Nor Jnl Ling 33.2, 105–143 © Nordic Association of Linguists 2010 doi:10.1017/S033258651000017X

Bohnacker, Ute. 2010. The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics. In Ute Bohnacker & Marit Westergaard (eds.), *The Nordic Languages and Second Language Acquisition Theory*, special issue of *Nordic Journal of Linguistics* 33(2), 105–143.

The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics

Ute Bohnacker

In a recent study of the clause-initial position in verb-second declaratives (the prefield), Bohnacker & Rosén (2008) found significant differences between native Swedish and German concerning the frequencies with which constituents occurred in the prefield, as well as qualitative differences concerning the mapping of information structure and linear word order: Swedish exhibited a stronger tendency than German to place new information, the so-called rheme, later in the clause. Swedish-speaking learners of German transferred these patterns from their L1 to German. Their sentences were syntactically well-formed but had Swedish-style prefield frequencies and a strong pattern of Rheme Later, which native Germans perceive as unidiomatic, as an acceptability judgment and a rewrite-L2texts task showed. The present study extends Bohnacker & Rosén's work in three ways. Learners of the reverse language combination (L1 German, L2 Swedish) are investigated to see whether similar phenomena also manifest themselves there. Secondly, written and oral data from highly advanced learners are examined to see whether the learners' persistent problems can be overcome by extensive immersion (3, 6 and 9 years of L2 exposure). Thirdly, besides investigating theme–rheme (old vs. new information), some consideration is given to another information-structural level, background vs. focus. The learners are found to overuse the prefield at first, with non-Swedish, German-style frequency patterns (e.g. low proportions of clause-initial expletives and high proportions of clause-initial rhematic elements). This is interpreted as evidence for L1 transfer of information-structural or discourse-pragmatic preferences. After 6 and 9 years, a substantial increase in clause-initial expletive subjects, clefts and lightweight given elements is indicative of development towards the target. The findings are related to current generative theorizing on the syntax-pragmatics interface, where it is often maintained that the integration of multiple types of information is one of the hardest areas for L2 learners to master.

Keywords clause-initial, cleft, discourse pragmatics, expletive, German L1, information structure, prefield, Swedish L2, syntax–pragmatics interface, theme–rheme, V2 declarative

Ute Bohnacker, Department of Linguistics and Philology, Uppsala University, Box 635, SE-751 26 Uppsala, Sweden. ute.bohnacker@lingfil.uu.se

1. INTRODUCTION

This paper investigates the interaction of word order and discourse-pragmatic constraints concerning the clause-initial position of declaratives in the advanced second language (L2) Swedish of native German speakers.

Advanced learners are often found to master the syntactic constraints of their L2 but to subtly diverge from native speakers in the ways they employ these syntactic means to structure oral and written text. This has been documented for L2 oral narratives, descriptions and written texts by a number of functionalist-oriented researchers, concerning, for example, referent introduction and maintenance, narrative perspective and textual cohesion (e.g. Carroll & von Stutterheim 1993, 2003; Mauranen 1996; Carroll & Lambert 2003). It has even been suggested that while L2 learners may acquire the language-specific linguistic means that have an impact on discourse pragmatics, their role in information organisation and information structuring will never become fully targetlike: They retain core principles and patterns of their native language (Carroll & Lambert 2003; Carroll & von Stutterheim 2003:372, 394–398; von Stutterheim 2003:202).

In the generative language acquisition literature there is also a growing body of evidence that L2 learners master pure syntax before they are able to put that syntax to appropriate discursive use (see White 2009). For instance, learners of a null-subject language such as Italian or Spanish start to produce and accept null subjects relatively early on and master intricate syntactic constraints to do with the null-subject property, even though null subjects are ruled out in their L1. On closer inspection, however, the learners' discourse distribution of null vs. overt subject pronouns is not fully appropriate (e.g. Pérez-Leroux & Glass 1999; Sorace & Filiaci 2006; Rothman 2009). The contextually appropriate distribution of null and overt subjects is regulated by discourse-pragmatic constraints, located at the interface of syntax and discourse/pragmatics according to current theorising. Discourse-pragmatic constraints also feature in recent studies on the use of postverbal subjects by L2 learners. For instance, Italian and Spanish exhibit Subject–Verb (SV) and Verb–Subject (VS) word orders depending on verb type (unaccusative/unergative), but VS word order is preferred by native speakers for presentationally focused subjects irrespective of verb type. L2 learners of Italian and Spanish acquire VS word order relatively easily but have protracted problems with VS in focused contexts: Unlike native speakers, L2 learners fail to produce VS in focused contexts or accept both VS and SV in equal proportions (e.g. Hertel 2003; Lozano 2006; Belletti, Bennati & Sorace 2007). A third example of inappropriate discourse patterns in L2 learners is word order variation concerning the clause-initial position, the so-called prefield, in Germanic Verb Second (V2) languages: Swedish-speaking learners of German produce syntactically well-formed V2 declaratives in their L2 and L3 German with a variety of constituents in clause-initial position, but organise information in these clauses in ways that diverge from native German, with Swedish-style frequencies for prefield constituent types and a strong pattern of new information being realised postverbally (Bohnacker 2005, 2006; Rosén 2006; Bohnacker & Rosén 2007, 2008). Bohnacker & Rosén attribute these findings to transfer of information-structural

patterns from the L1. Again, word order alternations that are influenced by discourse-pragmatic factors seem hard to acquire.

Work within the Chomskyan Minimalist program has suggested that the reason for such non-convergence might be that interfaces are involved (see e.g. Sorace 2005). Interfaces, a term adopted from chemistry and electronics, are currently in vogue in both grammatical theory and acquisition research. Interfaces are points at which different components or modules of a system interact with each other or with other, external systems. For language this means that different components of the linguistic computational system interact with each other or with other cognitive systems. In generative frameworks that revolve around grammar, this translates into the well-known diagram in (1), where Logical Form (LF) and Phonetic Form (PF) are the outward-looking, grammar-external, interfaces to the conceptual-intentional and articulatory-perceptual systems (see Chomsky 1993; White 2009). Not featured in this diagram but also assumed are a number of grammar-internal modules (phonology, morphology, syntax, semantics and the lexicon) and inward-looking, grammar-internal interfaces between these modules.

(1)

Conceptual-intentional system	LF	Grammar (computational system & lexicon)	PF	Articulatory-perceptual system

L2 learners may struggle at these interfaces when having to integrate different kinds of grammatical knowledge, such as phonology with morphology, morphology with syntax, syntax with the lexicon, or syntax with discourse. However, there is no a priori reason to assume that all interfaces are equally problematic – some interfaces may continue to cause problems for L2 learners at advanced proficiency levels, others may not.

Grammar-external interfaces are sometimes assumed to be more problematic for learners than grammar-internal ones, though this is a matter of dispute, as pointed out by White (2009).[1] In particular, it has been claimed that the (external) interface between syntax and other cognitive systems, notably discourse pragmatics, is more difficult than other interfaces and that it is here where interlanguages exhibit optionality, instability and residual L1 effects for the longest (e.g. Sorace 2005; Sorace & Filiaci 2006; Tsimpli & Sorace 2006; Valenzuela 2006; Belletti et al. 2007; Sorace & Serratrice 2009; see also Hulk & Müller 2000 and Platzack 2001). Congruent with this claim are the results from some of the above-mentioned empirical studies on the acquisition of discourse-driven distribution of null subjects, focused subjects and clause-initial word order variation. However, finding optionality and residual L1 effects in the discourse–syntax domain is not proof that this is the area hardest to master, unless other linguistic domains (say, inflectional morphology or

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

vocabulary) are also investigated for the same learner(s), which is rarely done. It is also an open question whether L2 discourse-pragmatic problems are surmountable at advanced proficiency levels, or whether they are inevitable and prone to fossilising.[2] Concerning the topic of the present paper, namely word order variation in the clause-initial position of L2 learners of a Germanic V2 language, we have some indication of discourse-pragmatic problems in advanced learners, but we do not know whether delays in this domain can eventually be overcome. This is because so far only tutored foreign language learners have been studied who had up to six years of classroom training in their home country (Rosén 2006; Bohnacker & Rosén 2007, 2008). The learners had not spent any extended period of time in a naturalistic L2 environment. One therefore wonders whether the observed transfer of L1 information-structural patterns would have persisted if the learners had been exposed to larger quantities (and possibly higher quality) of input in the target language. The present study of advanced native German learners of Swedish who have been immersed in the target language for three, six and nine years may be a first step towards answering this question.

This paper is set up in the following manner: Section 2 discusses different approaches to the interaction of word order and discourse pragmatics. Section 3 extends and applies this discussion to the prefield in Swedish and German. Section 4 provides background information on informants and data collection. In Section 5, quantitative and qualitative results are provided on how German-speaking learners of Swedish after three, six and nine years of exposure vis-à-vis native speakers make use of the prefield, concerning subjects and expletives as well as objects. Section 6 offers concluding remarks.

2. SYNTAX, DISCOURSE PRAGMATICS AND INFORMATION STRUCTURE

The syntax of a language is commonly described as encompassing a set of rules, parameters or constraints on which constituent orderings are possible irrespective of context. In a particular context, certain constituent orderings (e.g. preverbal and postverbal subjects) may be more likely or more felicitous than others. This variation is typically not ascribed to pure syntax, but to discourse-pragmatic and semantic factors. Discourse pragmatics covers many phenomena including politeness marking and language choice in multilingual contexts, but for present purposes another area of discourse pragmatics is more relevant, namely how speakers/writers of a particular language organise and present information. Such information management can be studied at a global or text level (e.g. Halliday & Hasan 1976; von Stutterheim & Klein 1989; Tomlin et al. 1997:66–77) and at a local level, i.e. that of the utterance or the clause. I will be considering mainly this latter, local level here – with

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

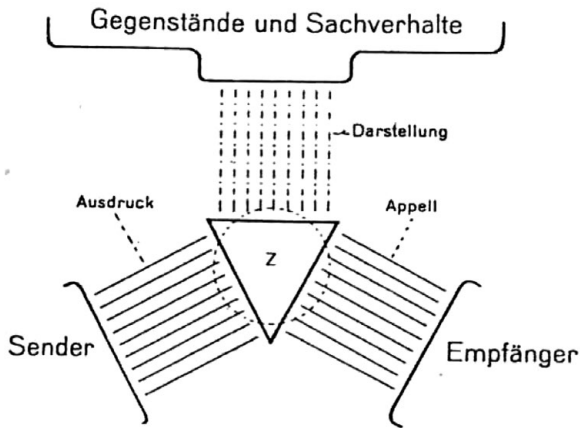

Figure 1. Bühler's communicative triangle (Bühler 1934:28).

its potentially universal or language-specific information-structural influence on constituent ordering.

Information structure concerns the division of information into more or less salient or relevant and its packaging and presentation with the help of linguistic structure (e.g. Chafe 1976; Prince 1981). The information structure of a particular utterance very much depends on the context that utterance occurs in. Speakers/writers structure utterances according to what they regard as the main point of the utterance and what is being said about what. They also make assumptions about what listeners/readers know and what they are thinking about. They linguistically encode the degree to which they regard the content of what they say to be accessible to the listener, and structure and present information accordingly.

Research on information structure is characterised by a bewildering heterogeneity in terminology. Perhaps the most widely used notions are topic and focus, where topic is often defined in terms of aboutness or as information already presented in the context, and focus is often defined in terms of new or prominent information. However, many scholars do not offset topic against focus but regard the two as a conflation of different levels, necessitating other notions beyond topic and focus at several distinct levels of information structure (e.g. Krifka 2007:41). I share this view and will assume that there are three separate levels of information structure. Here it is worth reminding ourselves that the way we view information structure depends on our model of communication. A widely known model is Bühler's (1934:28) communicative triangle in Figure 1, where content (*Gegenstände und Sachverhalte*) is communicated by the sender (*Sender*) to the receiver (*Empfänger*).

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

One might say that the communicative effect of information structure is to foreground certain aspects of the message and to background others, IN THE CONTEXT OF PREVIOUS DISCOURSE. If we are mainly interested in one of the three sides of the communicative triangle, our theory of information structure will be founded on this particular perspective, perhaps resulting in a single level model, for instance one where topic is offset against focus (e.g. Erteschik-Shir 2007). If we are interested in all three sides of the triangle, our theory of information structure will reflect this and presumably include several separate information-structural levels. This in turn may influence how researchers define particular information-structural notions. My aim here is not to evaluate or compare different models or to propose my own, but simply to put into context the information-structural terms that I will be using throughout the paper. Readers are welcome to substitute their own favourite terms.

