

RESEARCH

Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account

Thomas Hörberg^{1,2}¹ Department of Linguistics, Stockholm University, SE-106 91 Stockholm, SE² Department of Psychology, Stockholm University, SE-106 91 Stockholm, SE

thomas.hoerberg@psychology.su.se

In Swedish, grammatical functions are primarily encoded by word order. In prototypical transitive sentences, the subject precedes the direct object. However, Swedish also allows for fronting of the direct object, although such sentences are potentially ambiguous with respect to grammatical functions. This study therefore investigates direct object fronting in written Swedish with respect to 1) which functions this construction serves and 2) whether the use of direct object fronting is dispreferred when the grammatical functions cannot be determined on other information types. These questions are investigated on the basis of quantitative differences in the distribution of NP prominence properties (e.g., givenness and animacy) and formal, morphosyntactic cues to grammatical functions (e.g., case marking and verb particles) between OVS and SVO sentences, and between OVS sentences and passives. The results indicate that direct object fronting is used when the object either is topical and highly discourse prominent, or when it is contrastive. I also argue that direct object fronting is used to introduce new topics into the discourse. Subjects are more frequently high in discourse prominence in object-initial sentences than in subject-initial sentences. I suggest that this stems from a motivation to keep the information in object-initial sentences following the sentence-initial object “informationally light” and predictable. Unambiguous formal markers of grammatical functions are used more frequently in OVS sentences than in SVO sentences, but less frequently in passives than in SVO sentences. OVS sentences also more frequently contain an animate subject and an inanimate object than SVO sentences, and in passives, animate subjects and inanimate objects are even less frequent. Writers therefore seem to prefer the structurally unambiguous passive construction over the potentially ambiguous object-initial construction, when grammatical functions cannot be determined on the basis of other formal markers or an NP argument animacy difference. Further, sentences with two animate arguments more frequently contain formal markers than sentences with at most one animate argument. These findings indicate that writers actively avoid direct object fronting when it potentially results in an ambiguity, and provide evidence for the hypothesis that writers are inclined to actively avoid ambiguities more generally.

Keywords: object-initial word order; topicalisation; argument prominence; corpus linguistics; ambiguity avoidance; Swedish

1 Introduction

An important aspect of both spoken and written communication is the ability to express how participants and entities are involved in activities and events, such as being responsible for an activity or being affected by it. In transitive sentences, such information is provided by the grammatical functions of the NP arguments. In the prototypical case, the subject NP refers to the participant responsible for the action denoted by the verb, and the direct object NP (henceforth simply referred to as the *object*) to the participant or entity that is affected by that action. Information about grammatical functions is

in many languages encoded with word order or morphology, such as case marking or agreement.

In Swedish, word order is of particular importance. Subject-initial, SVO word order is by far the most common order in Swedish transitive sentences, and the grammatical functions of NP arguments in transitive sentences are in most cases assigned on the basis of their relative ordering (i.e., on the assumption that the subject is the initial NP of the sentence). However, although uncommon, the object may also be placed sentence-initially. Since Swedish is a verb-second language, sentence-initial objects must be directly followed by the finite verb (Teleman, Hellberg & Andersson 1999 (4): 688; but see Josefsson 2012 for exceptions). The word order is therefore OVS in object-initial sentences. Such sentences may be locally ambiguous with respect to grammatical functions. Consider, for instance, the following sentences in Swedish:

- (1) a. Tjejen gillar jag inte.
 girl.the like 1SG.NOM not
 'I don't like the girl.'
- b. Dig gillar jag inte.
 2SG.ACC like 1SG.NOM not
 'I don't like you.'

In (1a), the initial NP consists of a noun. Since nouns lack case marking in Swedish, it is morphologically ambiguous with respect to its grammatical function. Because of the high prevalence of sentence-initial subjects in Swedish, however, the initial NP is preferably interpreted as the subject. The second NP consists of a first-person subject pronoun and is therefore morphologically unambiguous regarding its grammatical function. The initial interpretation of the sentence as subject-initial can therefore not be maintained once the second NP is encountered, and the sentence must be re-interpreted as object-initial. In (1b), on the other hand, the initial NP consists of an unambiguous object pronoun, and the sentence must be interpreted as object-initial directly. In contrast to (1b), (1a) is therefore locally ambiguous with respect to grammatical functions. On the basis of an experimental study using event-related brain potentials, Hörberg, Kallioinen and Tamm (2013) showed that written OVS sentences such as (1a) are problematic to interpret, in that readers need to revise their initial interpretation of the sentence as subject-initial upon encountering the disambiguating post-verbal subject NP. Yet, both speakers and writers occasionally do use OVS sentences. It is therefore of interest to investigate the conditions under which OVS word order is used in Swedish, and whether OVS word order is dispreferred when the grammatical functions cannot be determined on other information types.

To this end, the present corpus study investigates quantitative differences in the distribution of NP properties, such as animacy and givenness, and morphosyntactic sentence properties, such as the presence of auxiliary verbs and verb particles. This is done by comparing OVS sentences to SVO sentences and passives in different text genres. In particular, the study aims to shed further light on the following two research questions:

1. How does the distribution of NP properties of animacy, person, definiteness and givenness differ between OVS and SVO sentences? What does this say about when OVS word order is used in written Swedish, and, by extension, what functions OVS word order serve?
2. Do writers more frequently provide formal, morphosyntactic information regarding grammatical functions (e.g., case marking) in OVS sentences in

comparison to SVO sentences, on the one hand, and in comparison to passives, on the other?

In the following, I give a brief overview of earlier accounts of the functional motivations behind object fronting (Section 2.1), and continue with a summary of studies investigating whether both speakers and writers adopt their productions in order to avoid ambiguities (Section 2.2). I then give a summary of relevant aspects of Swedish grammar, starting with the sentence structure of SVO, OVS and passive sentences (Section 2.3.1), and then present means of identifying grammatical functions in Swedish (Section 2.3.2). In Section 2.4, I give an overview of previous corpus studies investigating Swedish with respect to object fronting (Section 2.4.1), on the one hand, and with regard to ambiguity avoidance in object-initial sentences, on the other (Section 2.4.2). Finally, I present the method of the present study (Section 3), its findings (Section 4), and discuss their implications (Section 5).

2 Background

2.1 Grammatical functions and the function of the object-initial construction

Grammatical functions in transitive sentences are commonly assumed to express or to be associated with role-semantic properties of NP arguments (e.g., Foley 2011; Hörberg 2016: 9–10), henceforth referred to as *participant roles*. Participant roles concern the kind of involvement that the NP argument referents have in the event expressed by the sentence. Traditionally, notions such as *agent* and *patient* have been used to define participant roles (e.g., Fillmore 1968; Chomsky 1981). However, the degree to which participant roles are associated with agent or patient properties ultimately depend on the kind of event that is expressed by the main verb of the sentence. It is therefore not always possible to clearly and unambiguously assign a particular role to an argument (see Dowty 1991). Participant roles can instead be conceived as cluster or prototypicality concepts. In transitive sentences, the main verb may assign both agent and patient properties to the NP arguments, but the arguments will always be differentiated from each other in terms of possessing the most agent or patient properties, respectively. Participant roles can therefore be subsumed under two general categories, the *Actor* and the *Undergoer* role (see, e.g., Foley & Van Valin 1984; Dowty 1991; Van Valin & LaPolla 1997; Van Valin 2005; Primus 2006; Bornkessel-Schlesewsky & Schlewsky 2009; Bickel 2010), the former being expressed by the subject, and the latter by the object. As such, the subject NP refers to the participant that is responsible for the activity expressed by the verb, and the object NP to the participant or entity that is affected by that activity.

Many verbs require an Actor that is volitional and/or sentient, and therefore human or animate, but do not put the same constraints on the Undergoer role (Van Valin & LaPolla 1997: 305; Dahl 2008). Animacy is therefore strongly correlated with grammatical functions in natural discourse. As a multitude of corpus studies show, subjects are much more frequently animate than objects (see Dahl & Fraurud 1996 and Dahl 2000 for Swedish; Kempen & Harbusch 2004 for German; Øvrelid 2004 for Norwegian; Bouma 2008 for Dutch). For instance, in his corpus of spoken Swedish, Dahl (2000) found that 93.2% of the subjects and 9.9% of the objects were animate. According to Dahl (2008: 145), animacy is “important for determining what can be said about an entity” in the sense that many predicates require an animate Actor argument. Animacy can therefore function as a semantic cue to grammatical functions, and thereby render overt formal marking of grammatical functions (such as case marking or other morphosyntactic information) redundant. The present study will test this assumption by investigating whether writers less frequently use formal markers of grammatical functions when grammatical functions can be determined on the basis of an animacy difference between two NPs.

Grammatical functions are also related to the information structure of individual sentences in a discourse. Information structure concerns how writers and speakers structure their sentences to their addressees, with respect to the information provided in the previous discourse and in the situational context. Most (but not all) declarative sentences are used to express an assertion about some referent that is assumed to be known by the addressee. This referent is expressed by the sentence *topic* (e.g., Reinhart 1981; Lambrecht 1994; Erteschik-Shir 2007). On some views, topics are seen as a means of storing information about the discourse model (Reinhart 1981; Erteschik-Shir 2007). Accordingly, topics serve as referential entries under which propositions are stored. Crucially, what constitutes the topic of a sentence depends on the discourse that the sentence occurs in. Two equivalent sentences may have different topics depending on the context. For example, as pointed out by Reinhart (1981), in a sentence such as

(2) Max saw Rosa yesterday.

Max constitutes the topic when (2) is answered to the question *Who did Max see yesterday?*, but *Rosa* is the topic when it is the answer to *Did anybody see Rosa yesterday?*. The information of a declarative sentence that the writer or speaker assumes to be new to the addressee, on the other hand, is called the sentence *focus* (Lambrecht 1994; Gundel & Fretheim 2004; Erteschik-Shir 2007). The focus can be seen as new information that is added to the discourse model by the sentence at hand. What is considered to be focus of a sentence is also context-dependent. When (2) is answered to *Who did Max see yesterday?* then *Rosa* is focus, and when (2) is the answer to *Did anybody see Rosa yesterday?*, *Max* is focus. Focus is often conceived of as highlighting of information that stands in opposition with, or is contrasted against, a (possibly open-ended) set of information¹ that potentially could constitute the focused information (e.g., Rooth 1992; Gundel & Fretheim 2004; Molnár 2006; Erteschik-Shir 2007; Krifka 2007; Molnár & Winkler 2010). In (2), for example, the focused referent (Max or Rosa) can be seen to be contrasted against a possibly open-ended set of other individuals. However, there are good reasons to differentiate the notion of focus from the notion of *contrast*, which involves contrasting some information against a limited number of contextually given alternatives (Molnár 2002; 2006; Molnár & Winkler 2010). Both topics and foci can be contrastive, and, as illustrated in Examples (3) and (4) from Prince (1984), English topicalisation is in some situations only valid when the topicalised information is contrasted against a limited number of contextually determined alternatives:

(3) A: Why are you laughing?
B: #Annie Hall I saw yesterday. I was just thinking about it.

