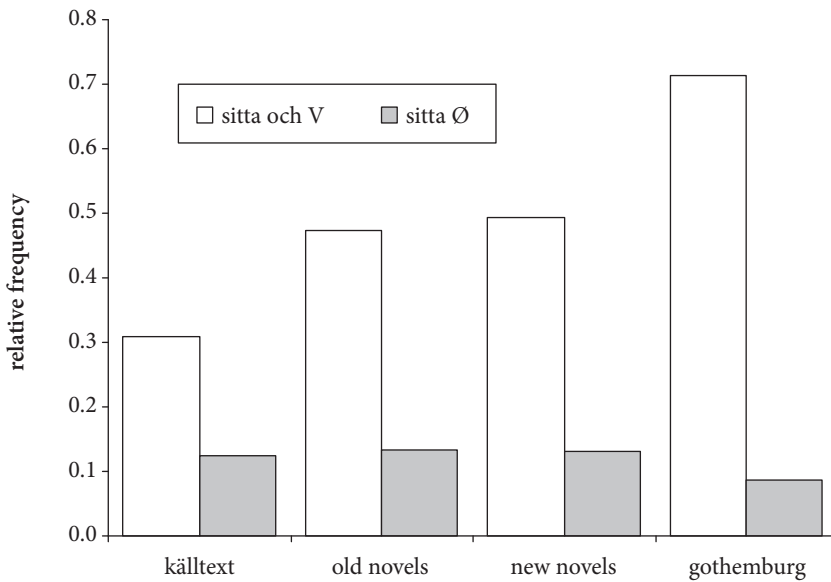


Figure 2. Relative frequency of zero-elaboration of *sitta* inside and outside of the SPC

relatively constant across the four corpora.² Indeed, we would not expect any changes here because the argument structure of lexical *sitta* and remained unchanged. By contrast, the white bars indicate a steady increase in “bare” uses of *sitta* in the SPC. These changing preferences suggest that there was a change in argument structure of *sitta* that is specific to its function within the SPC. As predicted, in the course of continuing grammaticalization, there is an increasing tendency to use *sitta* in its most compact form, which backgrounds its locative meaning by not further specifying the location or manner of sitting.

The differences between the four corpora are statistically significant overall ($df=3$, $\chi^2=71.66$, $p<.001$). In keeping with our earlier observation that the data need to be interpreted cautiously, we need to point out that time is only one parameter by which the four corpora differ. Still, the fact that significant differences exist is suggestive of a gradual change.

We now turn to our second prediction, viz. that adverbials should increasingly be used to modify the entire verbal complex and therefore occur before or after the two verbs, rather than between them. As not all instances of the SPC contain adverbial modification, this parameter can only be applied to those examples which actually are adverbially modified (about 15% of the data). Again, the hypothesis is borne out. Figure 3 shows that the relative frequency of adverbials placed between