(2) *Three information-structural levels/perspectives*
   a. Topic vs. comment
   b. Background vs. focus
   c. Theme vs. rheme

Bühler's triangle yields three separate information-structural levels, here called (2a) topic vs. comment, (2b) background vs. focus, and (2c) theme vs. rheme. At the level of content (*Gegenstände und Sachverhalte*), information is typically divided into topic and comment. Here, the topic of a sentence is understood to be the point of departure and to provide the referential frame with respect to which the predication is evaluated. Information ABOUT the topic is encoded in the comment (see Reinhart 1982 and her well-known metaphor of topic as a file card or entry in a library catalogue; Hockett 1958:201; Lambrecht 1994; Tomlin et al. 1997; Krifka 2007).

At another level, that of the *Sender*, the speaker considers which referent(s) and proposition(s) s/he wants to bring to particular attention. Information can be divided into background and focus, depending on the speaker's choice of demoting some information (background) and highlighting some other information (focus) that s/he regards as the most relevant part of the utterance; moreover, focus can also extend over the entire utterance (e.g. Jacobs 1984; Krifka 2007). What the speaker regards as the most relevant part of the utterance might also be new information for the listener, so that focus coincides with rheme. However, focus can also occur independently of the theme–rheme (given–new) status of constituents, e.g. when a given referent (theme) is placed in a new, unpredictable or not yet settled constellation.[3]

Finally, at the level of the listener (*Empfänger*), the speaker monitors which referents and propositions seem to be available to the listener and which ones require introduction or re-introduction. Information is structured into given and new, or theme and rheme in the terminology of Ammann (1928:2–3), who coined these terms, and representatives of the Prague School, who developed and disseminated them (e.g. Daneš 1970; Beneš 1971; see also Ekerot 1979; Prince 1981). THEME here stands

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

for what the speaker/writer assumes the listener/reader to know; it is old, maintained or given information in the sense that it has previously been explicitly mentioned or is inferable with recourse to the linguistic discourse or the discourse situation. RHEME stands for what the speaker assumes to be new information for the hearer. Dividing the clause into given and new is not always straightforward, as clauses may contain several given elements, and some contain none but are informationally all-new. Thematicity/givenness may also be viewed as a graded property, where recency of mention and other factors influence how accessible a thematic/given element is.[4]

A pragmatic tripartite approach to information structure such as the one in (2) allows for the different information-structural levels to be kept apart, but also for them to correlate, and this they often do. For instance, focused information is frequently encoded by the same expression as new information and tends to occur in the part of the utterance that contains the comment. Also, topic (point of departure) and theme (old information) frequently coincide in an utterance. However, not all topics are old information, and not all old information is a topic. Likewise, what is focused in an utterance often is, but need not be, new information, since focus and rheme relate to different information-structural levels. Importantly, then, the levels are not isomorphic. In Section 3, these notions are related to the German and Swedish prefield, but first I will make some general comments on the relation (and the interface) between syntax and discourse pragmatics.

How to view and formalise the relation between syntactic form and discourse function is much debated, and my understanding of the literature leads me to distinguish three major lines of approach. One is the in essence functionalist view that the grammatical form directly follows from the communicative function of a sentence, in short that there is iconicity (see Kuno 1987).

Another line of approach is the 'traditional' generative view that syntax is autonomous and discourse function is external to syntax (e.g. Chomsky 1965; Prince 1981, 1998; Chomsky in Stemmer 1999:400; Jackendoff 2002; Fanselow 2007; Féry 2007). Prince (1998:281) puts it as follows: '[T]he relation between syntactic form and discourse function is no less arbitrary than, say, the relation between phonological form and lexical meaning'. Structural possibilities are provided by the grammar INDEPENDENTLY of discourse pragmatics, and discourse-pragmatic notions do not play a role in the identification of syntactic slots or categories, nor in the triggering of syntactic operations. A multitude of grammatical devices (phonological, morphological and lexical markers, syntactic structures and surface positions) may be employed to support different discourse functions. However, proponents of the autonomous-syntax view point out that there is great cross-linguistic variation in this regard and that particular discourse functions do not invariably correlate with any grammatical reflex (e.g. Prince 1998; Féry 2007). Any mapping between language-specific form and pragmatic function can thus only be indirect and takes place not in

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

syntax but in separate cognitive components (Lambrecht 1994; Neeleman & van de Koot 2008).[5]

This view contrasts with a third approach, where syntax is assumed to encode discourse functions. Information-structural notions such as topic, focus and givenness are incorporated into formal theories of syntax via an articulated hierarchy of functional projections and corresponding 'syntactic' features. Topic and focus are regarded as morphological notions in some abstract sense, attracting relevant syntactic constituents to a specifier in order to check a morphological feature (e.g. Brody 1990; Rizzi 1997; É. Kiss 1998; Belletti 2004). Particular discourse-related interpretations are licensed in the specifier of a designated functional projection – in other words, 'syntactic positions – ultimately word order – directly affect aspects of the interpretation, which can thus be read off the syntactic configuration' (Belletti et al. 2007:659). This 'cartographic' approach may be attractive when there are discrete morphological markers for particular discourse functions but becomes rather abstract for languages that do not have such morphological markers or any fixed designated information-structural slots. While cartographic approaches are gaining in popularity, proponents do not agree on the details of this proliferated phrase structure. As Benincà & Poletto (2004:52) put it, 'there is no limit, in our view, as to how many of these projections there will ultimately be'. This may raise questions about learnability and concerns that formal theories of syntax try to account for phenomena that would better be handled in semantic, pragmatic or processing terms (Polinsky & Kluender 2007:277).

There is thus little agreement on whether and how much discourse pragmatics should be represented in the syntax, and I am reluctant to take a stand on this matter here. However, I would like to point out that depending on which line of approach is chosen, the locus or type of the interface in one's model of language knowledge may change. In the generative tradition, linguistic competence is mentally represented by means of an abstract linguistic system, the grammar. In this grammar, different components or modules interface with each other grammar-internally, and they also interface with other, grammar-external domains such as the conceptual-intentional system. A 'discourse-free' syntax approach thus necessarily involves an external interface with an interpretive module. A 'discourse-laden' syntax approach (as in the cartographic models), on the other hand, strives to treat discourse-pragmatic notions essentially as syntactic and as part of the computational system of the grammar. But if they are part of the computational system, this suggests a grammar-INTERNAL interface, notwithstanding the existence of a grammar-external interface with an interpretive module.

Recent L2 research points fairly consistently to learner problems associated with phenomena that involve a relationship between syntax and discourse pragmatics (e.g. Hertel 2003; Lozano 2006; Tsimpli & Sorace 2006; Belletti et al. 2007; Rothman 2009). In the past, generative linguists have often relegated these problems to domains

outside the grammar proper, e.g. to pragmatics, rhetoric or stylistics (Liceras 1988). Nowadays, they tend to be described as syntax–pragmatics or syntax–discourse interface problems. Alternatively, they are argued to arise within the computational system itself – as representational deficits in functional categories or in features associated with these categories (e.g. Valenzuela 2006; Belletti et al. 2007:659, 676), but to do so makes sense only within a cartographic theory that assumes discourse-pragmatic features to exist in the syntax. On an autonomous-syntax view no such features are part of the syntax. So it seems to me that there is no theory-neutral answer to the question whether L2 learners have greater problems at grammar-external interfaces than at grammar-internal interfaces, and whether the problems are pragmatic or grammatical in nature, since the answer very much depends on the formal theory of syntax adopted.

3. THE SWEDISH AND GERMAN PREFIELD AT THE INTERFACE OF SYNTAX AND DISCOURSE PRAGMATICS

3.1 Quantitative differences and similarities concerning constituents

Both Swedish and German adhere to the verb-second (V2) constraint that requires the finite verb in declaratives to be the second constituent. In main clauses that are not subject-initial, inversion of the subject and the verb is required, and V3 is generally ungrammatical. The position to the left of the finite verb is called the PREFIELD (e.g. Drach 1937:17–18; Reis 1980), see the examples in (3). In principle, in German and Swedish, the prefield can contain virtually any constituent, irrespective of syntactic category, complexity and semantic function, some modal particles excluded (e.g. Zifonun et al. 1997:1576–1644; Teleman, Hellberg & Andersson 1999:431–434, 689–690). Word orders seem to be interchangeable without any obvious difference in grammaticality or meaning.

(3)

	PREFIELD (Spec CP)	FINITE VERB (C)	MIDFIELD AND REMAINDER OF DECLARATIVE CLAUSE (IP/VP domain)					
a.	Nu	har	väl	alla	fått	ett	ex.	(Swedish)
	now	*have*	*well*	*all*	*got*	*a*	*copy*	
b.	Jetzt	haben	ja	wohl	alle	eins.		(German)
	now	*have*	*yes*	*well*	*all*	*one*		

'I suppose everyone's got a copy now.'

Much of the traditional work on Swedish and German and generative transformational approaches to clause structure share the assumption that the prefield of declarative main clauses is filled by 'fronting' an element from the midfield, in a secondary step so to speak after the midfield has been generated (e.g. Grewendorf 1988:64–67;

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

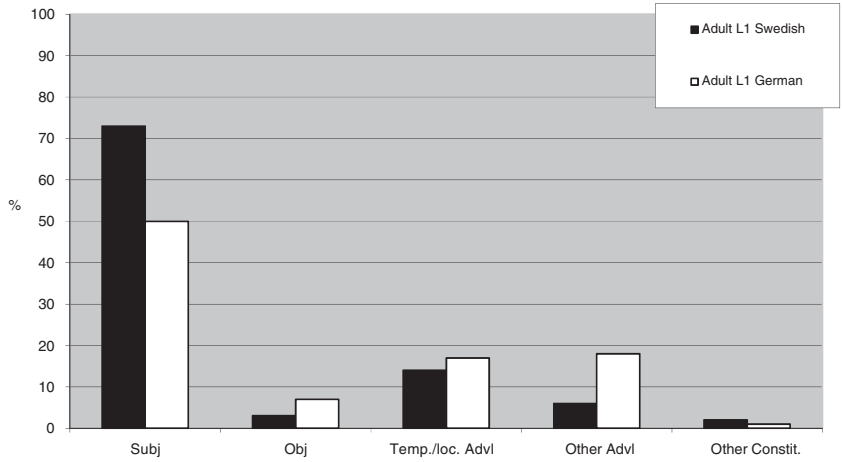

Figure 2. Overt constituents in the prefield, written L1 data (Bohnacker & Rosén 2008).

Teleman et al. 1999:406). As an alternative to fronting, expletives and elements not subcategorised for are commonly assumed to be base-generated in the prefield, but fronting of a constituent is the default. Which constituent is chosen to be fronted to the prefield is usually considered to be a matter of discourse factors, although lexical-semantic content and phonological weight have sometimes been suggested to play a (secondary) role.[6]

On the view that a description of the syntax of a language should aim to specify the possible strings, regardless of the likelihood that such strings will sound felicitous in a particular discourse context (e.g. Prince 1998), we can say that the syntactic constraints on the prefield and on how to start a V2 declarative clause in Swedish and German are the same. Not surprisingly, these two languages, along with other Germanic V2 languages, have been assumed to behave alike concerning the function and frequency of prefield constituents, with a distribution of 70% or 60% subject-initial vs. 30% or 40% non-subject-initial, though such figures are usually estimates, not based on corpora counts (e.g. Håkansson 1997:50). However, when Christina Rosén and I surveyed existing written text corpora, we found that V2 languages may differ substantially in the way they make use of the prefield, both quantitatively and qualitatively (Bohnacker & Rosén 2007:34–35). Subject-initial clauses were consistently more frequent in Swedish than in German. The informal written corpus data we collected ourselves confirmed this; as shown in Figure 2, Swedish has a stronger subject-initial preference (73%) than German (50%); objects are fronted more often in German (7%) than in Swedish (3%); and adverbials other than temporal and locational are fronted more frequently in German (18%) than in Swedish (6%).[7]

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

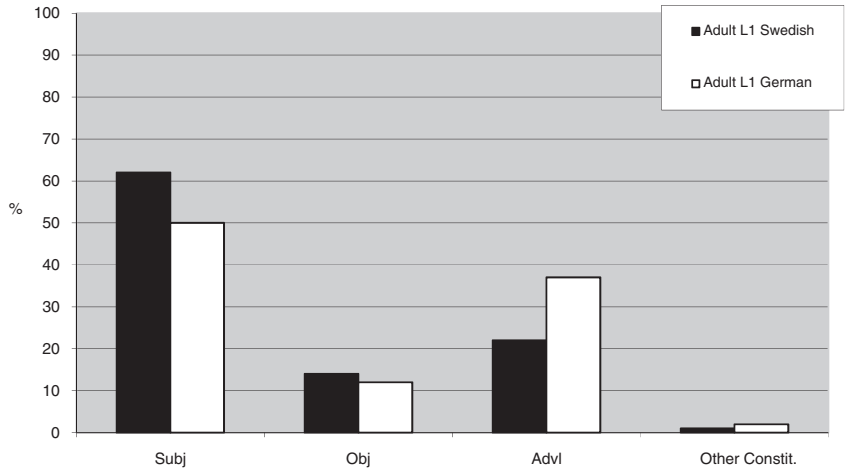

Figure 3. Overt constituents in the prefield, oral L1 data (Jörgensen 1976; Bohnacker corpus).