(4) A: You see every Woody Allen movie as soon as it comes out.
B: No – Annie Hall I saw (only) yesterday.

Whereas the positioning of the NP *Annie Hall* in the sentence-initial position is inappropriate in (3), in (4), where the NP *Annie Hall* is contrasted against all other Woody Allen movies, the NP may be positioned sentence-initially. Since *Annie Hall* is contrastive in (4), it may serve as the topic of the sentence. Contrast therefore appears to be an information structural notion that is orthogonal to topic and focus, but that shares features with both.

¹ According to Rooth (1992), this “set of information” would constitute “the set of propositions obtainable [...] by making a substitution in the position corresponding to the focused phrase”.

As with focus, it involves the highlighting of information that it is in opposition to, and as with topic, it facilitates discourse coherence in that it relates information to the previous discourse context (Molnár 2002; 2006; Molnár & Winkler 2010).

Topical NPs typically refer to entities that are presupposed, either by virtue of being introduced earlier on in the discourse, or by being known by the interlocutors. Topical referents tend to be maintained through longer stretches of discourse through the use of anaphoric expressions (i.e., *topic-chaining* or *topic-continuity*, see Givón 1983; Erteschik-Shir 2007: 44–45; Engdahl & Lindahl 2014). NPs that are part of the sentence focus, on the other hand, more frequently refer to entities that are unknown by the addressee and that are new in the discourse. NP referents which are presupposed are sometimes called *given*, and referents which are discourse new and assumed to be unknown, are called *new*. Some suggest that topical arguments always are given (Erteschik-Shir 2007: 20), others that they at least have to be familiar to the addressee (Lambrecht 1994: 262). Focused arguments, on the other hand, tend to be new since they often introduce new referents into the discourse (Lambrecht 1994: 262).

This *givenness* distinction is generally assumed to concern the cognitive status or accessibility of NP referents in the discourse model, and therefore only to be indirectly related to linguistic expressions (Ariel 1990; Gundel et al. 1993; Gundel & Fretheim 2004). However, the givenness status of a referent is often reflected in the form of NP arguments (Ariel 1990: 69; Gundel et al. 1993; Lambrecht 1994: 105). For example, whereas NPs consisting of personal pronouns or definite nouns refer to entities that are active in the discourse, or assumed to be known by the addressees, indefinite NPs refer to new entities which have not been previously introduced into the discourse. The linguistic features definiteness, pronominality and person are therefore highly correlated with givenness. Definiteness is used to signal whether a referent is uniquely identifiable in the discourse, either by virtue of being anaphoric or through association (Lyons 1999). Definite pronouns tend to be anaphoric and refer to highly discourse prominent referents that have already been introduced into discourse. First- and second-person pronouns are inherently high in givenness in the discourse by virtue of referring to the speaker/writer, and the listener/reader, respectively.

Although there is no one-to-one mapping between subjects and topics (as Example (1) illustrates), sentence-initial subject NPs constitute the topic in most sentences, and object NPs are more frequently part of the sentence focus (Reinhart 1981; Lambrecht 1994; Erteschik-Shir 2007). There is therefore a correlation between grammatical functions and the givenness of NP referents, which in turn are reflected in their forms. Several corpus studies have therefore found that subjects are much more frequently definite, first/second person, and pronominal than objects (see Dahl 2000 for Swedish; Øvrelid 2004 for Norwegian; Bouma 2008 for Dutch; and Du Bois 2003 for a review of studies on differences in referentiality and givenness between subjects and objects). For example, Dahl (2000) found that 60.7% of the subjects but only 2% of the objects were first- or second-person pronouns.

The object-initial construction is generally assumed to be used when it is the object referent, that is, the Undergoer, rather than the subject referent, that is, the Actor, that is topical. Since topical NPs also tend to be highly discourse prominent, several corpus studies therefore show that sentence-initial objects frequently are high in discourse prominence in terms of pronominality and definiteness, in comparison to objects positioned in their canonical position (See Weber & Müller 2004 for German; Øvrelid 2004 for Norwegian; Snider & Zaenen 2006 for English; Bouma 2008 for Dutch). The present study will investigate whether this also holds in written Swedish. More specifically, I will investigate whether topicalisation is the only motivation for positioning the object in the sentence-initial position,

as implied by earlier investigations of object-fronting in written Swedish (Rahkonen 2006; Bohnacker & Rosen 2008; 2009; Bohnacker 2010). Before presenting these previous findings on object fronting in Swedish, I first give a summary of studies investigating whether speakers and writers adapt their productions in order to avoid ambiguities, and provide a short presentation of relevant aspects of Swedish grammar.

2.2 *The object-initial construction and ambiguity avoidance*

As illustrated in Section 1, object-initial sentences in Swedish are potentially ambiguous with respect to grammatical functions. A question that arises, therefore, is whether speakers and writers are more inclined to provide unambiguous information regarding grammatical functions in OVS sentences in comparison to SVO sentences or other unambiguous alternatives, such as the passive. The passive is similar to the object-initial construction in that it, too, can be used when it is the Undergoer rather than the Actor that is the topical participant (e.g., Givón 1979; Pitz 2006; Foley 2011). In the passive, the Undergoer is realised as an intransitive subject NP, and the Actor argument is optionally realised as the argument of a prepositional *av* ('by') phrase. The passive therefore differs from the object-initial construction in Swedish with respect to potential ambiguity. Whereas OVS sentences potentially are ambiguous with respect to grammatical functions, in the passive, the initial NP can only function as an (intransitive) subject which expresses the Undergoer role. It is therefore of interest to investigate whether writer's choice between the object-initial construction and the passive is influenced by whether the grammatical functions of the NPs can be determined on other morphosyntactic or semantic information.

Results from previous studies regarding whether speakers and writers tend to avoid ambiguities during speaking and writing are mixed. On the one hand, several studies have either directly or indirectly provided evidence for ambiguity avoidance. For instance, some studies have found that the animacy difference between subjects and objects is more pronounced in object-initial sentences in both written and spoken language (Øvrelid 2004; Snider & Zaenen 2006; Bouma 2008; Bader & Häussler 2010). That is, speakers and writers more frequently use the potentially ambiguous OVS word order when the grammatical functions of the NPs can be determined on the basis of an argument animacy difference. It has also been found that speakers more frequently use morphological information regarding the argument functions in situations where they cannot readily be inferred on the basis of referential, semantic or plausibility information. In Korean and Japanese, overt realization of morphological case marking is optional in colloquial speech (Lee 2006; Kurumada & Jaeger 2015). In a corpus study of spoken Korean, Lee (2006) found overt nominative case marking to occur more frequently on subjects that are inanimate or low in discourse prominence in terms of person and definiteness. Overt accusative case marking was, on the other hand, more frequent for animate or discourse prominent objects (Lee 2006). In a sentence recall study, Kurumada and Jaeger (2015) showed that Japanese speakers more frequently use overt object case marking in transitive sentences with two animate NP arguments, and in semantically implausible sentences in which the assignment of participant roles is unexpected (e.g., *The criminal arrested the police officer*). These results suggest that speakers are inclined to provide additional morphological information regarding grammatical functions in situations where they are harder to determine on the basis of other information types. In a corpus study of written English, Temperley (2003) found the use of optional relative pronouns and complementisers to be more frequent in situations where their omission potentially would result in a syntactic ambiguity.²

² Such as in a sentence as *The lawyer who Smith hired was fired* in which the omission of the relative pronoun *who* would make it possible to initially interpret the proper name *Smith* as the head matrix clause subject (i.e., *The lawyer Smith*) rather than the subject of the object relative clause.

Haywood et al. (2005) found participants engaged in a dialogue task to more frequently avoid preposition phrase attachment ambiguities in situations where such ambiguities would potentially confuse their addressees.³ Taken together, these results indicate that both speakers and writers actively avoid ambiguities in order to accommodate the understanding of their addressees.

Several other studies have, however, failed to find any evidence for ambiguity avoidance (see, e.g., Ferreira & Dell 2000; Arnold et al. 2004; Jaeger 2006; 2010; Roland et al. 2006; and Ferreira 2008 for a review). For instance, in a spoken sentence production experiment, Ferreira and Dell (2000) did not find the optional complementiser *that* to be used more frequently in cases where its omission would lead to a direct object-subject ambiguity.⁴ Similar findings were done by Jaeger (2010) on the basis of spoken corpus data. In contrast to the findings of Haywood et al. (2005), Arnold et al. (2004) did not find any tendency for participants in an on-line production experiment to avoid prepositional phrase attachment ambiguities. It has been suggested that ambiguities that appear to be problematic in isolated sentences might in most cases be unproblematic at the discourse level (Rahkonen 2006; Ferreira 2008). There is also some evidence suggesting that sentence level ambiguities in fact can be beneficial for communication when they can be resolved in context (Piantadosi et al. 2012).

Because of these contradicting findings, it is still unclear whether people do actively avoid sentence level ambiguities. This study therefore investigates whether there is a tendency for writers to avoid potentially ambiguous object-initial sentences such as (1a).

2.3 Sentence structure and grammatical functions in Swedish

2.3.1 Subject-initial, object-initial and passive sentences in Swedish

The three sentence types investigated in this study are subject-initial sentences with SVO-word order (Example (5)), object-initial sentences with OVS-word order (Example (6)), and passive sentences in which the Undergoer is realised as the subject and the Actor is overtly mentioned as the argument of a prepositional *av* ('by') phrase (Example (7)).