2. This suggests that the four corpora are, indeed, comparable in this respect.

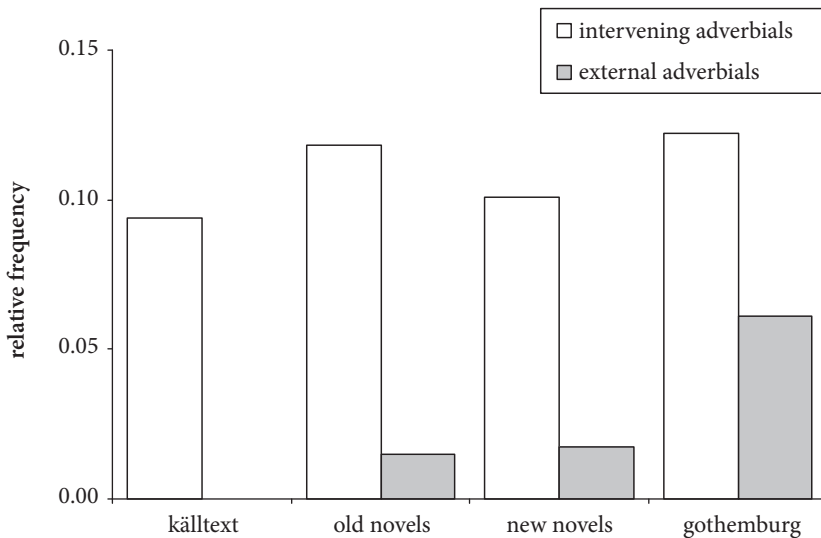


Figure 3. Internal and external adverbials in the SPC

the two verbs and adverbials external to the verbal complex changed in the expected direction (for the absolute frequencies, please see the Appendix).

The asymmetry in the distribution of intervening and external adverbials is statistically significant across the four corpora, ($df = 3$, $\chi^2 = 18.35$, $p < .001$).

Our third prediction was that an increase in the syntactic unity of the SPC should be reflected in a greater rate of extracted objects from the second, lexical verb. Again, only a subset of the data is relevant, as not all verbs take objects. About 40% of all retrieved examples contain a verb that takes either a direct or prepositional object. In cases where the second verb is transitive, the object can be extracted for the purpose of topicalization, or in order to function as the head of a relative clause. Over time, we expect speakers to make more use of this possibility. This prediction is borne out, as shown by Figure 4 (as previously, all absolute numbers are listed in the Appendix).

There is a significant difference in the relative frequency of extraction phenomena across the four corpora ($df = 3$, $\chi^2 = 33.1$, $p < .001$). As we expected, extraction is very uncommon in the older data. In the oldest period, thirty-one examples occur with object-taking verbs, but none of these are extracted. However, extractions are attested at the second stage, as shown in (20).

- (20) *Det var kanske den bibeln husmodern satt och läste i, då de kom ...*
 it was maybe that bible house.mother.the sat and read in when they came
 “Maybe it was that bible that mother sat and read when they came ...”



Figure 4. Increase of extracted objects of V2 in the SPC

It needs to be pointed out that even in synchronic data the relative frequency of extracted objects is still not very high overall. Only in about 7% of the possible cases does extraction occur.

6. Conclusion

The observed tendencies allow the following conclusions. The Swedish Pseudo-coordination construction with *sitta* has, even after it established itself as a grammatical construction, undergone a number of gradual changes. As part of the SPC, the verb *sitta* has changed its argument structure, and the construction showed an increase in syntactic cohesion.

The present paper also bears on a disputed point brought up in the study of light verbs. Butt & Lahiri (1998) compare complex predicates from Sanskrit, Old Bengali, Old Hindi, and Vedic to corresponding constructions in present-day Indo-Aryan languages. They find that their structure has remained essentially the same in spite of three millennia of language change. Butt & Lahiri conclude that auxiliaries are more prone to change than light verbs. While the process of grammaticalization turns lexical verbs into auxiliaries through semantic bleaching and morphosyntactic reduction, which eventually leads to affixation and complete deletion, light verbs are claimed to be historically stable (1998: 51). Butt & Lahiri explicitly reject proposals by Hook (1991, 1993), who proposes that light verbs underlie the same principles as other grammaticalizing elements.

The present analysis covers a far shorter period of time than the study by Butt & Lahiri, but yields tentative counterevidence against their claim. It has been shown that *sitta* displays characteristics of a light verb, as it takes on a type of aspectual meaning and clearly does not undergo auxiliation. Contra Butt & Lahiri, the data suggest that *sitta* has not been historically stable, but was instead subject to gradual change. So far this change has only affected the argument structure of *sitta* as part of the SPC; we do not know whether it will ultimately lead to morphophonemic reduction. However, the development so far does not suggest that light verbs behave any differently than auxiliaries, a conclusion also reached by Hook & Pardeshi (2006).