A similar asymmetry can be found in informal speech (Figure 3). I show this for a corpus of colloquial spoken German that I collected and transcribed myself (see Section 4 for details) and a spoken Swedish corpus from the Talbanken project at Lund University (Jörgensen 1976). Again, Swedish has a stronger subject-initial preference (62%) than German (50%) and fronts adverbials less often (22%) than German (37%), though there appears to be no pronounced difference concerning the frequencies for fronted objects (14%, 12%) in the spoken data from the two languages.[8,9]

These differences in frequency between Swedish and German led Rosén and myself to investigate the prefield in the two languages more closely. We were struck by the fact that Swedish speakers more often than German speakers placed phonologically light elements in clause-initial position, especially elements that had low or no informational value. Concerning subjects, it was particularly interesting to see that expletive *det* 'it' in the prefield in Swedish was much more frequent than expletive *es* 'it' in German. In Rosén's (2006) corpora of informal letters, 22% of all subject-initial sentences begin with an expletive in Swedish, but only 11% in German, as shown in Table 1. This difference is significant ($\chi^2 = 48.00$, $p < .001$).

Interestingly, the same asymmetry concerning expletive subjects can be found in informal speech (Table 2). I compare two corpora of spoken Swedish (Jörgensen 1976) with new corpus data from spoken colloquial German (Bohnacker corpus). In the two Swedish corpora, 16% and 19% of the subject-initial declaratives start with an expletive, but in the German data only 3% do so. The difference between the Swedish and German data is again significant ($\chi^2 = 221.08$, $p < .001$).

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

	Expletives out of all subjects	Expletives out of all overt prefield constituents
Adult L1 Swedish	22% (85/388)	16% (85/535)
Adult L1 German	11% (66/587)	6% (66/1173)

Table 1. Expletive subjects in the prefield, written data, informal letters (Bohnacker & Rosén 2008:520).

Source of data	Expletives out of all subjects	Expletives out of all overt prefield constituents
Adult L1 Swedish, Jörgensen A corpus	16% (99/632)	10% (99/979)
Adult L1 Swedish, Jörgensen B corpus	19% (578/3068)	13% (578/4610)
Adult L1 German, Bohnacker corpus	3% (48/1570)	2% (48/3001)

Note: Jörgensen A corpus – conversations and discussions between academics; Jörgensen B corpus – informal interviews; Bohnacker corpus – colloquial spoken German (see Section 4.2 below).

Table 2. Expletive subjects in the prefield, informal oral data (Jörgensen 1976; Bohnacker corpus).

Constituents other than subjects in the prefield would also deserve to be properly investigated (see Bohnacker & Rosén 2007), but due to space constraints I cannot do so here but only make a brief comment concerning objects. The definite inanimate pronoun *det* 'it/that' is by far the most common fronted object in Swedish, a fact that has also been noted in previous corpus studies (e.g. Rahkonen 2000, 2006). Such *det* is more frequent in the prefield than its German equivalent *das* 'it/that'. German speakers front a wider range of objects, both lexical and pronominal. For instance, in the informal written L1 texts collected by Rosén, *det* makes up 82% of all fronted object pronouns, but *das* only 24% (Rosén 2006:99–102). Such a large gap does not exist for fronted object *det* and *das* in the oral data, but it still is the case that native German speakers front a wider range of objects than the native Swedes do, including personal pronouns and lexical noun phrases.

The precise percentages of elements in the prefield (see Figures 2–3 and Tables 1–2) may be different for corpora of other text types, but my point here is that when keeping genre constant, there is a clear asymmetry between German and Swedish. I will argue that this is likely to be due to different tendencies in the two languages concerning the mapping of syntax and information structure.

3.2 Discourse-pragmatic similarities

The prefield is particularly important for communication as it anchors the clause in discourse. At the inter-sentential level, the prefield contributes to textual coherence by linking up with preceding discourse; at the intra-sentential level, it often establishes

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

the topic identified by the speaker, about which s/he then provides information (Reinhart 1982; Lambrecht 1994: Chapter 4; see Section 2 above). Processes of thought and communication motivate strategies by which the speaker as well as the listener identifies the topic as early as possible. Yet this does not mean that the prefield is a slot reserved for topics; topics can also occur elsewhere and many adverbials that are non-topics occur in the prefield as well. As regards theme and rheme, both Swedish and German have a tendency towards given before new, a tendency attested for many languages, which again may have to do with ease of online processing. This tendency, coupled with the V2 constraint of the two languages, gives rise to clauses where the prefield contains an element of low informational value. New (rhematic) information is usually provided later, after the finite verb (see Daneš 1970; Beneš 1971; Ekerot 1979; Hoberg 1981:174–176; Ekberg 1997:105–106; Zifonun et al. 1997:1640–1643; Teleman et al. 1999:53–64). Alternatively, the prefield may also host a focused element that the speaker wants to draw particular attention to, coded prosodically via stress and pitch contours (see e.g. Zifonun et al. 1997:1641–1642). The focused element may or may not be contrasted with other members of some evoked set of alternatives (recall note 3). Both Swedish and German mark focus by prosodic prominence and this can be done anywhere in the clause, so the prefield cannot be considered a designated focus position. These observations are not new and suggest that the two languages are information-structurally similar: There are certain word order tendencies, but little evidence for any DIRECT impact of information structure on Swedish and German syntax. Neither language appears to have a fixed slot for elements with a particular information-structural function, in contrast to what has sometimes been argued for other languages, such as a preverbal focus position in Korean or Hungarian (e.g. É. Kiss 1998:170–171; see Féry 2007 for an alternative view).

3.3 Discourse-related linearisation differences: Subjects and expletives

Despite these similarities, Bohnacker & Rosén's (2007, 2008) comparative Swedish and German corpus data (as well as an acceptability judgment task not reported on here) indicate that V2 languages may differ in the way they make use of the prefield, both quantitatively and qualitatively. They suggest that Swedish has a stronger tendency than German to keep informationally new (i.e. rhematic) material out of the clause-initial position and instead places it further to the right, i.e. postverbally. This can be achieved by filling the prefield with given (i.e. thematic) information, or with an element of no informational value, such as an expletive subject, or by leaving the prefield empty, as in V1 declaratives (not discussed here). We might thus say that Swedish linear syntax follows the information-structural principle of RHEME LATER, schematised in (4), more faithfully than German. The examples in (5) illustrate this:

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

Swedish disprefers clause-initial rhematic subjects; rhematic subjects (e.g. *en massa folk* 'lots of people') are nearly always postverbal and the prefield is filled by an expletive subject. No such tendency can be discerned for German – it is perfectly acceptable to start off with a rhematic subject (e.g. *ne Menge Leute* 'lots of people' in (5′a)) and doing so is more common in Bohnacker & Rosén's data than filling the prefield with an expletive (5′b).

(4) *Rheme Later*

PREFIELD	FINITE VERB	MIDFIELD, ETC.
Expletive or given information		New information

(5) A: 'Anything happened this morning?'
B: a. **Det** har ringt en massa folk till dig. (Swedish, preferred)
EXPL has called a lot people to you
b. **En massa folk** har ringt till dig. (Swedish, dispreferred)
'Lots of people have been calling you.'

(5′) a. **Ne Menge Leute** haben dich angerufen. (German, preferred)
a lot people have you called
b. **Es** haben dich ne Menge Leute angerufen. (German, dispreferred)
EXPL have you a lot people called

Swedish has a range of constructions with an expletive in the prefield, like the presentational sentences in (5), existential constructions and clefts, where new information is introduced postverbally (see e.g. Ekberg 1997:105–106, Teleman et al. 1999:53–64).

Corresponding expletive-initial constructions do exist in German but are much less common. I will illustrate this for the case of clefts, as they also feature in the discussion of the L2 data later in this paper. Clefts consist of an expletive, a copula, a clefted constituent and an embedded clause. While the clefted constituent is often a nominal phrase in both Swedish and German, Swedish allows more types of morphological and syntactic categories to be clefted, including pronominals and adverbial phrases (see Dyhr 1978:99, 188; Huber 2002:84–94). Corpus studies indicate that Swedish (and Danish and Norwegian) generally have much higher rates of clefts than German (and English) – double, four or five times as high (Dyhr 1978:166, 178; Kiese 1993:42–48; M. Johansson 2001:560–561; Gundel 2002; S. Johansson 2007:Chapter 12). The same obtains for the corpora in the present study: In the Swedish data from Rosén (2006) and Jörgensen (1976), expletives introducing clefts make up, respectively, 2% and 3% of all prefields. The German data (Rosén, Bohnacker) on the other hand only contain one single expletive-initial cleft (0.02% of all prefields). In naturalistic German discourse, clefts are not only exceedingly rare but also largely restricted to cases of strong contrast, with minimal focus on the clefted constituent. In Swedish, expletive-initial clefts are commonplace and occur

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

with a wide range of functions (e.g. Huber 2002:175–184). Swedish clefts are not primarily used for contrast, but often simply as a way of placing rhematic information to the right of the verb, as in the following authentic examples (*någon* 'somebody' in (6), *nåt* 'something' in (7), *först i juni i år* 'only this June' in (8)). German prefers not to use a cleft construction here but rather places the rhematic constituent preverbally in the prefield (see *jemand* in (6′), *irgendetwas* in (7′) and *erst im Juni diesen Jahres* in (8′)).

(6) **Det** är någon som vill tala med dej. (Swedish)
EXPL is somebody that wants talk with you
'Somebody would like to see you./There's someone to see you.'

(6′) Jemand möchte dich sprechen. (German)
somebody would.like you talk

(7) **Det** är nåt som inte stämmer här. (Swedish)
EXPL is something that not is-right here
'Something isn't right here./There's something wrong here.'

(7′) Irgendetwas stimmt hier nicht. (German)
something is-right here not

(8) [Context: Divers discovered a sunken vessel off the Swedish coast last year.]
Det var först i juni i år som dyklaget lyckades (Swedish)
EXPL was only in June in year that diving.team.the managed
identifiera den.
identify it
'The diving team only managed to identify it this June.'

(8′) Erst im Juni diesen Jahres konnten es die Taucher identifizieren. (German)
only in June this year could it the divers identify

Note that the Swedish clefts in (6)–(8) are not contrastive: Clefted *någon* 'somebody' in (6) is not contrasted with some other person who doesn't want to see the addressee. There is no contrastive or minimal focus on *någon*. Rather, the information provided in (6) is all new, with maximal focus on the entire cleft sentence (*någon som vill tala med dig*). Similarly, the clefted constituent *först i juni i år* 'only this June' in (8) is not contrastive, it does not carry minimal focus, but rather, the entire cleft sentence in (8) is new information. Similar 'all-new' clefts also regularly occur in English, Norwegian and French (see e.g. Hedberg 2000; Gundel 2002; Robach 2003).

3.4 Discourse-related linearisation differences: Objects

In the preceding sections, I have suggested that while both languages have a general tendency of theme before rheme, Swedish linear syntax follows the information-structural principle of Rheme Later more faithfully than German does. This observation also pertains to objects. Clause-initial rhematic objects are rare in

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

Swedish – if fronted, objects are nearly always themes (old information) and very often simply consist of an anaphoric definite pronoun, especially *det* 'it/that', as in (9).

(9) A: 'We've got a special offer today – vegetable lasagne for 3.99.'
B: **Det** tar vi.
that take we
'We'll have that.'