- (5) a. Barnen får inte äta upp all glass innan middan.
children.the can not eat up all ice-cream before dinner
'The children can't eat all the ice cream before dinner.'
- b.⁵ Barnen inte får äta upp all glass innan middan.
children.the not can eat up all ice-cream before dinner
'The children can't eat all the ice cream before dinner.'
- (6) a. All glass får barnen inte äta upp innan middan.
all ice-cream can children.the not eat up before dinner
'The children can't eat all the ice cream before dinner.'
- b. All glass får inte barnen äta upp innan middan.
all ice-cream can not children.the eat up before dinner
'The children can't eat all the ice cream before dinner.'
- (7) a. All glass får inte ätas upp av barnen innan middan.
all ice-cream can not eaten up by children.the before dinner
'The children can't eat all the ice cream before dinner.'

³ In sentences such as *Put the dog in the basket on the star* in which the prepositional phrase *in the basket* initially can be interpreted as the goal of the action rather than as an attribute of *the dog*.

⁴ Such as in *The coach knew (that) you missed practice* in which the second person pronoun (*you*) can initially be mistaken for the direct object of the matrix verb (*knew*) rather than the subject of the subordinate clause.

⁵ Note that this structure are only permissible in embedded clauses.

- b.⁶ All glass inte får ätas upp av barnen innan middan.
 all ice-cream not can eaten up by children.the before dinner
 ‘The children can’t eat all the ice cream before dinner.’

In subject-initial, SVO-sentences (as in Example (5)), the direct object follows all the verbs, the sentential adverbial and the verb particle. SVO sentences are by far the most common transitive sentence in Swedish, but other constituents may be placed sentence-initially. Since Swedish is a verb-second language, only one constituent must precede the finite verb in declarative main clauses (Teleman et al. 1999 (4): 688; but see Josefsson 2012 for exceptions).⁷ The word order is therefore OVS in object-initial sentences (Example (6)).

The passive can either be constructed with the suffix *-s*, as in Example (7), or periphrastically, with the verb *bli* together with a passive participle (e.g., *All glass blev inte uppäten av barnen* – ‘all the ice cream was not eaten by the children’). In s-passives, the Undergoer is realised as an (intransitive) subject, and the Actor may be expressed as the argument of prepositional *av* (‘by’) phrase, positioned sentence-finally. It is important to note that this prepositional *av* (‘by’) phrase is omitted in about 90% of all s-passives (see, e.g., Kirri 1975; Silén 1997; and Laanemets 2012). The type of passives included in the present study – s-passives with overt Actors – therefore make up only a small part of all passives in the corpus. Since the purpose of the study is to investigate distributional differences of properties between Actors and Undergoers, I chose to make this restriction for two reasons. First, in order to make such a comparison in the first place, obviously both the Actor and the Undergoer argument must be included in the target sentences. Passive sentences without overt Actors are therefore excluded. Second, the periphrastic passive has been shown to be used only in situations in which the subject – which functions as the Undergoer in passives – to some extent is in control of the described event (e.g., Engdahl 2001; 2006; Engdahl & Laanemets 2015). The subject argument in periphrastic passives therefore takes on role-semantic properties that are associated with both Actors and Undergoers, and the Actor-Undergoer dichotomy does therefore not apply in periphrastic passives. These are therefore also excluded.

In sentences with auxiliary verbs, a sentence adverbial and/or a verb particle, alternative word orders are permissible. In SVO sentences and passives, sentential adverbials follow the finite verb in main clauses (Example (5a) and (7a)) but precede it in subordinate clauses (Example (5b) and (7b)) (Teleman et al. 1999 (4): 7). In OVS sentences, the subject precedes the non-finite verb(s) and the verb particle (as in Examples (6a) and (6b)) (Teleman et al. 1999 (3): 39–40), and sentential adverbials either follow the subject NP (Example (6a)) or precede it (Example (6b)).

2.3.2 Means of identifying grammatical functions in Swedish

When no other means of identifying grammatical functions are available, grammatical functions are assigned on the basis of word order (a phenomenon referred to as *word order freezing*, see e.g., Bouma 2011; Mahowald 2011). In such cases, the first NP argument is assigned the subject function and the second the object function. Grammatical functions can also be determined on the basis of the case forms of personal pronouns (Sköld 1970). Swedish lacks noun case marking and verb agreement, but has subject and object forms for personal pronouns (Teleman et al. 1999 (2): 296). The paradigm is illustrated

⁶ Note that this structure are only permissible in embedded clauses.

⁷ In the remainder of the paper, I will use the term *clause* to differentiate between main or subordinate clauses, or to talk about specific types of subordinate clauses. I will use the term *sentence* when referring to a clause independent of whether it constitutes a sentence on its own, is embedded, or contains another embedded clause.

in Table 1. As pointed out by an anonymous reviewer, the masculine third-person subject pronoun *han* was historically used as the object form, and is therefore often used for both the subject and the object function in various dialects and in colloquial Swedish more generally.⁸ Both spoken and informal written Swedish displays widespread syncretism for the third-person plural pronoun, the form *dom* being used instead of *de* and *dem*. The generic pronoun *man* is only used in the subject function and therefore unequivocally identifies the subject.

Also the relative word order between non-finite verbs and verb particles, on the one hand, and the second NP, on the other, can disambiguate OVS sentences (Sköld 1970; Rahkonen 2006). In SVO sentences, the object follows all the verbs, the sentential adverbial and the verb particle. In OVS sentences, the subject instead precedes the non-finite verb(s) and the verb particle (Teleman et al. 1999 (3): 39–40). This is illustrated in Examples (8) and (9) (adopted from Rahkonen 2006). In (8a), the initial NP must function as the subject, because the final NP follows the main verb. In (8b), it must instead be the object, because the final NP precedes the main verb.

- (8) a. Den äldsta av rävarna har lurat jägaren.
 the oldest of foxes.the has cheated hunter.the
 ‘The oldest of the foxes has cheated the hunter.’
 b. Den äldsta av rävarna har jägaren lurat.
 the oldest of foxes.the has hunter.the cheated
 ‘The oldest of the foxes, the hunter has cheated it.’

In 9, the NP functions are instead unambiguously determined by the relative ordering between the final NP and the verb particle. The particle precedes the final NP in subject-initial sentences such as (9a), but follows it in object-initial sentences as (9b).

- (9) a. En av gästerna kastade ut värden.
 one of customers.the threw out inkeeper.the
 ‘One of the customers threw out the inkeeper.’
 b. En av gästerna kastade värden ut.
 one of customers.the threw inkeeper.the out
 ‘The inkeeper threw out one of the customers.’

Transitive sentences that contain more than one verb or a verb particle are therefore always morphosyntactically unambiguous regarding grammatical functions.

In addition to these formal markers – that is, case forms and the presence of auxiliary verbs and verb particles – grammatical functions can also in some cases be determined

Table 1: Personal pronouns in written Swedish, differentiating between nominative and accusative forms.

	First pers.		Second pers.		Third pers.			Third pers. generic
	SG.	PL.	SG.	PL.	SG. FEM.	SG. MASC.	PL.	
Subject	jag	Vi	du	ni	hon	han	de	man
Object	mig	oss	dig	er	henne	honom	dem	.

⁸ Recently, the gender neutral third-person *hen* was officially accepted in Swedish, as evident by its recent inclusion in the glossary of the Swedish Academy (SAOL 2015). However, as the data in the present corpus study consist of texts written from 1961 to 1993, no occurrences of *hen* was found.

on the basis of discourse semantics and animacy. As discussed in Section 2.1, many verbs require a volitional or sentient and therefore human or animate Actor, but do not put the same constraints on the Undergoer. Animacy can therefore sometimes function as a cue to grammatical functions. For instance, in the final sentence in Example (10) (taken from the corpus used in this study), the initial NP *den saken* ('that thing') functions as the object. But this can only be determined on the basis of the semantic relationship between the verb, the second NP, and its relationship to the previous discourse. Because the first NP is inanimate and refers back to the event of being stabbed in the head, the only semantically plausible interpretation is that it was the thick-headed Olof who survived, and therefore constitute the subject of the sentence.⁹

- (10) Det stämmer att Olof Ersson blivit huggen i huvudet med en yxa. En
 it is.true that Olof Ersson been stabbed in head.the with an axe an
 yxhammare, närmare bestämt. Den rev upp ett fem centimeter långt sår
 axhammer closer decided it tore up a five centimeter long wound
 och krossade hjässbenen. **Men den saken överlevde den tjockskallige**
 and crushed crown.bones.the **but that thing survived the thickheaded**
Olof.
Olof
 'It's correct that Olof Ersson was stabbed in the head with an ax. An axhammer,
 to be precise. It tore up a five centimeter long wound and crushed the skull. But
 that thing the thick-headed Olof survived.'

Because of this, writers might avoid using the potentially ambiguous OVS word order in cases where both arguments are animate, and their functions therefore cannot be determined on the basis of an animacy difference. As mentioned in Section 2.2, several corpus studies in other languages have found evidence that this is the case. That is, OVS sentences are used less frequently when both the subject and object is animate (Øvreliid 2004; Snider & Zaenen 2006; Bouma 2008; Bader & Häussler 2010).

2.4 Previous studies on Swedish

2.4.1 Object fronting in Swedish

Previous studies have shown that sentences with fronted direct objects are highly infrequent in Swedish. For instance, in a 87,000-word corpus consisting of newspaper texts, high school textbooks, magazines, and governmental brochures, written in between 1962 and 1971, Westman (1974) found that only about 2% of all main clauses were object-initial. Similar results are reported by Jörgensen (1976), who found about 1.6% of all main clauses in formal and scripted radio news to be object-initial. Nordman (1992) found that around 2% of all main clauses were object-initial in her 45,000-word corpus of technical and academic texts. In Bohnacker & Rosén's (2008) 17,500 word corpus of informal Swedish, written by high-school and university students, about 3% of all sentences were object-initial. In informal, spoken Swedish, object fronting is somewhat more frequent. In a corpus of discussions between Swedish academics, Jörgensen (1976) found 9% of all main clauses to be object-initial, and in a collection of informal interviews, as much as 14% of all main clauses were object-initial. These findings show that although the object-initial word order is highly infrequent in formal, written Swedish, it is fairly frequent in colloquial speech.

⁹ As pointed out by an anonymous reviewer, it is possible to interpret an inanimate NP as the subject of a sentence with the verb *survive* in particular metaphorical contexts. However, such an interpretation is impossible in the present context.