As a final conclusion, we find that the traditional approach of studying grammaticalization through crosslinguistic comparisons and grammaticality judgments can be usefully supplemented by an approach based on quantitative data whenever this is possible. While many languages have no written record of great time depth, the languages for which such data are available provide opportunities to test hypotheses that not only bear on language-specific questions but also on general principles of language change. Ultimately, the richest accounts of grammatical phenomena will be those that are based on different types of evidence.

References

- Aikhenvald, Alexandra. 2006. "Serial verb constructions in a typological perspective." *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Aikhenvald & R.M.W. Dixon, 1–87. Oxford: Oxford University Press.
- Alsina, Alex, Joan Bresnan, & Peter Sells, eds. 1997. *Complex Predicates*. Stanford: CSLI.
- Alsina, Alex. 1997. "A Theory of Complex Predicates: Evidence from causatives in Bantu and Romance." *Complex Predicates*, ed. by Alex Alsina, Joan Bresnan, & Peter Sells, 203–246. Stanford: CSLI.
- Austin, Peter. 1998. "Crow Is Sitting Chasing Them: Grammaticalisation and the verb "To Sit" in the Mantharta Languages, Western Australia." *Case, Typology and Grammar: In honour of Barry J. Blake*, ed. by Anna Siewierska, & Jae Jung Song, 19–36. Amsterdam: John Benjamins.
- Barlow, Michael & Suzanne Kemmer, eds. 2000. *Usage-based models of language*. Stanford: CSLI.
- Bergs, Alexander. 2005. *Social Networks and Historical Sociolinguistics. Studies in Morphosyntactic Variation in the Paston Letters (1421–1503)*. Berlin: Mouton de Gruyter.
- Bowern, Claire. This volume. "The Diachrony of Complex Predicates."
- Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI.
- Butt, Miriam & Aditi Lahiri. 1998. "Historical stability vs. historical change." Ms. Konstanz: University of Konstanz.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

- Bybee, Joan L., Revere D. Perkins, & William Pagliuca. 1994. *The Evolution of Grammar: Tense, aspect and mood in the languages of the world*. Chicago: University of Chicago Press.
- Bybee, Joan L. & Paul Hopper, eds. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Di Sciullo, Anna Maria & Sara Thomas Rosen. 1990. "Light and semi-light verb constructions." *Grammatical Relations: A cross-theoretical perspective*, ed. by Katarzyna Dziwirek, Patrick Farrell, & Errapel Mejías-Bikandi, 109–125. Stanford: CSLI.
- Foley, William & Mike Olson. 1985. "Clausehood and verb serialization." *Grammar Inside and Outside the Clause*, ed. by Johanna Nichols & Anthony Woodbury, 17–59. Cambridge: Cambridge University Press.
- Gahl, Susanne & Susan Garnsey. 2004. "Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation." *Language* 80:4.748–775.
- Givón, Talmy. 2006. "Multiple routes to syntactic complexity: The diachronic foundations of (any) synchronic typology." Paper presented at the 11th Biennial Rice Linguistics Symposium: Intertheoretical Approaches to Complex Verb Constructions.
- Harris, Alice. This volume. "Light verbs as classifiers in Udi."
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Hook, Peter E. 1991. "The emergence of perfective aspect in Indo-Aryan Languages." *Approaches to Grammaticalization*, ed. by Bernd Heine & Elizabeth C. Traugott, 59–89. Amsterdam: Benjamins.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Hook, Peter E. 1993. "Aspectogenesis and the Compound Verb in Indo-Aryan." *Complex Predicates in South Asian Languages*, ed. by Manindra Verma, 97–113. Delhi: Manohar.
- Hook, Peter & Prashant Pardeshi. 2006. "Are vector verbs eternal?" Paper presented at the 11th Biennial Rice Linguistics Symposium: Intertheoretical Approaches to Complex Verb Constructions.
- Hooper, Joan B. 1976. "Word frequency in lexical diffusion and the source of morphophonological change." *Current Progress in Historical Linguistics*, ed. by William Christie, 96–105. Amsterdam: North Holland.
- Hopper, Paul J., & Elizabeth C. Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.
- Koops, Christian. 2004. "Emergent aspect constructions in Present-Day English." *Studies in Linguistic Motivation*, ed. by Günter Radden & Klaus-Uwe Panther, 121–154. Berlin: Mouton.
- Kuteva, Tania. 1999. "On sit/stand/lie auxiliation." *Linguistics* 37:2.191–213.
- Kuteva, Tania. 2001. *Auxiliation — An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University Press.
- Lightfoot, David W. 1991. *How to set Parameters: Evidence from Language Change*. Cambridge, Mass.: MIT Press.
- Lødrup, Helge. 2002. "The syntactic structures of Norwegian pseudocoordinations." *Studia Linguistica* 56:2.121–143.
- Svorou, Soteria. 1994. *The Grammar of Space*. Amsterdam: Benjamins.
- Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik*. Stockholm: Norstedts.