Pronominal *det* makes up 91% (288/318) of the prefield objects in the informal Swedish speech data (from the Talbanken Jörgensen B corpus). Its German equivalent *das* appears in the prefield too (47% (141/302) of the fronted objects in the informal speech data), but German speakers regularly front a wider range of pronominal and lexical objects as well, such as *Kissen und einen blauen Flickenteppich* 'cushions and a blue rug' in (10). Swedish speakers would instead start with a thematic subject (*jag* 'I') and postpone the object *kuddar och en blå trasmatta*, as is shown in (10′).

(10) Gestern war ich bei IKEA und hab zwei Regale besorgt. **Kissen und einen blauen Flickenteppich** hab ich auch gekauft. (German)
yesterday was I at IKEA and have two shelves got cushions and a blue rug have I also bought
'Yesterday, I went to IKEA and got two shelves. I also bought cushions and a blue rug.'

(10′) a. ?**Kuddar och en blå trasmatta** köpte jag också. (Swedish, dispreferred)
cushions and a blue rug bought I also
b. Jag köpte också kuddar och en blå trasmatta. (Swedish, preferred)

In addition to these slightly divergent tendencies concerning theme–rheme, there may also be different word order tendencies at another information-structural level when fronted objects are considered. As mentioned above, Rosén (2006) found that inanimate definite *det* 'it/that' made up the bulk of fronted pronominal objects in informal written Swedish (82%), and this holds even more strongly for the oral data (96%), whereas German regularly fronts other pronouns. Yet why would Swedish front pronominal objects other than *det* less frequently than German? Such a difference cannot be accounted for straightforwardly by saying that Rheme Later is stronger in Swedish.

Pronouns in both languages typically encode old information (the theme) as they refer back to a textually accessible antecedent, and they also serve to build topic continuity. It is not surprising, then, that they often occur in the prefield. For instance, both object pronouns *die* 'her' and *henne* 'her' in (11)–(11′) function as themes, as they refer back to the previously-mentioned Louisa. But why is it, then, that a fronted animate object is fine in German (11a) but dispreferred in Swedish (11′a)?[10]

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

(11) A: 'And what're we gonna do about Louisa?'
B: a. **Die** hab ich schon gefragt. (German, preferred)
her have I already asked
b. Ich hab die schon gefragt. (German, dispreferred)
'I've already asked her.'

(11′) a. ?**Henne** har jag redan frågat. (Swedish, dispreferred)
her have I already asked
b. Jag har redan frågat henne. (Swedish, preferred)

The answer might be found at another information-structural level. Constituents that encode new information for the listener (rheme) are often also considered by the speaker to be the most relevant (focus) and realised with greater prosodic prominence. But rheme and focus do not always coincide – old/given/thematic referents can be focused too, e.g. by being placed in a new, unpredictable or as yet unsettled relationship. In such cases, thematic constituents such as pronouns are in focus, and then bear sentence stress. This analysis works well for both German and Swedish when there is a stressed thematic animate object pronoun in the prefield, as in (11″) and (11‴).

(11″) **DIE** hab ich schon gefragt.
'HER, I've already asked.'

(11‴) **HENNE** har jag redan frågat.
'HER, I've already asked.'

Accenting this already given referent induces an interpretation of narrow/minimal focus on 'her'; a set of alternatives is created and there is felt to be a salient opposition in what is predicated of them, for instance, she (i.e. Louisa) has already been asked, but some other person(s) have not. Animate personal object pronouns in the Swedish prefield are thus not impossible per se, but appear to be focused and realised with greater prosodic prominence than inanimate pronouns (e.g. *det*), which easily function as neutral themes.[11] For a native speaker of Swedish, fronted *henne* in (11′) evokes a situation where the speaker contrasts having asked Louisa while not having asked one or several other persons. When no such narrow-/minimal-focus reading is intended, the object pronoun remains both unstressed and unfronted (11′b). Hazarding a guess, I would expect that narrow/minimal focus on pronominal objects is not very common in text corpora, and if this turns out to be true, it could explain why fronted personal object pronouns are infrequent in the Swedish data. In German, there is nothing wrong with unstressed personal object pronouns in the prefield, so the numbers of fronted personal object pronouns are higher. (Of course, this still does not tell us why there might be such a difference between Swedish and German.) These speculations will need to be investigated more thoroughly in future work.[12]

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

4. INFORMANTS, DATA COLLECTION AND METHOD

4.1 Learners

The L2 learners are six adult native speakers of German who had all had a monolingual childhood in Germany or Austria. At school they had 7–9 years of English as a foreign language (from age 10/11 years), and some years of Latin or French. As regards their Swedish, they were adult learners; none had been exposed to Swedish before the age of 20. At the time of the study, the informants were long-term residents of Sweden and they used Swedish every day, in the workplace, with friends and/or at home. They were university graduates in their early twenties to late thirties, employed at schools, universities and with the local council, as teachers, researchers, cleaners and therapists. While all had been exposed to classroom Swedish, most of their acquisition was naturalistic. Three had attended classes in Swedish as a foreign language in Germany (2 hours per week for one year) and began to work immediately upon arrival in Sweden, without taking further classes. The other three learners had no previous knowledge of Swedish before arriving in Sweden. They attended Swedish classes for immigrants (4–10 hours per week) for one year, after which acquisition continued untutored. The learners were advanced in the sense that they were communicating fluently and had passed the respective Swedish university-entry language proficiency exams (Rikstestet/TISUS) before data collection started.[13] The learners stated that they felt at ease when speaking Swedish but less confident when writing the language.

The data from these informants were collected during the 1990s and 2000s at three-year intervals. From Ulrike and Steffen data were collected 3, 6 and 9 years after arrival, and from Stella, Nicole, Ellen and Stefanie after 3 and 6 years in Sweden.[14] All the data are naturalistic production, spoken and written. The oral data consist of a 45-minute recording of the informant narrating events of their life in conversation with an experimenter, and for some of them an additional 45-minute recording of the informant teaching a class or giving a seminar at the workplace in the absence of experimenters. Each recording consists of 5000–7000 words and contains both dialogue and monologue passages. Additionally, each informant supplied 5000 words of unedited written text (informal emails/letters). Word order and constituent type were classified and coded by hand (Bohnacker 2007). Oral and written data are investigated separately, so as not to mask potential information-structural differences between the two modes. However, I have collapsed oral narrative and oral teaching into one informal oral category, as I could not detect any substantial differences concerning prefield use between them. Most of the L2 data were originally collected for a study on verb placement and verb particle constructions (Bohnacker 2007), but they can also be used to study the prefield of declaratives. There are 9,563 declaratives or instances of a filled or potentially filled prefield, 3,423 for the written data and

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

6,140 for the oral data. The learners placed the finite verb in a targetlike manner, with only 0.026% violations of the V2 constraint (Bohnacker 2007:24–26).

4.2 Native speakers

The native informants are adults and roughly of the same age group as the L2 learners. L1 text types have been matched as closely as possible with those of the L2 learners. The written L1 corpora comprise compositions (informal letters, summaries) by 70 native speakers of German (28,500 words) and 80 native Swedish speakers (17,500 words), and were collected by Christina Rosén between 1999–2005 (see Rosén 2006).

The oral L1 data consist of conversations between native speakers that include both dialogue and more monologue-like narrative passages. Here I am using a new, previously unpublished corpus of oral native German, which contains 30,700 words of colloquial South German dialogue (Bohnacker corpus). Three female informants age 25–35 and three age 60–70, all from the greater Ulm area, were recorded for four hours in one-to-one conversations with a local experimenter.[15] The recordings were made between 1994 and 2000 and transcribed by myself. Word order and constituent type were manually classified and coded.

For Swedish, I perused Jörgensen's (1976) corpus study of recordings made in 1968. This includes (i) conversations and discussions between academics (8 informants, 3 hours of recording, Jörgensen A), and (ii) 32 informal interviews of 30–45-year-old employees on the topic of immigrants and immigration (8–9 hours of recording, Jörgensen B). The conversations and discussions comprise 11,200 words, and the interviews 45,000 words. These data were collected and analysed as part of the Lund University Talbanken project during the 1970s. I noted Jörgensen's (1976) counts but also carried out manual searches of a 30,700-word portion of the informal interviews, comparable in size to the oral native German data.

5. RESULTS: HOW GERMAN-SPEAKING LEARNERS OF SWEDISH MAKE USE OF THE PREFIELD

5.1 Subjects and expletives

The overall frequencies of subject-initial clauses out of all V2 declaratives in the L2 Swedish productions are closer to native German (50% in the written and 50% in the oral data) than to native Swedish (73% in the written, 62% in the oral data). For ease of exposition, the data have been aggregated for the learners at 3 vs. 6 and 9 years of residence in Sweden. As shown by the black bars in Figures 4 and 5, in the L2 writing, the percentages of subject-initial declaratives hover between 37% and 50%, and between 49% and 60% in the oral L2 data. (Exact raw figures are provided in the

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

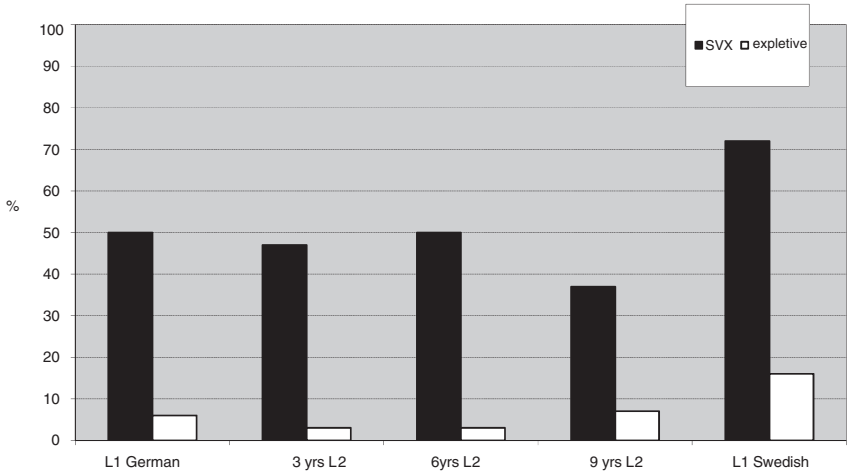

Figure 4. Prefield subjects and expletives in L1 German, L2 Swedish and L1 Swedish, informal written data (letters).

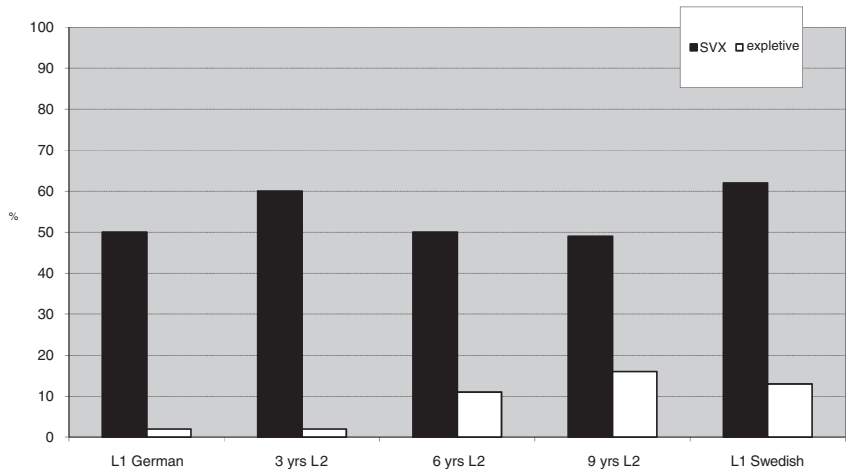

Figure 5. Prefield subjects and expletives in L1 German, L2 Swedish and L1 Swedish, informal speech.

appendix.)[16] The white bars represent expletive-initial clauses and will be discussed shortly.

When we investigate the informants' subject-initial declaratives more closely, developmental tendencies emerge. At 3 years, informationally new and phonologically heavy subjects regularly occur in the prefield for all six speakers (see examples (12)–(13)), at 6 years they are found in some speakers, and at 9 years occasionally only. While not ungrammatical, these heavy clause-initial subjects are

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

In Sweden	Expletives out of all subjects	Expletives out of overt prefield constituents
3 years, 6 learners	5% (35/648)	3% (35/1380)
6 years, 6 learners	5% (29/564)	3% (29/1121)
9 years, 2 learners	19% (46/240)	7% (46/654)

Table 3. L2 expletive subjects in the prefield, informal written data.