In line with the discussion presented in Section 2.1, direct object fronting in Swedish is most commonly assumed to be used when it is the object, rather than the subject, that is topical and therefore highly discourse prominent (Teleman & Wieselgren 1970; Teleman et al. 1999 (4): 432; Rahkonen 2006; Bohnacker & Rosen 2008; 2009; Bohnacker 2010). Discourse prominent NPs are given and often definite. As such, they are often anaphoric and refer back to something in the previous discourse, or deictic and refer to something in the situational context (such as speech act participants). Fronted objects that consist of inanimate third-person pronouns most often either refer back to a new referent or a proposition that was introduced in the previous sentence (*focus/rheme-topic chaining*), or to a previously introduced discourse topic (*topic-topic chaining*) (Engdahl & Lindahl 2014). Using the Stockholm-Umeå Corpus (SUC), a one-million word corpus of texts from various genres published in between 1990 to 1993, Rahkonen (2006) found a strong relationship between the discourse prominence of direct objects and object fronting. Whereas 73.7% of all definite and anaphoric objects occurred in sentence-initial position, only 23.2% of the non-anaphoric and indefinite objects were located sentence-initially.

Object-initial word order is particularly common when the object consists of a neuter pronominal or demonstrative object (i.e., *det* and *detta* – ‘that’ – henceforth referred to as *text deictic* objects) that refer back to a proposition in the immediate left context (Rahkonen 2006; Bohnacker & Rosen 2008; 2009; Bohnacker 2010). In Example (11), for instance, the sentence-initial pronominal object *det* refers back to the proposition of the previous sentence.

- (11) Samtidigt som vi andas måste vi nämligen spänna ett antal
 simultaneously as we breathe must we namely flex a number
 småmuskler i svalget så att slemhinnorna inte ska dras in och
 small.muscles in throat.the so that mucosa.the not will pulled in and
 förminska luftkanalen. **Det** kan man visa med ett enkelt experiment.
 diminish air.duct.the **that** can one show with a simple experiment
 ‘At the same time as we breathe, we have to flex a number of small muscles
 in the throat, in order for the mucosa not to get pulled in and diminish the
 air duct. That can be shown with a simple experiment.’

Rahkonen (2006) found that text deictic objects more frequently occur in the sentence-initial position than in the post-verbal position. In a 90,000 word corpus consisting of essays written in Swedish, 66.7% of all text deictic objects occurred in the sentence-initial position, and in the SUC corpus, 57.5% of all text deictic objects were sentence-initial. On the basis of these results, Rahkonen (2006: 47) concluded that “the primary reason for [object-initial word ordering] in written Swedish is to accord an anaphoric [i.e., a highly discourse prominent] non-subject early placement”.

Bohnacker & Rosén (2008) came to similar conclusions. They compared the frequency of the sentence-initial text deictic *det* in Swedish to the frequency of the corresponding sentence-initial demonstrative pronoun *dass* in German. In their Swedish corpus, *det* made up 82% of all sentence-initial objects, whereas in their German corpus, only 24% of all sentence-initial objects consisted of *dass* (see also Bohnacker & Rosén 2009). On the basis of these findings, Bohnacker & Rosén (2008) concluded that the preference to start a sentence with highly discourse prominent, given or thematic information is much stronger in Swedish than it is in German. This is particular the case when the object is highly thematic and refers back to a proposition of the previous discourse. In such cases, the object is in most cases fronted in order to “enhance textual cohesion” (Bohnacker & Rosén 2008: 521).

However, as illustrated by the following example from Rahkonen (2006), not all sentence-initial objects in Swedish are high in discourse prominence:

- (12) En härligt örtekryddad soppa käkade vi bland annat.
 a delightful herb-spiced soup ate we among other
 ‘What we ate among other things was a soup delightfully spiced with herbs.’

In (12), the sentence initial object NP *En härligt örtekryddad soppa* (‘a delightful herb-spiced soup’) is indefinite and new in the discourse. Although Rahkonen (2006: 45) argued that such sentences “cannot be considered to represent the typical information structure of OVS sentences”, clearly their existence calls into question whether the object-initial construction is used solely when it is the Undergoer that is discourse prominent rather than the Actor.

The present study investigates this by comparing the distribution of NP prominence properties of argument NPs of both subject- and object-initial sentences. More specifically, below I will show that sentence-initial objects are less frequently discourse given, definite and new than sentence initial subjects. On the basis of these findings, I will argue that object fronting in Swedish is not only used when the object is high in discourse prominence and thematic, but also either when the object is contrastive or when it introduces a new topic into the discourse. Before moving on to presenting the method and results of the study, I first discuss ambiguity avoidance in Swedish object-initial sentences.

2.4.2 Ambiguity avoidance in Swedish transitive sentences

The only corpus-based study that has investigated ambiguity avoidance with respect to grammatical functions in Swedish is that of Rahkonen (2006). He investigated whether case marked personal pronoun subjects are used more frequently in potentially ambiguous OVS sentences than they are in SVO sentences (in which the argument functions are assigned on the basis of word order), and in object relative clauses with an OSV word order (which Rahkonen assumes to be unambiguous with respect to the argument functions).¹⁰ Indeed, he found case marked subjects to occur more frequently in OVS sentences in comparison to both SVO sentences and OSV object relative clauses. More specifically, 65.1% of all subjects in OVS sentences consisted of personal pronouns, in comparison to only 33% of all subjects in SVO sentences, and 41.2% in all OSV sentences. These findings suggest that writers are prone to avoid OVS word order when grammatical functions cannot be determined on the basis of the morphological form of the subject, and consequently, that writers to some extent avoid potentially ambiguous OVS sentences.

However, Rahkonen (2006) also found that the Actor argument in passive sentences, that is, the argument of the prepositional *av* (‘by’) phrase (see Example (7)), almost never is realised as a personal pronoun. He therefore suggested that the high frequency of subjects consisting of personal pronouns in OVS sentences stems from a dispreference for using personal pronouns in prepositional *av* (‘by’) phrases, rather than reflecting a tendency to avoid potential ambiguities. When writers want to signal that it is the Undergoer

¹⁰ I.e., relative clauses such as

- (i) boken som pojken läste
 book.the that boy read
 ‘the book that the boy read’

in which the relative pronoun *som* cannot function as subject and the NP *pojken* must function as subject (Rahkonen 2006).

of a transitive event that is topical, rather than the Actor, they can do so either by using the object-initial construction or the passive (e.g., Foley 2011). According to Rahkonen, writers tend to resort to the former construction whenever the Actor of the event is highly discourse prominent and therefore needs to be realised as a personal pronoun. This is to avoid the use of a prepositional *av* ('by') phrase with a personal pronoun. It is, however, rather unclear *why* such prepositional phrases would be avoided in the first place. Since the passive construction is unambiguous with respect to grammatical functions, it instead seems more plausible to assume that writers tend to resort to the passive whenever the Actor argument cannot be realised as a personal pronoun, in order to avoid a potential ambiguity. In the present study, I provide evidence for this alternative hypothesis by showing that not only case marking, but also other formal means of disambiguation, namely auxiliary verbs and verb particles (see Examples (8) and (9)), occur more frequently in OVS sentences than in both SVO sentences and passives.

Rahkonen (2006) also investigated whether case marked personal pronouns are used more frequently in what he referred to as “semantically ambiguous” OVS sentences in comparison to “semantically unambiguous” OVS sentences. In semantically ambiguous sentences, both NP arguments were animate and could therefore fill both participant roles of the verb. In semantically unambiguous sentences, the object argument instead was text deictic (i.e., consisted of *det* and *detta* – ‘that’), and therefore could only fill the Undergoer role of most verbs. Rahkonen (2006) found case marked subjects to be somewhat more frequent in the former sentences in comparison to the latter. In sentences with two animate arguments, 90% of all subjects consisted of a case marked pronoun, and in sentences with text deictic objects, 76.8% of the subjects were case marked. These findings provide some additional support for the hypothesis that writers actively avoid OVS sentences in the face of potential ambiguity (although Rahkonen (2006) himself argued against this account, see Section 5.2 below), although the difference in percentage is rather small. In this study, I provide more substantive support for the ambiguity avoidance hypothesis, by showing that the full range of formal markers of grammatical functions presented in Section 2.3.2 are used more frequently in sentences with two animate arguments than in sentences with at least one inanimate argument.

3 Method

3.1 The corpus

The sentence materials used in the study were collected from the treebank *Svensk trädbank*, a syntactically annotated version of the *Stockholm-Umeå Corpus* (SUC) (Gustafson-Capková & Hartmann 2006) with materials from the *Talbanken* (TB) corpus (Einarsson 1976a; b). The SUC corpus, which also was used by Rahkonen (2006), consists of 500 published texts from various genres that were published in between 1990 to 1993. The TB corpus consist of 85 *professional prose* texts from four different genres, written in between 1960 to 1971, originally compiled for the aforementioned study of Westman (1974). Most of the texts in these two corpora have undergone editing and proofreading. A list of the main genres of the corpora as well as the number of texts and words of each genre is shown in Table 2. The treebank is morphologically and syntactically annotated in Tiger-XML format.

Searches were conducted with TIGER search 2.1. (König et al. 2003), using the TIGER search query language (König & Lezius 2003). Data collection was done with the aim to minimise the constraints on the structural variation of the sentences. Following Jaeger (2011), search patterns were constructed to avoid false exclusions rather than false inclusions. False hits were excluded from the initial data at a later stage. The target sentences

Table 2: SUC and TB corpora main genres and their respective text, sentence and word frequencies. Number and percentage of search hits for each sentence type in each genre is also shown.

Corpus	Genre	Texts	Sentences	Hits						Total N
				SVO		OVS		Passive		
				N	%	N	%	N	%	
SUC	Press: reportage	44	7278	1149	86.3	68	5.1	47	3.5	1264
	Press: Editorial	17	2385	385	81.7	32	6.8	22	4.7	439
	Press: Reviews	27	3961	536	79.8	51	7.6	34	5.1	621
	Skills, Trades and Hobbies	58	8933	1343	82.2	118	7.2	55	3.4	1516
	Popular Lore	48	6525	1160	85.4	55	4.0	72	5.3	1287
	Biographies and Memoirs	26	3598	627	83.8	51	6.8	35	4.7	713
	Miscellaneous	70	10847	1239	78.5	50	3.2	145	9.2	1434
	Learned and Scientific Writing	83	9633	1398	83.1	63	3.7	159	9.4	1620
	General fiction	82	13028	2527	86.2	185	6.3	35	1.2	2747
	Mysteries and Science fiction	19	4070	665	84.2	58	7.3	9	1.1	732
	Light reading	20	2908	611	84.6	53	7.3	5	0.7	669
	Humor	6	1071	183	76.6	27	11.3	2	0.8	212
TB	Brochure texts	25	1733	298	83.2	20	5.6	20	5.6	338
	Newspaper texts	28	1669	277	85.8	16	5.0	14	4.3	307
	Educational texts	14	1624	292	88.0	10	3.0	20	6.0	322
	Debate articles	18	1134	259	89.0	12	4.1	8	2.7	279
Total		585	80397	12949	83.7	869	5.6	682	4.4	14500

were SVO sentences (Example (5)), OVS sentences (Example (6)) and passives (Example (7)), including both main and subordinate clauses. Search strings allowed for NPs of any length and sentences with up to three auxiliary verbs in addition to the main verb (Teleman et al. 1999 (3): 278), one verb particle, and one sentential adverbial phrase.