Wiklund, Anna-Lena. 1996. "Pseudocoordination is subordination." *Working Papers in Scandinavian Syntax* 56.29–54.

Résumé

Ce travail trace le développement historique de la construction suédoise de pseudo coordination avec le verbe de position *sitta* 'être assis'. En suédois, un nombre restreint de verbes qui inclut les verbes de position tels *sitta* sont utilisés en coordination avec un autre verbe pour indiquer que l'événement décrit est en train de se dérouler, ou dure de façon particulièrement longue. L'analyse quantitative de données historiques suggère que la construction a continué à subir une série de modifications graduelles pendant cinq cents ans, soit longtemps après qu'elle a été établie comme une construction grammaticale. Sa participation à la construction a contraint le verbe *sitta* à modifier la structure de ses arguments, et la construction a gagné en cohésion syntaxique.

Zusammenfassung

Die vorliegende Studie behandelt die historische Entwicklung der sog. Pseudo-Koordination im Schwedischen. Pseudo-koordinierende Konstruktionen erlauben es, bestimmte Verben, darunter *sitta*, mit einem anderen Verb zu kombinieren um eine Situation als zeitlich ausgedehnt oder im Verlauf begriffen zu beschreiben. Durch eine quantitative Erhebung dieser Konstruktion basierend auf mehreren historischen Korpora wird gezeigt, dass die Pseudo-Koordination mit *sitta* bereits im 14. Jhd eine grammatikalisierte Konstruktion darstellte, die darüber hinaus in den folgenden Jahrhunderten eine Reihe weiterer, gradueller Veränderungen erfahren hat. Das Verb *sitta* veränderte als Teil dieser grammatischen Konstruktion seine Argumentstruktur, und die syntaktische Kohäsion der Konstruktion insgesamt nahm zu.

Authors' addresses

Martin Hilpert
Freiburg Institute for Advanced Studies
(FRIAS)
Albertstr. 19
79104 FREIBURG
Germany

Christian Koops
Rice University
6100 Main Street
HOUSTON, TX 77005
ckoops@rice.edu

martin.hilpert@friass.uni-freiburg.de

Appendix

Table 2. Elaborated and non-elaborated instances of *sitta* in the SPC and elsewhere

	SPC		Lexical <i>sitta</i>	
	unelaborated	elaborated	unelaborated	elaborated
Källtext	30	66	53	373
Old Novels	879	1,174	440	2,829
New Novels	919	1,219	398	2,614
Gothenburg	198	80	36	380

Table 3. Intervening and external adverbials in the SPC

	Intervening Adverbials	External Adverbials	All SPC Examples
Källtext	9	0	96
Old Novels	243	31	2,053
New Novels	216	37	2,138
Gothenburg	34	17	278

Table 4. Extracted and in-situ objects of V2 in the SPC

	Extracted	In-Situ
Källtext	0	31
Old Novels	7	987
New Novels	12	996
Gothenburg	8	105