In Sweden	Expletives out of all subjects	Expletives out of overt prefield constituents
3 years, 6 learners	3% (45/1306)	2% (45/2179)
6 years, 6 learners	22% (210/962)	11% (210/1917)
9 years, 2 learners	33% (185/561)	16% (185/1158)

Table 4. L2 expletive subjects in the prefield, informal oral data.

unusual in L1 Swedish, where one would preferably start with a light expletive subject and place the rhematic subject postverbally, as in (12′) and (13′).[17]

(12) Tack för korten. **En riktigt fin sjungvecka** var det med er!
thanks for pictures.the a really nice sing.week was it with you
'Thanks for the pictures. It's been a really nice singing week with you!'
(Swedish L2; Ellen, 3 years, written)

(12′) **Det** har varit en riktigt fin körvecka med er. (Swedish L1, preferred)
EXPL has been a really nice choir.week with you

(13) 'Here you can see that it increases with age, but you can't see it so well here.'
Men **en bättre diagram** är på sida 67.
but a better chart is on page 67
'But there's a better chart on page 67.'
(Swedish L2; Ulrike, 3 years, oral, teaching)

(13′) Men **det** finns ett bättre diagram på sidan 67. (Swedish L1, preferred)
but EXPL is a better chart on page.the 67

There is a clear trend in the learner data concerning clause-initial expletive *det*, plotted as white bars in Figures 4 and 5. At 3 years, the learners rarely begin a declarative clause with expletive *det* (2% written, 3% oral), which is substantially lower than the native Swedish speakers (16% written, 13% oral), but similar to native German (6% written, 2% oral). Exact figures are provided in Tables 3 and 4. The difference between the L2 learners at 3 years and the native Swedish speakers is significant both for the written condition ($\chi^2 = 117.01$, $p < .001$) and the oral condition ($\chi^2 = 194.71$, $p < .001$).

From 6 years onwards, the proportion of clause-initial *det* rises, though this is evident at first in the oral data only (Tables 3–4). Individual results are plotted in Figure 6. All six L2 learners show increased production of clause-initial *det* from the

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

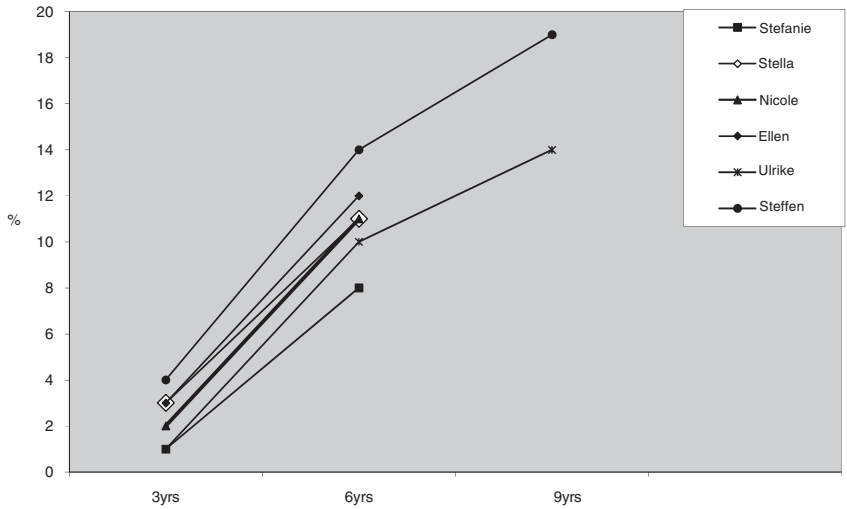

Figure 6. L2 expletive subjects in the prefield, informal speech.

3-year to the 6-year data point. However, only two learners (Ulrike and Steffen) have been studied at 9 years, so it remains to be seen whether the other four learners show the same developmental pattern for the entire nine-year period.

Clause-initial expletives introducing clefts are not found at all in the learners at 3 years. Recall that there were hardly any in the L1 German data either (1 instance, 0.02%), but 2% and 3% in the L1 Swedish data. This absence of clefts in the L2 data, as well as the rarity of clause-initial expletives in general, is unlikely to be a sampling artefact, as the learners are producing a large number of declaratives (3,904) at this data point (3 years). Rather, I believe, it is to be interpreted as transfer of the L1 German pattern to L2 Swedish.

At 6 years, *det*-clefts begin to appear as well, though again largely in the oral data. This development is not only visible in the aggregated group data but also in the individual data. Compare for instance the near-minimal pairs from one and the same learner at 3 and 6 years and at 3 and 9 years (examples (14)–(16)).

(14) Swedish L2, Steffen (3 years, oral) and interviewer (I)
S: Till henne har ja(g) sagt att JAG skall fixa det.
to her have I said that I shall fix it
'I've told her that I'll sort it out.'
I: Men de(t) är väl HON som ska göra de(t).
but it is well she that shall do it
'But SHE's the one who's supposed to do it, right?'
S: Nej, **JAG** skall göra de(t). (Swedish L1, dispreferred)
no I shall do it
'No, *I*'m supposed to do it./No, it's *ME* who's gonna do it.'

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

(14′) Swedish L2, Steffen (6 years, oral) and interviewer (I)
I: Å då hamnar de(t) väl på hans bord.
and then lands it well at his desk
'And then it'll be his responsibility.'
S: Nej, **de(t)** är JA(G) som ska göra de(t). (targetlike)
no EXPL is I that shall do it
'No, *I*'m supposed to do it./No, it's *ME* that'll be doing it.'

(15) Swedish L2, Ellen (3 years, oral)
Nånting stämde inte där. (Swedish L1, dispreferred)
something was-right not there
'Something wasn't right there./There was something that wasn't right there.'

(15′) Swedish L2, Ellen (6 years, oral)
De(t) är ju nånting som är lurt här. (targetlike)
EXPL is well something that is fishy here
'There's something that isn't right here.'

(16) Swedish L2, Ulrike (3 years, oral, teaching)
[Context: Interrupts her teaching, listens to faint noises outside.]
Nån vid dörren knackar. (Swedish L1, dispreferred)
someone at door.the knocks
'Someone's knocking at the door./There's someone knocking at the door.'

(16′) Swedish L2, Ulrike (9 years, oral, teaching)
De(t) är nån som är därute. Är **de(t)** nån som kan
expl is someone that is there.out is EXPL someone that can
gå å öppna? (targetlike)
go and open
'There's someone out there. Can someone go and open the door?'

Over the years, the proportion of expletive-initial declaratives increases manifold, from 3% at 3 and 6 years in the written data to 7% at 9 years and from 2% at 3 years in the oral data to 11% and 16% at 6 and 9 years, respectively. I suggest that this change is indicative of the learners' growing awareness of the Swedish frequency distributions and information-structural patterns, with a strong preference for rhematic information being placed later in the clause. The increased use of clause-initial expletives achieves just this.

5.2 Objects

In general, the learners produce more clause-initial objects than native Swedish speakers do, as shown by the black bars in Figures 7 and 8. They do so significantly more in their writing (Figure 7) than in informal speech, where the difference between learners and native Swedes is not significant (Figure 8). Recall also that informal spoken Swedish does not exhibit fewer fronted objects than German does, only fewer types of fronted objects. (Exact figures are provided in the appendix.) The white bars in Figures 7 and 8 represent *det*-initial clauses and will be discussed shortly.

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

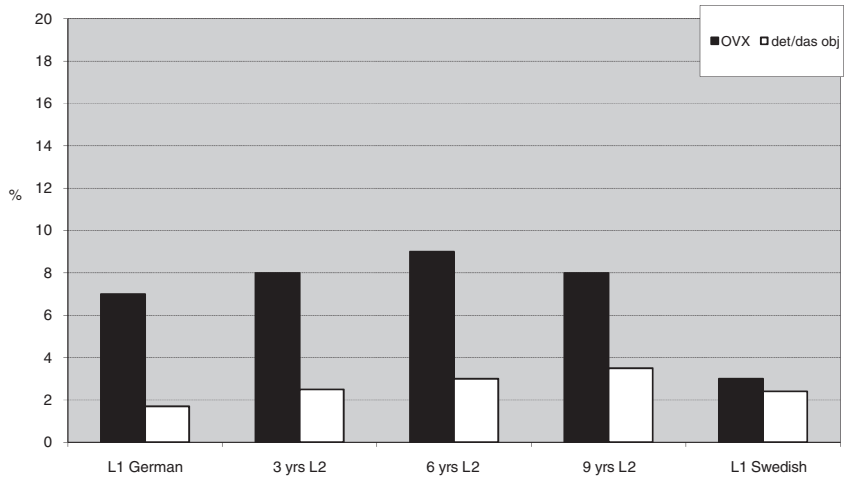

Figure 7. Prefield objects and object *det/das* in L1 German, L2 Swedish and L1 Swedish, informal writing.

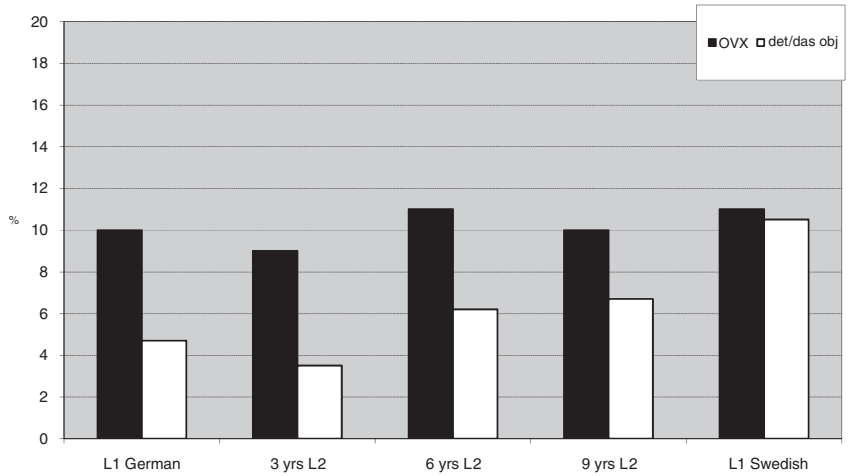

Figure 8. Prefield objects and object *det/das* in L1 German, L2 Swedish and L1 Swedish, informal speech.

Higher rates of object fronting alone do not tell us whether the learners diverge from the information-structural patterns of native Swedish. On closer scrutiny, however, the learners do appear to produce slightly different types of clause-initial objects and arguments than the native speakers. One such difference concerns the

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

fronting of the thematic inanimate object pronoun *det* 'it/that', another concerns the fronting of heavy rhematic objects, and a third the fronting of animate personal object pronouns. I will briefly discuss these in turn. For practical reasons, it has not been possible to investigate these issues in a quantitative way, so in some cases I only present some representative examples from a cursory survey of the data with regard to the information-structural properties associated with OVS sentences.

A commonly fronted object in native Swedish is *det* 'it/that'. As indicated by the white bars in Figure 7, the L2 learners also front *det*. Although it is not the predominant type of fronted object in their writing, it constantly makes up roughly a third (27%–38%) of all fronted objects. In the oral L2 data, pronominal object *det* occurs at substantially higher rates and these rates increase over time. As Figure 8 shows (white bars), *det* is common already at 3 years, making up 38% (77/200) of all fronted objects, rising at 6 years to 60% (113/188) and at 9 years to 67% (87/130) of all fronted objects, which suggests that the more advanced learners come to behave much like native speakers. An example is given in (17).

(17) Swedish L2, Stella (6 years, oral) and interviewer (I)
[Context: Discussing Swedish and German bread baking with interviewer.]
S: **De(t)** tycker inte JAG.
it think not I
'*I* don't think so.'
I: . . .

S: Eller om man tar färsk jäst, ja(g) menar . . .
or if one take fresh yeast I mean
'Or if you take fresh yeast, I mean . . .'
S: **De(t)** gör JA(G) i alla fall.
it do I in any case
'At least *I* do that.'

Besides *det*, the learners front a variety of objects, both pronominal and lexical. It is those that lead to the comparably high proportion of fronted objects, shown by the black bars in Figures 7 and 8. These include fronted objects that are informationally new, as in (18)–(19). Native Swedish speakers often perceive such sentence openings as unidiomatic, heavy, stilted, old-fashioned and un-Swedish, and would instead start with a light subject pronoun (*jag* 'I', *du* 'you') and place the rhematic object postverbally, see (18′), (19′).