Data exclusion was guided by finding sentences that 1) are declarative or passive, 2) contain argument NPs that refer to the participants involved in the event denoted by the sentence predicate, 3) contain argument NPs that serve as arguments of that predicate only,¹¹ and 4) do not contain any idiomatic or lexicalised constructions.

A more detailed description of the search queries and the exclusion procedure, together with a list of excluded sentence types can be found in Appendix 1. The actual search strings are provided in Appendix 2.

A total of 14,500 sentences remained after data exclusion, out of which 12,949 sentences were SVO sentences, 869 were OVS sentences, and 682 were passives. In other words, only about 5.6% of all transitive sentences are object-initial. The rather low number of passives is due to the constraint that each passive must contain a prepositional *av* ('by') phrase.

Table 2 also shows the percentage of each sentence type of each genre. It is worth noting that there is quite a bit of variation in the proportion of sentence types across genres. However, genre differences is beyond the scope of the present paper and will not be discussed further. A more thorough investigation and discussion about genre differences in the frequency of use of subject-versus object-initial sentences can be found in Hörberg (2016: 105–107).

¹¹ As opposed to, for instance, secondary predication or causative constructions. See Appendix 1.

3.2 Properties under investigation

The NP properties under investigation in this study are givenness, animacy, definiteness, pronominality, and case marking (i.e., whether or not the NP at hand consists of a personal pronoun). The annotation of these properties were, to the extent that it was possible, based upon the morphosyntactic annotation available in Svensk Trädbank. In the following, I give brief descriptions and definitions of these properties as operationalised in the present study. A more detailed description of the annotation procedure of these properties can be found in Hörberg (2016: 79–88).

Givenness. As discussed in Section 2.1, the form of an NP argument is commonly assumed to correspond to a specific level of givenness or accessibility (e.g., Gundel et al. 1993). NPs were therefore annotated for givenness on the basis of their form. I used the four-level givenness scale shown in Table 3. The scale aims to capture the givenness/accessibility distinctions suggested by Prince (1981), Ariel (1990), and Gundel et al. (1993). It differentiates between *new/type identifiable* NPs, *token identifiable* NPs, *familiar* NPs and *given* NPs, as defined in Gundel et al. (1993). The categories largely correspond to the categories of Prince (1981), and to the NP forms of the Accessibility Hierarchy of Ariel (1990).

Conceptually, *new/type identifiable* NPs (e.g., indefinite NPs and a few grammatically definite NPs with generic reference) designate referents of a new type that are introduced in the discourse under the assumption that the referent in question is unknown to the addressee. The addressee needs to “create” and determine the type of referent without being able to determine which specific referent it pertains to.¹² *Token identifiable* NPs (e.g., full names with modifiers and definite descriptions) designate a specific referent which is assumed to be unknown to the addressee. The addressee needs to be able to either “construct” a unique referent with no previous knowledge of it, or to identify the referent on the basis of the expression, either through inference or by the information provided in the expression. *Familiar* NPs (e.g., names, kinship terms and short definite NPs) designate a

Table 3: Givenness categories and their correspondence to the givenness distinctions proposed by Prince (1981) and Gundel et al. (1993), and the Accessibility Hierarchy of Ariel (1990).

Category		Prince (1981)	Gundel et al. (1993)	Ariel (1990)
Discourse New	New/type identifiable	Brand new	Type identifiable	–
	Token identifiable	(indefinite) Inferrables	Referential	full name + modifier; long definite description (introduces unique referent)
(definite) Inferrables		Uniquely identifiable	long definite description (identifies unique referent)	
Discourse Given	Familiar	(demonstrative) Inferrables	Familiar	full name; short definite description; last name; first name; distal demonstrative + modifier; proximal demonstrative + modifier; distal demonstrative + NP
		Unused		
		Evoked (lexical NP:s)		
	Given	Evoked (pronominal NP:s)	Activated	proximal demonstrative + NP; naked distal demonstrative; naked proximal demonstrative
			In focus	personal pronomina

¹² The category does not have any correspondence in the Accessibility Hierarchy which only concerns degree of accessibility/givenness and therefore only cover argument NPs with definite reference.

specific referent which is assumed to be known by the addressee. *Given* NPs (e.g., personal pronouns or short NPs with a proximal demonstrative) finally, designate a specific referent which is active in the discourse. The speaker therefore assumes that it is represented in the short term memory of the addressee, possibly at the center of attention.

In most of the analyses below, I use the dichotomous distinction between *discourse new* NPs, consisting of New/type identifiable NPs and token identifiable NPs, and *discourse given* NPs, consisting of Familiar and Given NPs. I will also differentiate between arguments as being *high* or *low* in discourse prominence, using discourse prominence as an umbrella term for the extent to which arguments possess properties associated with discourse givenness (e.g., such as being definite, pronominal and first- or second-person).

Animacy. A dichotomous distinction between animate and inanimate was used. NPs referring to humans and non-human animate beings, groups of animate beings, as well as organizations and politically organised geographic regions that depend on human organization in some sense (e.g., rock bands, political parties, municipals, countries etc) were annotated as animate. All other NPs were annotated as inanimate.

Definiteness. Definite and indefinite NPs were classified on the basis of the grammatical or semantic definiteness of either the head or the prephrasal element (e.g., the article, determiner or initial pronoun, Teleman et al. 1999 (3): 15), as annotated in the corpus. Indefinite NPs also included NPs without a grammatically definite/indefinite head or prephrasal element (i.e., indefinites with weak reference, see Teleman et al. 1999 (4): 175–176).

Pronominality. A two-way distinction between pronominal NPs and lexical NPs, consisting of proper and common noun NPs, was done on the basis of the word class of the NP head.

Case/person. As discussed in Section 2.3.2, the case marking system of Swedish is limited to personal and generic pronouns (see Table 1). As such, the case distinction coincides with a distinction between personal pronouns and other NPs. Swedish case distinguishes subjects from direct objects and obliques. All NPs that consisted of or was headed by the pronouns shown in Table 1 were classified as case marked/person – *person* here referring to personal pronouns generally. All others were classified as unmarked/non-person, *non-person* referring to all NPs not consisting of or being headed by a personal pronoun.

Text deixis. As discussed in Section 2.4.1, object-initial word order is especially common in Swedish when the object is text deictic, consisting of a third-person neuter pronoun or a neuter demonstrative (i.e., *det* and *detta*, Rahkonen 2006; Bohnacker & Rosén 2008), and therefore often refers to an earlier or an upcoming proposition of the discourse (Levinson 2004). The present study also investigates the distribution of text deictic NPs. NPs in which the head of the NP either consisted of a third-person singular neuter pronoun (i.e., *det*), a singular neuter demonstrative pronoun (i.e., *detta*, *det här* or *det där*) or the neuter relational pronoun *detsamma* were categorised as text deictic (see Teleman et al. 1999 (2): 5). All other NPs were classified as non-text deictic.

3.3 Statistical analyses

I conducted all statistical analyses that are presented below with the statistical software R (R Core Team 2016).¹³ The significance of distributional differences of some property between two or more categories (e.g., the difference in the proportion of Discourse Given NPs between subjects and objects in SVO sentences, see Table 4), were tested on the basis of χ^2 -tests, testing the null hypothesis that the property at hand has the same distribu-

¹³ All analyses were conducted across both main and subordinate clauses. However, in order to determine whether the inclusion of subordinate clauses bias the results, I conducted the same analyses across main clauses only. These yielded the same results as those reported below, showing that the inclusion of subordinate clauses is unproblematic.

Table 4: The distribution of animacy and referential properties for subjects and objects in subject-initial sentences (total N: 12949).

Property	Subject		Object		p
	N	%	N	%	
Discourse Given	9167	70.8%	4054	31.3%	<.0001
Definite	10428	80.5%	6426	49.6%	<.0001
Pronominal	6744	52.1%	2272	17.5%	<.0001
Animate	9637	74.4%	2526	19.5%	<.0001
Case marked/Person	5796	44.8%	1436	11.1%	<.0001

tion across the categories (e.g., the proportion of Discourse Given NPs being the same for subjects and objects in SVO sentences). Multiple χ^2 -tests were therefore conducted on the same data set. Since the risk for a false positive (i.e., falsely rejecting a null hypothesis in favour of the research hypothesis) goes up with the number of tests that are conducted, all p-values from these χ^2 -tests were corrected for multiple comparisons on the basis of the method suggested by Holm (1979).

Logistic mixed effects modelling was done with the `glmer()` function in the *lme4* package (Bates et al. 2015). All of the analyses, apart from the logistic mixed effects model analysis, are conducted across all genres of both corpora without taking any potential differences between genres into account.

4 Results

4.1 The distribution of animacy and referential properties in SVO-sentences

As discussed in Section 2.1, several corpus studies have found that subjects are more frequently animate and high in discourse prominence than objects. This asymmetry stems from the functional difference between subjects and objects. Whereas subjects express the Actor role and in the prototypical case are topical, objects refer to Undergoers that commonly are focused and therefore new in the discourse. Differences in the distribution of animacy and referential properties between subjects and objects in prototypical SVO sentences was also investigated in the present study. The results are shown in Table 4.¹⁴ The table shows that subjects of SVO-sentences are animate (rather than inanimate) to a much greater extent than objects. Subjects are also much more frequently high in discourse prominence in terms of givenness, definiteness, pronominality and person than objects. These findings are rather uncontroversial and confirm the findings of previous studies (Dahl & Fraurud 1996; Dahl 2000; Du Bois 2003; Kempen & Harbusch 2004; Øvrelid 2004; Bouma 2008).