(18) I lördags har jag varit på IKEA och köpt två bokhyllor.
on Saturday have I been at IKEA and bought two bookcases

En blå trasmatta och en kudde har jag också köpt.
a blue rug and a cushion have I also bought
'On Saturday, I went to IKEA and bought two bookcases. I also bought a blue rug and a cushion.' (Swedish L2; Ulrike, 3 years, written)

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

(18′) Jag köpte även en blå trasmatta och en kudde. (Swedish L1, preferred)
I bought also a blue rug and a cushion

(19) Hoppas att du har haft roligt igår. **En cykeltur eller en längre tur till havet** har du gjort, kanske?
hope that you have had fun yesterday a bike.trip or a longer trip to sea.the have you made perhaps
'I hope you had fun yesterday. You went for a bike ride or a longer trip to the seaside perhaps.' (Swedish L2; Nicole, 3 years, written)

(19′) Du (kanske) gjorde en cykeltur eller en längre tur till havet (kanske)?
you maybe made a bike.trip or a longer trip to sea.the maybe (Swedish L1, preferred)

Clause-initial heavy rhematic objects such as in (18)–(19) regularly occur in the oral and written data of all six learners at 3 years. At later data points, however, such object fronting is largely restricted to the learners' writing. It seems that the learners in their oral productions at 6 and 9 years have moved closer to the Swedish pattern, where rhematic information is realised postverbally and where the clause-initial position typically contains light thematic or expletive elements. In their writing, however, this development appears to lag behind. (For some comments on this, see Section 6.)

Also of interest is the occurrence of fronted pronominal objects other than *det*. As noted in Section 3.2 above, Swedish hardly ever fronts animate personal object pronouns, except when they bear sentence stress and are in focus, thereby inducing an interpretation of contrast. In German, unstressed animate personal object pronouns front easily. They function as simple thematic elements, without minimal/narrow focus. As we will see, this creates a problem for the L2 learners. Consider (20), where Nicole fronts the (unstressed) prepositional object *till honom* 'to him', referring back to *din chef* 'your boss' mentioned earlier.

(20) [Context: About doing a training course abroad and how to go about telling one's boss.]
I: Å när ska du berätta de(t) för din chef?
and when shall you tell it for your boss
'And when're you gonna tell your boss?'
N: **Till honom** har ja(g) redan sagt det, men det blir kanske inte av i alla fall.
to him have I already said it but it become maybe not off in any case
'I've already told him, but it might not happen anyway.'
(Swedish L2; Nicole, 3 years, oral)

(20′) a. {**Dem/ihm**} hab ich das schon gesagt. (German L1, preferred)
him have I it already said
b. Jag har redan {sagt/berättat} det {till/för} honom. (Swedish L1, preferred)
I have already said/told it to/for him

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

For a native speaker of Swedish, fronted *till honom* evokes a situation where having spoken to the boss is contrasted with not having spoken to one or several other persons. No such interpretation appears to be intended by Nicole, and there is no stress on *till honom*, which makes the fronted object inappropriate and the sentence 'un-Swedish'. The same applies to the fronted objects in (21)–(22). I suggest that they are due to L1 transfer, since in the learners' native German, such fronted pronominal objects would be acceptable and simply interpreted as neutral, given information (cf. (21′), (22′)).

(21) Det gör inget att du inte kan nu. **Oss** ska sen pappa hjälpa.
it does not that you not can now us will then dad help
'It doesn't matter if you can't come (and help) now. Dad will help us later.'
(Swedish L2; Stella, 3 years, written)

(21′) a. Uns wird dann der Vater helfen. (German L1, preferred)
us will then the father help

b. Pappa kommer att hjälpa oss sen. (Swedish L1, preferred)
dad will to help us later

(22) Jag har inte hört nånting från Aisa.
'I haven't heard anything from Aisa.'
Henne har jag också skickat artikeln.
her have I also sent article.the
'I sent the article to her as well.' (Swedish L2; Ulrike, 6 years, written)

(22′) a. {**Der/ihr**} hab ich den Artikel auch geschickt. (German L1, preferred)
her have I the article also sent

b. Jag har skickat artikeln till henne också. (Swedish L1, preferred)
I have sent article.the to her also

After three years of immersion in a Swedish-dominant environment, the learners seem insensitive to the subtle interpretive effects of placing objects in clause-initial position, or the left periphery of the clause.[18] But again there is development over time. At 6 and 9 years, fronted animate pronominal objects have nearly vanished from the oral data, while occasional instances can still be found in the written learner data, such as that in (22).

6. DISCUSSION AND CONCLUSION

In this paper, I have presented naturalistic production data from six L1 German advanced learners of Swedish that suggest that these learners exhibit German-style frequency patterns in the prefield and information-structural patterns which are different from native Swedish. This can be seen most clearly for the earliest data discussed here, after three years of exposure. The learners generally produce fewer subject-initial clauses and more object-initial clauses than the Swedish native

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

speakers, but these overall differences in frequencies do not make the L2 productions off target per se. What sets the learners apart is their rare use of expletives and other clause-initial elements of no or low informational value and their placement of heavy rhematic elements in the prefield.

Swedish and German, which are typologically close, do not implement information-structural principles that are diametrically opposed but only slightly different from each other: Rheme Later is stronger in Swedish than in German, i.e. Swedish has stricter discourse-pragmatic constraints than German on what can (or does) occur in the prefield. Of course, discourse-pragmatic constraints on word order are not on–off, either–or, grammatical–ungrammatical: Adequacy depends on context. One can place a heavy rhematic constituent in the Swedish prefield (but one usually does not), and one can start a German declarative with an expletive (but one usually does not). What native speakers usually do and what they usually do not do are tendencies and preferences, and going against these will not result in ungrammaticality; it is simply pragmatically odd. At three years of exposure, the learners have not adjusted their German-style discourse-pragmatic constraints but are overusing the prefield position from a Swedish perspective.

This overuse of the prefield in the present study can be related to Bohnacker & Rosén (2008), who studied the reverse language combination (L1 Swedish learners of L2 German) and found that their informants underused the prefield from a German perspective, restricting it largely to elements of no or low informational value. In parallel with Bohnacker & Rosén's findings, the L2 production data at three years can be interpreted as evidence for L1 transfer of information-structural patterns or discourse-pragmatic preferences.

However, the data presented here also indicate development towards the target, in the sense that the learner productions at six and nine years show substantially higher rates of clause-initial expletive subjects, clefts and lightweight given elements (e.g. pronominal object *det*). Contextually inappropriate rhematic elements in the prefield decrease, as do thematic animate pronominal objects that give rise to an unwarranted narrow-/minimal-focus reading. This suggests that syntax–pragmatic difficulties can be overcome as proficiency develops.[19]

Interestingly, the development towards the target sets in earlier and more forcefully in speech than in writing for these learners. This may be a surprising result since in other studies (and in studies of other phenomena) L2 learners often do better in unspeeded writing tasks because they have time to monitor. Our learners may also have monitored their writing for a number of things (see Krashen 2003; Kroll 2003), yet it is unlikely that they monitored for differences concerning the interaction of information-structure and word order, as they were probably not even aware that Swedish differs from German in this regard. The word order preferences under discussion are not readily accessible to introspection as aspects of (pure)

syntax might be, where only one specific word order is grammatical (e.g. VO vs. OV, V2 vs. non-V2), or morphology, where a particular inflectional morpheme is obligatory and others are ungrammatical. Discourse-driven word order patterns are also largely ignored in descriptive grammars, teacher training and language teaching materials. My guess would be that residual L1 patterns show more strongly in the L2 writing because writing is what these learners of Swedish do the least – their acquisition and immersion has been mainly oral and aural (see Section 4 above).

The learners in this study do better than Bohnacker & Rosén's (2008) learners of the reverse language combination. This is likely to be an effect of increased exposure (input) during their three and six or more years in a dominant L2 environment vis-à-vis foreign language classroom learning for a maximum of six years in Bohnacker & Rosén's study. Even though it is impossible to measure input exactly, Bohnacker & Rosén's advanced group can be estimated to have received 1800–2000 hours of input during their six years of L2 studies. By contrast, even on a conservative estimate of only four daily hours of exposure for 300 days per year, the advanced learners in the present study would have had 3600 hours of communication in Swedish by three years, and 7200 hours by six years. (The actual number of hours is likely to be substantially higher.) The quality of input may also play a role. Taught foreign language learners (like Bohnacker & Rosén's informants) are almost invariably exposed to non-native input from their classroom peers and perhaps even at the level of instruction. L2 learners who receive naturalistic input in an immersion setting (as the informants in the present study) do not run that risk.

Now how do the results presented here relate to recent empirical L2 studies mentioned in Section 1 on the acquisition of discourse-driven distribution of null vs. overt subjects and postverbal focused subjects in languages such as Italian and Spanish or referent introduction and maintenance in L2 German narratives? Common to them all is that learners at high proficiency levels continue to grapple with the pragmatically appropriate distributional pattern (e.g. Carroll & Lambert 2003; Carroll & von Stutterheim 2003; Hertel 2003; Lozano 2006; Sorace & Filiaci 2006; Belletti et al. 2007; Rothman 2009). Since in many of these studies the highest-proficiency group diverges from the native controls, some researchers have drawn the conclusion that information structuring in L2 acquisition never becomes fully native-like (e.g. Carroll & von Stutterheim 2003; von Stutterheim 2003; Sorace & Filiaci 2006; Valenzuela 2006:301). Other researchers have found that the performance of their most proficient learner group is not significantly different from that of the native controls concerning the particular phenomenon under investigation. They therefore conclude that while syntax–pragmatic difficulties may be persistent, these can be surmounted as proficiency develops (e.g. Rothman 2009). This is a hopeful conclusion and on the basis of the present study, I would concur with it.

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

Much of the aforementioned current research on discourse-driven word order and subject realisation maintains that the interaction of syntax and discourse pragmatics is one of the hardest areas to master, and a commonly cited reason for this is that it involves a grammar-external interface, where multiple types of information have to be integrated with each other. At first sight, the results from the present study may support such a conclusion, since the learners at three years significantly differ from native speakers, even though they had been immersed in a Swedish-dominant environment and generally were at a high proficiency level, as determined by the Swedish-language university entry exam Rikstestet/TISUS, which all of them had passed prior to data collection.

However, discourse-appropriate use of the prefield is not the only area of language these learners have problems with. In an earlier study of the same learners (Bohnacker 2007) it was shown that they had not acquired the syntax of Swedish transitive verb particle constructions (VPCs) after six years of immersion. In that study, I argued that syntactic operations related to the topmost clause-structural level (CP), manifested by V2, are not universally the most difficult area of syntax to attain. Lower structural levels, such as VP, manifested by VPCs, may present equally severe or in fact more severe acquisition problems. At that time, a number of researchers (e.g. Hulk & Müller 2000; Platzack 2001) were proposing that syntactic phenomena involving the left periphery of the clause (the CP domain) were the hardest to acquire for all learners. CP was seen as particularly 'vulnerable' as it mediates between the propositional content of the clause and the linguistic discourse or discourse situation on the outside. This view is very similar to current theorising that considers the grammar-external interface to be the ultimate challenge for L2 learners. I am not so sure about this and believe that the jury is still out. I would contend that not everything falling into the large category of syntax–discourse interaction can be lumped together if we want to decide which phenomena present a particular challenge for L2 acquisition. Certain syntactic–pragmatic 'interface' phenomena may well be the hardest for some language combinations, but phenomena in other domains of language (e.g. phonology) might be the ultimate L2 challenge for other language pairings (see White 2009).

In my view, it is not the involvement of an 'interface' as such that makes the L2 acquisition of certain structures difficult, but rather the fact that language-specific possibilities and constraints can have subtle consequences for information distribution in discourse, and these are often gradual in nature. Recall that we are dealing with native speakers' PREFERENCES for information selection and distribution in particular contexts and discourse types, and when very advanced non-native speakers neglect to adopt these preferences, this will lead to a discourse accent but not really to ungrammaticality (see also Carroll & Lambert 2003; von Stutterheim 2003; Rothman 2009:968).[20]

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

As for my initial discussion of interfaces in Sections 1 and 2, the learners in the present study, like those in other recent studies (e.g. Lozano 2006; Tsimpli & Sorace 2006; Belletti et al. 2007), master pure syntax (V2) well before they master the appropriate discourse-pragmatic use of that syntax. One might want to capture this with a cartographic approach, where the language-specific information-structural differences and L2 problems would essentially be treated as SYNTACTIC and GRAMMAR-INTERNAL, and located inside the computational system (e.g. Belletti et al. 2007). In this case, we might assume 'interpretable' features such as [+Given/Theme], [+Focus] and corresponding left-peripheral functional projections corresponding to the prefield, with an interface mapping between these syntactic features and an interpretive component. The L2 learners could be said to have a representational deficit such that these features are underspecified (or specified differently), thereby allowing a wider range of possible mappings, e.g. both thematic and rhematic elements in the prefield, than native speakers do. I am not adopting such an approach here. One could also describe the results with a discourse-free theory of syntax (e.g. Prince 1998; Neeleman & de Koot 2008), where V2 is identical across languages, but where information-structural differences between Swedish and German and ensuing L2 problems are treated as PRAGMATIC and outside of grammar and purely located at an EXTERNAL interface with the conceptual-intentional system.