4.2 Object fronting and discourse topicalisation

As discussed in Section 2.4.1, object fronting in Swedish is commonly assumed to be used when the object is topical and therefore high in discourse prominence (Teleman & Wieselgren 1970; Teleman et al. 1999 (4): 432; Rahkonen 2006; Bohnacker & Rosen 2008; 2009; Bohnacker 2010). In line with this, Rahkonen (2006) found sentence-initial objects to more frequently be high in discourse prominence than post-verbal objects. Both

¹⁴ Note, that in this and the following tables, the n:s of each row do not sum to the total number of sentences, nor do the percentages sum to 100%. This is because the properties are not mutually exclusive (although they might be dependent). Each value refers to the number or the percentage of occurrences that possess the property at hand (e.g., are definite) in comparison to the occurrences that do not possess that property (e.g., are indefinite).

Rahkonen (2006) and Bohnacker & Rosen (2008) further found object fronting of text deictic objects to be particularly frequent. The findings of the present study in part confirm these results. The distribution of referential properties of both sentence-initial and post-verbal object NPs is shown in Table 5. Objects are more frequently high in discourse prominence when positioned in the sentence-initial position in comparison to when they occur in post-verbal position in terms of discourse givenness, definiteness and pronominality. Sentence-initial objects are also much more frequently text deictic in comparison to objects positioned in their canonical, post-verbal position.

These findings indicate that objects more often are positioned sentence-initially when they are highly discourse prominent, and therefore refer back to either a referent that has been previously introduced into the discourse, or a previously mentioned proposition. In particular, text deictic objects are almost exclusively positioned sentence-initially and very seldom occur in post-verbal position.

4.3 Other functions of object fronting

However, object fronting does not only take place when the object is high in discourse prominence. Sentence-initial objects are less frequently discourse prominent than sentence-initial subjects. This is illustrated in Table 6, which compares the distribution of referential properties of initial objects to that of initial subjects. Initial objects are less frequently discourse given, definite and new than initial subjects. If the only motivation for object fronting was to emphasise that it is the object, rather than the subject, that refers to the most discourse prominent referent of the two NPs, one would expect initial objects and subjects to be discourse prominent to an equal extent. In fact, as will be discussed below in Section 4.4, subjects are more frequently high in discourse prominence when positioned in the post-verbal position than when they are located in their canonical, sentence-initial position (see Table 8). Consequently, it is not the relative prominence difference between the NPs (independent of their functions) that determine their positioning either, such that it always is the most discourse prominent NP that is positioned sentence-initially. This

Table 5: The distribution of referential properties for sentence-initial (total N: 869) and post-verbal (total N: 12949) objects, respectively.

Property	Initial object		Post-verbal object		<i>p</i>
	N	%	N	%	
Discourse Given	548	63.1%	4054	31.3%	<.0001
Definite	656	75.5%	6426	49.6%	<.0001
Pronominal	402	46.3%	2272	17.5%	<.0001
Text Deictic	340	39.1%	332	2.6%	<.0001

Table 6: The distribution of referential properties for sentence-initial objects (total N: 869) and subjects (total N: 12949).

Property	Initial object		Initial subject		<i>p</i>
	N	%	N	%	
Discourse Given	548	63.1%	9167	70.8%	<.0001
Definite	656	75.5%	10428	80.5%	<.01
Pronominal	402	46.3%	6744	52.1%	<.01
Text Deictic	340	39.1%	298	2.3%	<.0001

Table 7: The percentage of SVO- (total N: 12949) and OVS-sentences (total N: 869) in which the initial NP either outranks, is equally ranked or is outranked by the second NP in terms of the prominence feature at hand.

Word Order	Property	NP1 > NP2	NP1 = NP2	NP1 < NP2	p
SVO	Givenness	59.5%	27.2%	13.3%	<.001
	Definiteness	41.2%	50.2%	8.6%	<.001
	Pronominality	49.0%	45.3%	5.7%	<.001
	Animacy	58.2%	38.6%	3.3%	<.001
	Case marking/Person	38.1%	57.4%	4.5%	<.001
OVS	Givenness	24.2%	36.0%	39.8%	<.001
	Definiteness	13.1%	66.4%	20.5%	<.001
	Pronominality	12.1%	45.2%	42.7%	<.001
	Animacy	0.3%	16.6%	83.1%	<.001
	Case marking/Person	0.1%	33.5%	66.4%	<.001

is illustrated in Table 7. The table shows the percentage of SVO- and OVS- sentences, respectively, in which the initial NP either outranks, is equally ranked or is outranked by the second NP in terms of the prominence feature at hand. In SVO-sentences, the sentence-initial subject (NP1) most frequently outranks or is equally ranked to the post-verbal object (NP2) in terms of all prominence properties (givenness, definiteness and pronominality). In OVS-sentences, on the other hand, the post-verbal subject (NP2) more frequently outranks or is equally ranked to the sentence-initial object (NP1) in terms of discourse prominence. These findings show that object fronting is not only used when the object is high in discourse prominence.

4.3.1 Contrast

Indeed, 37% of all fronted objects in the corpus data are discourse new. In many of these cases, object fronting is used to indicate that the object is contrastive (cf. Teleman et al. 1999 (4): 432; Molnár 2002; 2006; Molnár & Winkler 2010). As discussed in Section 2.1, contrastive NPs are contrasted against a contextually determined limited set of alternatives (e.g., Molnár 2002; 2006; Gundel & Fretheim 2004; Erteschik-Shir 2007; Molnár & Winkler 2011). Such an instance is exemplified in (13):

- (13) Dragon är gott i en omelett, basilika likaså – särskilt om omeletten
 tarragon is good in an omelette basil as.well particularly if omelette.the
 serveras med tomatsallad. **Oregano kan man också tänka sig.**
 serve-PASS with tomato.sallad **oregano can one also consider 3SG.REFL**
 ‘Tarragon is good in an omelette, basil as well – in particular if the omelette is
 served with tomato salad. Oregano can also be used.’

The fronted object *oregano* (‘oregano’) of the second sentence is contrasted against the set of herbs that also work well in an omelette. In Example (13), the sentence-initial object is therefore discourse new and focused, but positioned sentence-initially in order to emphasise that it is contrastive.

4.3.2 Topic shift

In some cases, object fronting also seems to be used to introduce a new topic into the discourse, and thereby to mark a discourse *topic shift* (James 1995). In such constructions, the object NP is discourse new but occupies the sentence-initial position in order to emphasize

that it contrasts with – but is contextually related to – the previous discourse topic. Following Molnár and Winkler (2010), I assume that such constructions serve to establish discourse coherence in that they signal relatedness to the previous discourse topic. Since the object NP is new and “informationally heavy”, the rest of the construction tends to be “informationally light”, in order to keep the information flow of the sentence as constant as possible (cf. Fenk-Oczlon 2001). As such, it contains a semantically weak main verb such as a possessive or copula verb, and a post-verbal subject that generally is topical, given and highly discourse prominent, and that serves as a reference point or “ground” for the introduction of the new discourse topic NP. This is exemplified in Example (14), taken from the corpus.

- (14) I ateljén står bland färger ett gammalt skrivbord fullt av vykort och
 in studio.the stand among paint an old desk full of post.cards and
 andra bilder av kvinnor i konsten, från Marilyn till Botticelli och Cranach.
 other pictures of women in art.the from Marilyn to Botticelli and Cranach
 En bok om Picasso är uppslagen och där finns en bild av Les
 a book about Picasso is opened and there is a picture of Les
 Demosielles d’Avignon. **Några kvinnliga konstnärssidoler har Cecilia inte,**
 Demosielles d’Avignon **any female favorite-artists have Cecilia not**
 tyvärr, säger hon. Stimulansen kommer från manliga konstnärer som
 unfortunately says she spur.the comes from male artists like
 Rousseau, Picasso, Cranach. Kvinnliga konstnärer som Derkert, Hjertén och
 Rousseau Picasso Cranach female artists like Derkert Hjertén and
 Delauney är mer moraliskt stöd.
 Delauney is more moral support
 ‘In the studio, among paint, stands an old desk cluttered with post cards and other
 pictures of women in the art, from Marilyn to Botticelli and Cranach. A book
 about Picasso is opened and in it, there is a picture of *Les Demosielles d’Avignon*.
 Unfortunately, Cecilia does not have any *female favorite artists*, she says. The spur
 comes from male artists such as Rousseau, Picasso, Cranach. Female artists such
 as Derkert, Hjertén and Delauney is more of a moral support.’

Example (14) is a short passage from a newspaper article about two female painters, Cecilia and Sissel, in the beginning of their career. The beginning of the passage is about how women in general serve as a source of inspiration for Cecilia’s artwork. In the critical sentence (which is boldfaced), the discourse new object NP *några kvinnliga konstnärssidoler* (‘any female favourite artists’) is positioned in the sentence-initial position in order to highlight its contrast with the contextually related discourse topic about inspirational women more generally. In the subsequent sentences, the discourse takes a turn towards a discussion of inspirational artists. The use of object fronting in this context therefore appears to signal a shift in discourse topic by means of contrasting the new discourse topic with the old. This is done by using the possessive verb *ha*, which in this context serves to establish a relationship between the already known and highly given referent Cecilia and the new discourse topic about inspirational artists.

In order to quantitatively investigate the hypothesis that sentences with a possessive verb can be used to introduce a new discourse topic by means of object fronting, I investigated whether fronted objects more frequently co-occur with possessive verbs when they are discourse new, in comparison to when they are discourse given.¹⁵ Indeed, in 46.5% (47 out of 86) of all object-initial sentences with a discourse new object, the verb is possessive. This

¹⁵ A detailed description of the annotation of verb classes can be found in Hörberg (2016: 84–86).

is significantly more than in object-initial sentences with a discourse given object, in which only 22.6% (274 out of 783) of all verbs are possessive, $\chi^2(1) = 22.38, p < .0001$. In other words, discourse new sentence-initial objects much more frequently occur with possessive verbs than discourse given sentence-initial objects. This provides support for the hypothesis that a possessive construction can be used to introduce new discourse topics.

4.4 Differences between sentence-initial and post-verbal subjects

In order to shed further light on functional differences between object- and subject-initial sentences, I also investigated differences in the distribution of NP properties between post-verbal and sentence-initial subjects. Table 8 shows the results. Subjects of transitive sentences are more frequently high in discourse prominence in terms of pronominality and person when positioned in the post-verbal position. Transitive subjects are also more frequently animate when positioned in the post-verbal position. An important contribution to these distributional differences is that post-verbal subjects more commonly occur as personal pronouns (such as in, e.g., Example (12)). Subjects therefore more often refer to highly discourse prominent first-, second- or third-person participants when occurring in the post-verbal position. This finding is consistent with the suggestion that, at least in object-initial sentences with discourse new objects, the remainder of the sentence tend to be “informationally light” and predictable. As a result, post-verbal subjects can be expected to be encoded with personal pronouns, which commonly refer to discourse participants that are highly given and therefore predictable in the discourse.

However, it is not the case that the predictability of the post-verbal subject (in terms of being realised as a personal pronoun) depends on the givenness status of the sentence-initial object. Table 9 shows the proportion between marked/person and unmarked/non-person post-verbal subjects, differentiated on the basis of the givenness of the sentence-initial

Table 8: The distribution of referential and animacy properties for post-verbal (total N: 869) and sentence-initial subjects (total N: 12949), respectively.

Property	Post-verbal subject		Sentence-initial subject		p
	N	%	N	%	
Discourse Given	641	73.8%	9167	70.8%	.19
Definite	709	81.6%	10428	80.5%	.446
Pronominal	622	71.6%	6744	52.1%	<.0001
Animate	782	90.0%	9637	74.4%	<.0001
Case marking/Person	587	67.5%	5796	44.8%	<.0001

Table 9: The percentage of marked/person versus unmarked/non-person post-verbal subjects (i.e., only data from OVS sentences, total N: 869), differentiated on the basis of the givenness of the sentence-initial object NP.

		NP1 Givenness							
		Given		Familiar		Token Id.		New	
		N	%	N	%	N	%	N	%
NP2 Case/Person	Marked/Person	276	69.5%	109	72.2%	64	61.5%	138	63.6%
	Unmarked/Non-person	121	30.5%	42	27.8%	40	38.5%	79	36.4%
TOTAL		397	100%	151	100%	104	100%	217	100%

object NP. Thus, Table 9 only includes data from OVS sentences. The table illustrates that the distribution of a personal pronoun and non-personal pronoun post-verbal subject NPs is roughly the same over the givenness levels of the sentence-initial object. This observation was confirmed by a χ^2 -test of independence which found no support for a relationship between the NP2 case marking/person and NP1 givenness in object fronted sentences, $\chi^2(3) = 5.45, p = 0.28$. The strong tendency to use a personal pronoun – and therefore highly predictable – subject NP in object fronted sentences is therefore independent of the givenness of the sentence-initial object. This suggests that it is not only when the fronted object is discourse new that writers aim to keep the remainder of the sentence predictable and therefore easier for the reader. This rather seems to be a more general tendency of all object-initial sentences.

4.5 Avoidance of potentially ambiguous OVS sentences

I now turn to the issue of whether writers are inclined to avoid potentially ambiguous object-initial sentences, first with respect to whether other formal, morphosyntactic information about the argument functions is available, and then with regard to whether the grammatical functions can be determined on the basis of an argument animacy difference. This is done by investigating the distribution of formal, morphosyntactic, means of disambiguation, on the one hand, and the distribution of animacy-based means of disambiguation, on the other, in terms of how they differ between object-initial sentences, subject-initial sentences and passives.

4.5.1 Formal means of disambiguation

As discussed in Section 2.4.2, Rahkonen (2006) found case marked personal pronoun subjects to be used more frequently in potentially ambiguous OVS sentences than in SVO sentences and in object relative clauses with an OSV word order. However, Rahkonen (2006) did not interpret these findings as a tendency for writers to avoid OVS sentences without a case marked subject. Instead, he suggested that the high frequency of personal pronoun subjects in OVS sentences stems from a dispreference for using personal pronouns to overtly express the Actor argument in passive sentences, that is, as the argument of a prepositional *av* ('by') phrase. It is, however, unclear why such prepositional phrases would be avoided in the first place. A more plausible hypothesis is that writers tend to resort to the passive – which is unambiguous with regard to grammatical functions – when there are no other formal means of identifying the argument functions available.

In order to test this alternative hypothesis, I investigated whether it only is case marking that occurs more frequently in OVS sentences than in SVO sentences and passives, or if this also applies to auxiliary verbs and verb particles (see Examples (8) and (9)). The results are shown in Table 10. The table shows the percentage of OVS, SVO and passive

Table 10: The percentage of OVS (total N: 869), SVO (total N: 12949) and passive (total N: 682) sentences with a case marked NP, an auxiliary verb, a verb particle, or any or several of these formal markers.

Property	OVS		SVO		Passive		p
	N	%	N	%	N	%	
Case marking	588	67.7%	6375	49.2%	59	8.7%	<.0001
Auxiliary verb	323	37.2%	3993	30.8%	167	24.5%	<.0001
Verb particle	121	13.9%	1335	10.3%	22	3.2%	<.0001
Any formal marker	713	82.0%	8429	65.1%	221	32.4%	<.0001

sentences, respectively, that contain at least one case marked NP, an auxiliary verb, a verb particle, or any or several of these formal markers (i.e., that are unambiguous by any means). These results show that the preference for the use of formal markers in OVS sentences and the dispreference for using them in passive sentences is not limited to case marking of the Actor argument. They therefore speak against Rahkonen's (2006) suggestion that the high frequency of personal pronoun subjects in OVS sentences is driven by a dispreference for using personal pronouns in prepositional *av* ('by') phrases.

Indeed, all three types of formal markers occur most frequently in OVS sentences, followed by SVO sentences, and least frequent in passives. These findings speak in favor of the idea that writers tend to use the passive rather than an object-initial sentence when the target sentence won't contain any formal markers. In other words, the choice between using an OVS sentence and a passive at least in part seems to be driven by a motivation to avoid potentially ambiguous OVS sentences.

4.5.2 Animacy-based disambiguation

As discussed in Section 2.2, several studies have found the animacy difference between subjects and objects to be more pronounced in object-initial sentences than in subject-initial sentences (Øvrelid 2004; Snider & Zaenen 2006; Bouma 2008; Bader & Häussler 2010). Speakers and writers less frequently use the potentially ambiguous OVS word order when the grammatical functions of the NPs cannot be determined on the basis of an animacy difference between the NPs.

The results of the present study show similar results. As illustrated in Figure 1, subjects are more frequently animate and objects more frequently inanimate in OVS sentences than in SVO sentences. In OVS sentences, the NP arguments are equal in animacy (i.e., both either animate or inanimate) in only 16.6% of all sentences, whereas in SVO sentences, they are equal in animacy in 38.6% of all sentences (see Table 7).

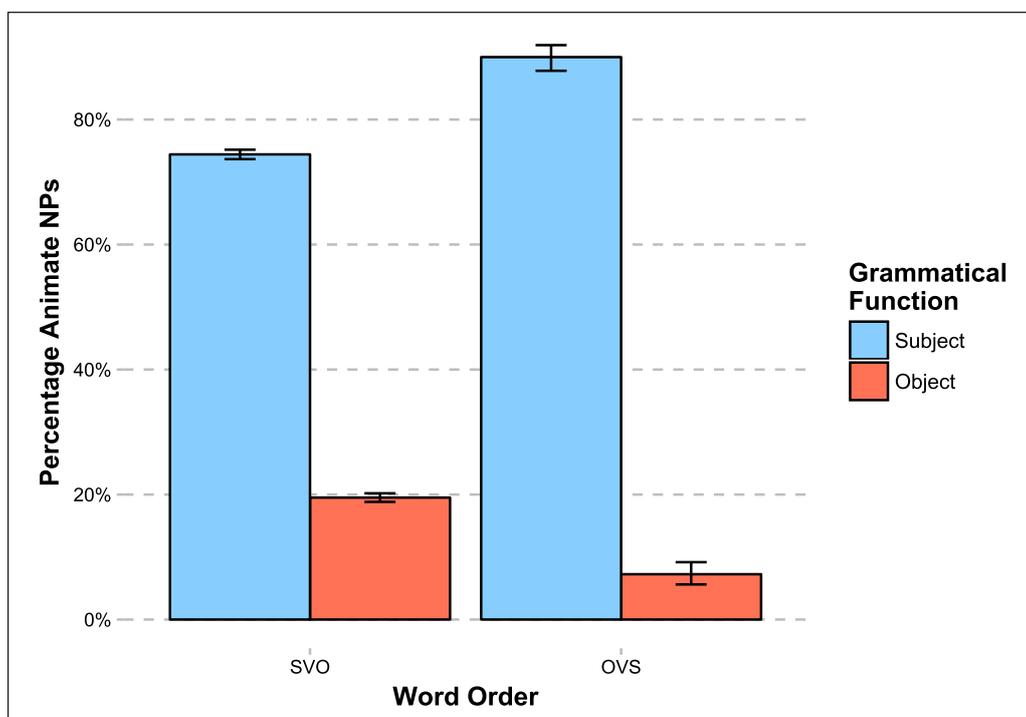


Figure 1: Percentage of animate NP arguments, comparing subjects and objects in SVO and OVS sentences, respectively. Error bars illustrate 95% confidence intervals, calculated on the basis of normal approximation.

These findings show that the animacy difference between subjects and objects is more pronounced in object-initial sentences than in subject-initial sentences also in written Swedish. They also indicate that writers are particularly prone to use an inanimate object (i.e., Undergoer) NP in object-initial sentences. As shown in Table 11, this is not the case in passives. Here, the sentence-initial Undergoer NP is animate equally often as it is in SVO sentences. This indicates that writers preferably use the object-initial construction in cases where the object argument is inanimate, so that the sentence can be assumed to be object-initial already at the presentation of the initial NP, or at least at the position of the verb. This in turn suggests that writers actively avoid OVS sentences in which the grammatical functions cannot be determined on the basis of the animacy of the initial NP (such as in (1a)).

Table 11 also shows that the sentence-initial Undergoer of passive sentences is high in discourse prominence in terms of givenness and definiteness in comparison to Undergoers in SVO sentences. This confirms the idea that the passive is, as the object-initial construction, used when the information structure is marked in terms of the Undergoer being discourse topical rather than the Actor (see Foley 2011 on the discussion of foregrounding passives).

However, it is important to note that the distribution of properties over the Actor argument is significantly different in passives in comparison to both SVO and OVS sentences. In passives, the Actor is much more often low in discourse prominence and inanimate. This is likely to reflect that the passive very often is used to background (or in most cases to completely omit) the Actor argument (Haspelmath 1990; Foley 2011). It is likely that this is more frequently done with discourse new Actor arguments which often are lexical and indefinite.