To my mind, determining what constitutes internal and external interface problems in L2 acquisition seems to have a lot to do with one's predilection for a particular formal theory of syntax, and not only with the empirical learner data. Thus, I will not take a stand on the interface issue here.

What I have tried to show in this paper is that there are cross-linguistic differences concerning the relationship between word order and discourse pragmatics in V2 declaratives in Swedish and German and that learners – subtly but persistently – transfer discourse-related word order patterns from their L1 into the L2. In my view, this evidence for transfer can and should stand independently of any particular theoretical slant. I also hope to have shown that learners can overcome pragmatically inappropriate word orders at advanced proficiency levels, but that it takes much time and input to do so.

ACKNOWLEDGEMENTS

I would like to thank the anonymous and non-anonymous *NJL* reviewers, especially Christine Dimroth, for their positive comments and helpful suggestions. Thanks also to Christina Rosén for fruitful collaboration and discussions at odd hours, and to the enthusiastic audiences in Barcelona, Canterbury, Lisbon, Lund and Tromsø, where earlier versions of this work were presented. This research was partially supported by the Erik-Philip Sörensen Foundation.

APPENDIX

In Sweden	Subjects & expletives	Objects	Temporal & locational adverbials	Other adverbials	Other constituent
3 years, 6 learners	47% (648/1380)	8% (115/1380)	21% (290/1380)	22% (304/1380)	2% (23/1380)
6 years, 6 learners	50% (564/1121)	9% (98/1121)	19% (210/1121)	21% (238/1121)	1% (11/1121)
9 years, 2 learners	37% (240/654)	8% (53/654)	23% (150/654)	29% (192/654)	3% (18/654)

Table A1. Constituents in the prefield, written L2 Swedish data, informal letters.

In Sweden	Subjects & expletives	Objects	Temporal & locational adverbials	Other adverbials	Other constituent
3 years, 6 learners	60% (1306/2179)	9% (200/2179)	24% (527/2179)	6% (125/2179)	1% (21/2179)
6 years, 6 learners	50% (962/1917)	10% (188/1917)	29% (555/1917)	9% (180/1917)	2% (32/1917)
9 years, 2 learners	49% (561/1158)	11% (130/1158)	26% (305/1158)	12% (144/1158)	2% (18/1158)

Table A2. Constituents in the prefield, informal oral L2 Swedish.

NOTES

1. The nature of 'external' and 'internal' interfaces would need to be investigated in greater detail than is currently done (see Section 2). Perhaps it is not as straightforward to distinguish between internal (inside the central grammar box) and external (outside the grammar box) as the diagram in (1) makes out. For instance, Jackendoff (2002) does not separate semantics from conceptual structure (as is done under a Chomskyan approach), but instead regards conceptual structure to be one of the three major generative components of the architecture of language, together with syntactic structure and phonological structure (2002:125, 282). These three are all connected with each other via interface components, and thus for Jackendoff the interface between conceptualisation and syntax would not be external but internal.
2. Also worth considering is the fact that bilingual speakers cannot be expected to perform like two monolinguals in the first place. On this view, very advanced L2 behaviour would not be regarded as 'deviant' from the (monolingual) target, but simply as characteristic of bilinguals.
3. This pragmatic treatment of focus as the MOST RELEVANT part(s) of the utterance from the viewpoint of the speaker is somewhat different from the treatment of focus in much of

https://doi.org/10.1017/S033258651000017X Published online by Cambridge University Press

the semantics literature (e.g. Jacobs 1984; Rooth 1992). There, an operator focus creates a set of alternatives, specifies a relation to this set of alternatives, and quantification-like, alternatives are chosen or excluded. My impression is that semanticists home in on focus types that can be captured via the set-of-alternatives approach and analysed in terms of logical relations, whilst pragmaticians often point out instances of focus which do not (seem to) specify relations to a set of alternatives, e.g. where the whole utterance is in focus, and which are only relevant at the pragmatic but not at the semantic level. Moreover, the literature abounds with proposals that split up focus into different subtypes, e.g. information focus, presentational focus, wide focus, narrow focus, identificational focus, exhaustive focus, contrastive focus, which I cannot go into here in the interests of space. 'Information focus' is sometimes used for new (non-contrastive) information, comparable to the use of the term rheme in the present paper. For overviews, see e.g. É. Kiss (1998:245–249) and Krifka (2007:18–36).

4. Halliday (1967) also adopts the theme–rheme terminology, but defines 'theme' as the leftmost constituent: 'Basically, the theme is what comes first in the clause' (1967:212); this also holds for later work of his. Defining a pragmatic function via linear position appears vacuous to me.
5. Due to effects of pressure on planning and language production in real time, there may also be processing explanations for correlating linguistic-form/discourse-function tendencies (e.g. Arnold et al. 2000). For instance, informationally old/given elements are often short and simple in form (e.g. pronouns), whilst new elements are often longer and more complex (e.g. lexical phrases, clauses). The placement of long and complex elements towards the end of the utterance may ease the burden of processing on the listener (and/or the speaker).
6. Language typologists working on word order have proposed a number of linearisation hierarchies in terms of lexical semantics, weight and discourse notions that might be relevant for the Swedish and German prefield. Typological linearisation hierarchies can be viewed as soft constraints, intended to capture the high frequency of and preference for certain surface constituent orderings by speakers of a language with variable word order. One such semantic linearisation hierarchy is the animacy constraint, according to which animate constituents occur before inanimate ones, and which is a simplified version of the person linearisation hierarchy (e.g. Silverstein 1976), according to which first person precedes second person which in turn precedes third person animate, third person inanimate, etc. Linearisation might also be determined by the lexical-semantic properties of the verb or the lexical-semantic properties of the arguments of the verb, for instance by a ranking of the semantic roles of the participants of an event, where agents precede patients, which in turn precede recipients, etc. (e.g. Siewierska 1993). Concerning weight linearisation, short, phonologically light and structurally simple constituents have been said to occur before long, phonologically heavy and structurally complex ones (e.g. Behaghel 1909/1910; Hawkins 1994). Well-known discourse-related linearisation hierarchies are definite before indefinite nominal phrases, given before new information, and topic before comment. However, it has proved difficult to reduce word order patterns to a single factor such as weight. Rather, different linearisation hierarchies may be at work simultaneously and work with or against each other. Also, a particular linearisation hierarchy might be stronger in one language than in another, or only be a determinant in a particular area of that language (e.g. concerning constituent ordering in embedded clauses vs. main clauses). Recent work in corpus linguistics implements these typological proposals to see how well certain linearisation hierarchies can account for word order variation in corpus data from a specific language. Concerning German, the ordering of

subjects and objects relative to each other has been studied, but as yet only for the genre of newspaper articles in electronically searchable corpora, and only for a small subset of objects that are easily identifiable (via overtly case-marked items). Despite these obvious limitations in scope, it is interesting to see that for some clausal positions, word order variation can be captured relatively well by some of the proposed linearisation hierarchies, but only badly for others. For instance, Bader & Häussler (2010) find a strong correlation of word order with two semantic hierarchies (animacy and agenthood) for the German midfield, i.e. for the postverbal ordering of subject and object relative to each other. Yet these semantic linearisation hierarchies do not play much of a role for the prefield – here, agenthood and animacy do not determine whether the subject or the object is placed in the prefield (preverbally) or in the midfield (postverbally). Bader & Häussler (2010:753–758) conclude that word order variation in the German prefield is not primarily determined by lexical semantic factors, but predominantly by discourse-related constraints (e.g. topic first). For similar observations, see Weber & Müller (2004) and Bouma (2008).

7. These differences are statistically significant for subjects and expletives ($\chi^2 = 75.797$, $p < .001$), objects ($\chi^2 = 15.216$, $p < .001$) and other adverbials ($\chi^2 = 58.951$, $p < .001$), but not for temporal and locational adverbials. Figure 2 compares only informal letters in order to avoid any potential confounding effects that different text types might cause.
8. Jörgensen (1976) does not provide counts for the different subclasses of adverbials in the spoken Swedish data. All adverbials in Figure 3 are therefore combined into one bar. The percentages cited are from Jörgensen (1976:101–105); he also provides totals for the declarative clauses out of which these percentages are calculated (pp. 70–71, 138–139). Jörgensen's detailed appendix (pp. 158–159) has allowed me to calculate the raw figures for prefield constituent types, though I have not been able to reconstruct his exact totals for the declarative clauses, but only approximations thereof. This is why the percentages in the bar chart in Figure 3 are not backed up with exact raw figures in the present paper. To do so, one would have to go through the entire Talbanken corpus manually.
9. Prefield objects are thus more frequent in informal spoken Swedish than in written Swedish, and on a par with German. Once subtypes of objects are considered however, there are some noteworthy differences, and the relatively high rate of prefield objects in spoken Swedish turns out to be entirely due to inanimate pronominal object *det* 'it/that' (see below).
10. Rahkonen (2000, 2006), investigating objects in written Swedish, also finds anaphoric inanimate *det* to be by far the most commonly fronted object, whereas animate pronominal objects hardly ever occur in clause-initial position.
11. Aafke Hulk (p.c., 21 June 2009) informs me that fronted animate personal pronouns in Dutch receive a focus reading too.
12. Factors outside discourse pragmatics may also contribute to the Swedish dislike of fronted unaccented animate object pronouns. For instance, Rahkonen (2000) suggests that the lexical-semantic factor of animacy plays a role because it helps us to differentiate the grammatical function of constituents. For two corpora of formal written Swedish (Gymnasistsvenska, i.e. school leavers' argumentative essays from the 1970s (Talbanken)) and a portion of the mixed-genre Stockholm-Umeå Corpus of the 1990s (SUC), Rahkonen finds that inanimate objects disproportionately often appear in the prefield, whilst animate personal object pronouns hardly ever do. Bouma (2008:256–262) has recently noted the same for a corpus of written Dutch. Both Rahkonen and Bouma speculate that an animate object in the prefield might mislead the listener into mistaking it for an agent subject (since animates are often agents and agents are often subjects), whilst an

inanimate object in the prefield will not mislead the listener (but be processed as non-agent straightaway). As speakers want to ease communication, they will happily front subjects and inanimate objects, but shirk away from fronting animate objects and animate personal object pronouns. Rahkonen's Swedish and Bouma's Dutch results may be taken to support this idea, but they do not appear to carry over to German, since there animate objects do occur in the prefield, see e.g. Bader & Häussler (2010).

13. TISUS (Test in Swedish for University Studies) is a two-day examination, testing reading comprehension, written composition, and oral communication in an interview with two examiners.
14. Data collection at nine years after arrival was planned for two of these learners during 2010.
15. The South German recordings were originally made with a different research topic (i.e. pronoun drop) in mind, but can also be investigated for clause-initial word order variation. The age difference between the 25–35 and 60–70-year-old informants might be a concern. I therefore analysed the productions of each individual separately, but could not detect any noticeable differences between the younger and older informants concerning the prefield.
16. There are more subject-initial clauses (60%) in the learners' oral data at 3 years than at any other data point. This is due to a preponderance of first person singular subjects (*jag* 'I') in some of the data files because of the way the interviewer interacted with the informant.
17. Moreover, the learners' choice of lexical items, grammatical gender and inflectional morphology sometimes differs from native Swedish. Such non-target features will not be commented on here.
18. This is reminiscent of Camacho's (1999) findings on L1 Peruvian Quechua speakers acquiring L2 Spanish. In Quechua, left-peripheral objects without clitics are information-structurally 'neutral' (i.e. they do not evoke narrow/contrastive focus). Transferring this neutral L1 pattern to the L2, the learners produced objects without clitics in the left periphery in their Spanish, unaware that native Spanish listeners assign a contrastive focus reading to them that is unintended by the L2 speaker.
19. It may, however, be the case that experiments that go beyond naturalistic production data would reveal a discrepancy between the learners and native speakers, e.g. with regard to acceptability judgments or processing in real time. This is a topic for future research.
20. Thanks to Christine Dimroth for suggesting the term discourse accent.

REFERENCES

Ammann, Hermann. 1928. *Die menschliche Rede. Sprachphilosophische Untersuchungen 2. Teil: Der Satz*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse structure on constituent ordering. *Language* 76, 28–55.

Bader, Markus & Jana Häussler. 2010. Word order in German: A corpus study. *Lingua* 120, 717–762.

Behaghel, Otto. 1909/1910. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25, 110–142.

Belletti, Adriana (ed.). 2004. *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3. New York: Oxford University Press.

Belletti, Adriana, Elisa Bennati & Antonella Sorace. 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory* 25, 657–689.

Beneš, Eduard. 1971. Die Besetzung der ersten Position im deutschen Aussagesatz. *Fragen der strukturellen Syntax und der kontrastiven Grammatik*, 160–182. Düsseldorf: Schwann.

Benincà, Paola & Cecilia Poletto. 2004. Topic, focus, and V2: Defining the CP sublayers. In Luigi Rizzi (ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures*, vol. 2, 52–75. Oxford: Oxford University Press.

Bohnacker, Ute. 2005. Nonnative acquisition of Verb Second: On the empirical underpinnings of universal L2 claims. In Marcel den Dikken & Christina Tortora (eds.), *The Function of Function Words and Functional Categories*, 41–77. Amsterdam: John Benjamins.

Bohnacker, Ute. 2006. When Swedes begin to learn German: From V2 to V2. *Second Language Research* 22(4), 1–44.

Bohnacker, Ute. 2007. On the 'vulnerability' of syntactic domains in Swedish and German. *Language Acquisition* 14(1), 1–43.

Bohnacker, Ute & Christina Rosén. 2007. How to start a declarative V2 clause: Transfer of syntax or information structure in L2 German. *Nordlyd* 34(3), 29–56.

Bohnacker, Ute & Christina Rosén. 2008. The clause-initial position in L2 German declaratives: Transfer of information structure. *Studies in Second Language Acquisition* 30(4), 511–538.

Bouma, Gerlof. 2008. *Starting a Sentence in Dutch: A Corpus Study of Subject- and Object-fronting*. Groningen: University of Groningen.

Brody, Michael. 1990. Some remarks on the focus field in Hungarian. *UCL Working Papers in Linguistics* 2, 201–225.

Bühler, Karl. 1934. *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.

Camacho, José. 1999. From SOV to SVO. *Second Language Research* 15, 115–132.

Carroll, Mary & Monique Lambert. 2003. Information structure in narratives and the role of grammaticised knowledge. In Christine Dimroth & Marianne Starren (eds.), *Information Structure and the Dynamics of Language Acquisition*, 268–287. Amsterdam: John Benjamins.

Carroll, Mary & Christiane von Stutterheim. 1993. The representation of spatial configurations in English and German and the grammatical structure of locative and anaphoric expressions. *Linguistics* 31, 1011–1041.

Carroll, Mary & Christiane von Stutterheim. 2003. Typology and information organisation: Perspective taking and language-specific effects in the construal of events. In Anna Giacolone Ramat (ed.), *Typology and Second Language Acquisition*, 365–402. Berlin & New York: Mouton de Gruyter.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and Topic*, 25–55. New York: Academic Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. A Minimalist program for linguistic theory. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, 1–52. Cambridge, MA: MIT Press.

Daneš, František. 1970. Zur linguistischen Analyse der Textstruktur. *Folia Linguistica* 4, 72–79.

Drach, Erich. 1937. *Grundgedanken der deutschen Satzlehre*. Frankfurt: Diesterweg.
Dyhr, M. 1978. *Die Satzspaltung im Dänischen und Deutschen. Eine kontrastive Analyse*. Tübingen: Narr.
Ekberg, Lena. 1997. Diskursiva och syntaktiska mönster i återberättelser hos invandrarbarn i Rosengård. In Gisela Håkansson, Lena Lötmarker, Lillemor Santesson, Jan Svensson & Åke Viberg (eds.), *Svenskans beskrivning 22. Förhandlingar vid Tjugoandra sammankomsten för svenskans beskrivning Lund den 18–19 oktober 1996*, 97–110. Lund: Lund University Press.
Ekerot, Lars-Johan. 1979. Syntax och informationsstruktur. In Kenneth Hyltenstam (ed.), *Svenska i invandrarperspektiv*, 79–108. Lund: Liber Läromedel.
É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2), 245–273.
Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax–Discourse Interface*. Oxford: Oxford University Press.
Fanselow, Gisbert. 2007. The restricted access of information structure to syntax: A minority report. In Féry et al., 205–220.
Féry, Caroline. 2007. Information structural notions and the fallacy of invariant correlates. In Féry et al. (eds.), 161–184.
Féry, Caroline, Gisbert Fanselow & Manfred Krifka (eds.) 2007. *Interdisciplinary Studies on Information Structure* 6, 161–184.
Grewendorf, Günther. 1988. *Aspekte der deutschen Syntax*. Tübingen: Narr.
Gundel, Jeanette K. 2002. Information structure and the use of cleft sentences in English and Norwegian. In Hilde Hasselgård, Stig Johansson, Bergljot Behrens & Cathrine Fabricius-Hansen (eds.), *Information Structure in a Cross-linguistic Perspective*, 113–128. Amsterdam & New York: Rodopi.
Håkansson, Gisela. 1997. Barnets väg till svensk syntax. In Ragnhild Söderbergh (ed.), *Från joller till läsning och skrivning*, 47–60. Malmö: Gleerups.
Halliday, M. A. K. 1967. Notes on transitivity and theme in English. Part 1 and 2. *Journal of Linguistics* 3, 37–81 & 199–244.
Halliday, M. A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76(4), 891–920.
Hertel, Tammy Jandrey. 2003. Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research* 19, 273–304.
Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Hueber.
Hockett, Charles. 1958. *A Course in Modern Linguistics*. New York: MacMillan.
Huber, Stefan. 2002. Es-*Clefts und* det-*Clefts*. Stockholm: Almqvist & Wiksell.
Hulk, Aafke & Natascha Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3, 227–244.
Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.
Jacobs, Joachim. 1984. Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91, 25–58.
Johansson, Mats. 2001. Clefts in contrast: A contrastive study of *it* clefts and wh clefts in English and Swedish texts and translations. *Linguistics* 39, 541–562.
Johansson, Stig, 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam & Philadelphia: John Benjamins.

Jörgensen, Nils. 1976. *Meningsbyggnaden i talad svenska*. Lund: Studentlitteratur.
Kiese, Jörn. 1993. *Fokussierende Sätze im Deutschen und Englischen*. Frankfurt: Peter Lang.
Krashen, Stephen. 2003. *Explorations in Language Acquisition and Use*. Porthsmouth, NH: Heinemann.
Krifka, Manfred. 2007. Basic notions of information structure. In Féry et al. (eds.), 13–55.
Kroll, Barbara (ed.). 2003. *Exploring the Dynamics of Second Language Writing*. New York: Cambridge University Press.
Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago: University of Chicago Press.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
Liceras, Juana. 1988. Syntax and stylistics: More on the pro-drop parameter. In James Pankhurst, Michael Sharwood-Smith & Paul Van Buren (eds.), *Learnability and Second Languages: A Book of Readings*, 71–93. Dordrecht: Foris.
Lozano, Cristóbal. 2006. Focus and split-transitivity: The acquisition of word order alternations in non-native Spanish. *Second Language Research* 22, 145–187.
Mauranen, Anna. 1996. Discourse competence: Evidence from thematic development in native and non-native texts. In Eija Ventola & Anna Mauranen (eds.), *Academic Writing*, 195–230. Amsterdam & Philadelphia: John Benjamins.
Neeleman, Ad & Hans van de Koot. 2008. Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Linguistics* 11(2), 137–189.
Pérez-Leroux, Ana Teresa & William Glass. 1999. Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research* 15, 220–249.
Platzack, Christer. 2001. The vulnerable C-domain. *Brain and Language* 77, 364–377.
Polinsky, Maria & Robert Kluender. 2007. Linguistic typology and theory construction: Common challenges ahead. *Linguistic Typology* 11, 273–283.
Prince, Ellen. 1981. Toward a taxonomy of given–new information. In Peter Cole (ed.), *Radical Pragmatics*, 249–264. New York: Academic Press.
Prince, Ellen. 1998. On the limits of syntax, with reference to left dislocation and topicalization. In Peter [W.] Culicover & Louise McNally (eds.), *The Limits of Syntax* (Syntax and Semantics 39), 281–302. New York: Academic Press.
Rahkonen, Matti. 2000. Topikalisering av icke-subjekt i svenska påståendesatser. *Språk och stil* 10, 37–85.
Rahkonen, Matti. 2006. Some aspects of topicalization in active Swedish declaratives: A quantitative corpus study. *Linguistics* 44(1), 23–55.
Reinhart, Tanya. 1982. Pragmatics and linguistics: A analysis of sentence topic. Bloomington, IN: Indiana University Linguistics Club, 1–38.
Reis, Marga. 1980. On justifying topological frames: 'Positional field' and the order of nonverbal constituents in German. *Documentation et Recherche en Linguistique Allemande Contemporaine* 22/23, 59–85.
Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar: Handbook of Generative Syntax*, 281–337. Dordrecht: Kluwer.
Robach, Inger-Britt. 2003. Satsklyvning i franska och svenska. In Eva Larsson Ringqvist (ed.), *Ordföljd och informationsstruktur i franska och svenska*, 45–72. Växjö: Växjö University Press.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1, 75–116.
Rosén, Christina. 2006. *Warum klingt das nicht deutsch? Probleme der Informationsstrukturierung in deutschen Texten schwedischer Schüler und Studenten*. Stockholm: Almqvist & Wiksell.
Rothman, Jason. 2009. Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax–pragmatics interface. *Journal of Pragmatics* 41, 951–973.
Siewierska, Anna. 1993. On the interplay of factors in the determination of word order. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, 826–846. Berlin: Walter de Gruyter.
Silverstein, Michael, 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
Sorace, Antonella. 2005. Syntactic optionality at interfaces. In Leonie Cornips & Karen Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 55–80. Amsterdam: John Benjamins.
Sorace, Antonella & Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22, 339–368.
Sorace, Antonella & Ludovica Serratrice. 2009. Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism* 13(2), 195–210.
Stemmer, Brigitte, 1999. An on-line interview with Noam Chomsky: On the nature of pragmatics and related issues. *Brain and Language* 68(3), 393–401.
Teleman, Ulf, Staffan Hellberg & Erik Andersson (eds.). 1999. *Svenska Akademiens grammatik*, vol. 4: *Satser och meningar*. Stockholm: NordstedtsOrdbok.
Tsimpli, Ianthi & Antonella Sorace. 2006. Differentiating interfaces: L2 performance in syntax–semantics and syntax–discourse phenomena. In David Bamman, Tatiana Magnitskaia & Colleen Zaller (eds.), *30th Boston University Conference on Language Development*, vol. 2, 653–664. Somerville, MA: Cascadilla Press.
Tomlin, Russell S., Linda Forrest, Ming Ming Pu & Myung Hee Kim. 1997. Discourse semantics. In Teun A. van Dijk (ed.), *Discourse as Structure and Process*, 63–111. London: Sage.
Valenzuela, Elena. 2006. L2 end state grammars and incomplete acquisition of Spanish CLLD constructions. In Roumyana Slabakova, Silvina Montrul & Philippe Prévost (eds.), *Inquiries in Linguistics Development*, 283–304. Amsterdam: John Benjamins.
von Stutterheim, Christiane. 2003. Linguistic structure and information organisation: The case of very advanced learners. In Susan Foster-Cohen & Simone Pekarek Doehler (eds.), *EUROSLA Yearbook* 3, 183–206. Amsterdam: John Benjamins.
von Stutterheim, Christiane & Wolfgang Klein. 1989. Text structure and referential movement. In Rainer Dietrich & Carl F. Graumann (eds.), *Language Processing in Social Context*, 39–76. Amsterdam: North Holland.
Weber, Andrea & Karin Müller. 2004. Word order variation in German main clauses: A corpus analysis. *20th International Conference on Computational Linguistics*, 71–77, Geneva.
White, Lydia. 2009. Grammatical theory: Interfaces and L2 knowledge. In William C. Ritchie & Taj K. Bhatia (eds.), *The New Handbook of Second Language Acquisition*, 49–68. Leeds: Emerald.
Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker et al. (eds.). 1997. *Grammatik der deutschen Sprache*, vol. 2. Berlin & New York: Walter de Gruyter.