4.5.3 The interplay between formal and animacy-based means of disambiguation

As discussed in Section 2.4.2, Rahkonen (2006) also found case marked personal pronoun subjects to occur somewhat more frequently in OVS sentences with two animate NP arguments, in comparison to OVS sentences with a text deictic object argument (unable to fill the Undergoer role of many verbs). Although the difference was small, these findings pro-

Table 11: The distribution of animacy and referential properties for Undergoers and Actors in SVO (total N: 12949), OVS (total N: 869) and passive (total N: 682) sentences.

NP	Property	SVO		Passive		OVS		p
		N	%	N	%	N	%	
Undergoer	Discourse Given	4054	31.3%	340	49.9%	548	63.1%	<.0001
	Definite	6426	49.6%	543	79.6%	656	75.5%	<.0001
	Pronominal	2272	17.5%	117	17.2%	402	46.3%	<.0001
	Text Deictic	332	2.60%	29	4.3%	340	39.1%	<.0001
	Animate	2526	19.5%	125	18.3%	63	7.2%	<.0001
	Case/Person	1436	11.1%	54	7.9%	11	1.3%	<.0001
Actor	Discourse Given	9167	70.8%	211	30.9%	641	73.8%	<.0001
	Definite	10428	80.5%	368	54%	709	81.6%	<.0001
	Pronominal	6744	52.1%	16	2.3%	622	71.6%	<.0001
	Text Deictic	298	2.3%	2	0.3%	3	0.3%	<.0001
	Animate	9637	74.4%	305	44.7%	782	90%	<.0001
	Case/Person	5796	44.8%	5	0.7%	587	67.5%	<.0001

vide some additional support for the hypothesis that writers avoid OVS sentences when the argument functions might not be able to be determined on the basis of animacy.

In the present study, I investigate distributional differences of all formal markers in sentences with two animate arguments versus sentences with one or two inanimate arguments, comparing OVS sentences to both SVO sentences and passives. Sentences in which both the subject and the object arguments are animate were classified as *animate argument sentences*, and all other sentences as *inanimate argument sentences*. In the former sentence type, the argument functions cannot be determined on the basis of animacy, since both arguments potentially can fulfil the Actor role.

The results are illustrated in Figure 2. The figure shows the percentage of formal marking in OVS, SVO and passive sentences, comparing inanimate argument sentences with animate argument sentences. The figure indicates that formal marking is more frequent in OVS sentences than in SVO sentences, but less frequent in passives than in SVO sentences. It also indicates that formal marking is more frequent in animate argument sentences than it is in inanimate argument sentences, both for OVS, SVO and passive sentences.

In order to further investigate these distributional differences, I analysed the data using logistic mixed effects regression modelling (Gelman & Hill 2006; Jaeger 2008). The logistic mixed effects model is a type of general linear model (Howell 2010) that is used for a binomially distributed dependent variable, and that accounts for random effects, such as differences between genres in a corpus. The model predicts the outcome of a dichotomous dependent variable in terms of log odds (i.e., *logits*) as a linear combination of a set of independent variables (i.e., fixed effects) and one or several random variables (i.e., random effects, see Gelman & Hill 2006; Jaeger 2008). In the present model, I predicted the proportion of formal marking in terms of log odds with Sentence Type (SVO vs. OVS vs. passive), Animacy (animate arguments vs. inanimate argument(s)) and the Sentence

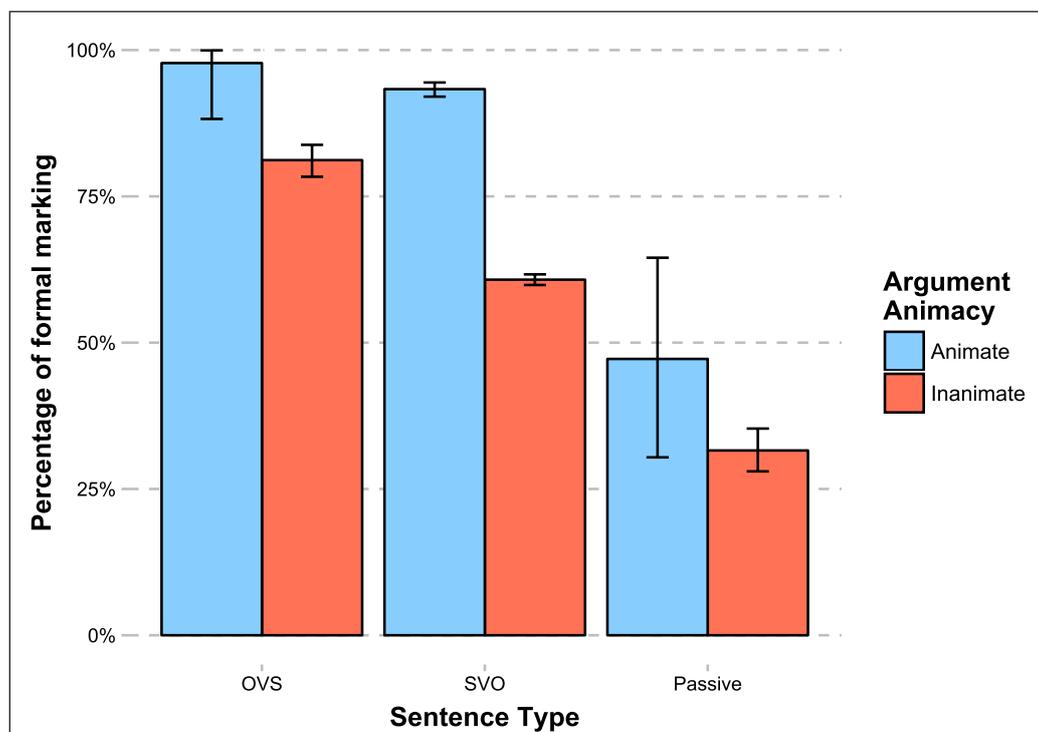


Figure 2: The percentage of formal marking in OVS, SVO and passive sentences, differentiating between animate and inanimate argument sentences. Error bars illustrate 95% confidence intervals, calculated on the basis of normal approximation.

Type \times Animacy interaction as fixed effects. Inanimate argument SVO sentences served as the reference sentence type, so that the model evaluates to what extent the formal marking of all other sentence types differ with respect to such sentences. The model also included a random intercept for Genre as well as a by-genre random slope for Sentence Type. The random effects structure was selected by fitting a model with a maximal random effects structure, and then subsequently eliminating random effects using backwards elimination. Only those effects that significantly improved the model's predictive ability were included in the final model.

A significant effect of Animacy was found, $\beta = 1.86$, $z = 18.47$, $p < .0001$. In other words, SVO sentences more frequently contain formal marking when they contain two animate arguments in comparison to when they at most contain one. There was also a significant effect of OVS word order, $\beta = 1.003$, $z = 7.48$, $p < .0001$, showing that inanimate argument OVS sentences are more frequently formally marked than inanimate argument SVO sentences. The OVS sentence \times Animacy interaction was not significant, however, showing that the effect of argument animacy is the same in SVO and OVS sentences: the frequency of formal marking is increased to the same extent in both SVO and OVS sentences. On the other hand, there was a negative effect of passive word order, $\beta = -0.99$, $z = 8.68$, $p < .0001$, showing that inanimate argument passive sentences are significantly *less* likely to contain formal markers than inanimate argument SVO sentences. There was also a negative Passive sentence \times Animacy interaction that is significant, $\beta = -1.26$, $z = 3.43$, $p < .001$, showing that the frequency of formal marking is *less* affected by argument animacy in passive sentences in comparison to SVO sentences.

These findings show that whereas OVS sentences generally are more likely to be formally marked than SVO sentences, passive sentences are less likely to contain formal marking. Further, all three sentence types are more likely to contain formal marking when they contain two animate arguments in comparison to when they contain at most one, although this effect of argument animacy on the frequency of formal marking is less pronounced in passive sentences.¹⁶

5 Discussion

Swedish is a language that lacks subject-verb agreement and that only has case marking on personal pronouns. Word order is therefore of particular importance for determining the grammatical functions of NP arguments. In the majority of transitive sentences, the subject precedes the direct object. Yet, Swedish allows for OVS word order, although such sentences are potentially ambiguous with respect to grammatical functions. This study therefore investigated the functional motivations behind the use of object-initial word order. It also investigated whether the object-initial word order is dispreferred when the grammatical functions cannot be determined on other information types, which would provide evidence for the hypothesis that writers tend to avoid OVS word order in the face of potential ambiguity. Using corpus data of written Swedish, these questions were investigated on the basis of quantitative differences in the distribution of NP prominence properties (i.e., givenness, definiteness, pronominality, person and animacy) and distributional differences of formal, morphosyntactic markers of grammatical functions (i.e., case marking, auxiliary verbs and verb particles), comparing OVS sentences to both SVO sentences and passives. In the following, I first discuss the findings regarding distributional differences of NP prominence properties and their implications for the functional motivations behind the use of OVS word order in Swedish. I then turn to the findings

¹⁶ Simple effects analyses show that the difference in the percentage of formal marking between animate argument passives and inanimate argument passives is significant.

about differences in the distribution of formal markers and what they imply for the question of whether speakers tend to avoid using object-initial word order when it potentially will result in an ambiguity.

5.1 The motivations for the object-initial word order in Swedish

As discussed in Section 2.1, word order variation is generally assumed to be motivated by information structural considerations. In particular, direct object fronting in Swedish is usually considered to be used when the object (rather than the subject) is topical and therefore high in discourse prominence (Teleman & Wieselgren 1970; Teleman et al. 1999 (4): 432; Rahkonen 2006; Bohnacker & Rosen 2008; 2009; Bohnacker 2010). Rahkonen (2006: 45) claimed that OVS sentences in which the object is low in discourse prominence do not “represent the typical information structure of OVS sentences”. Although the findings of the present study show that objects more frequently are high in discourse prominence – and in particular text deictic – when positioned sentence-initially, sentence-initial objects that are discourse new are quite common. In fact, sentence-initial objects are significantly less frequently discourse given than sentence-initial subjects. These findings show that the object-initial word order not only is used when the object is topical and therefore discourse prominent. They also to some extent call into question Bohnacker & Rosén’s (2008) claim that the preference to start a sentence with highly discourse prominent, given or thematic information is stronger in Swedish than in German.

In line with Molnár & Winkler (2010), I instead suggest that object fronting often is used when the object is contrastive, that is, that it is contrasted against a set of alternatives that have been introduced in the previous discourse. Following Molnár & Winkler (2010), I assume that contrast is orthogonal to topic and focus, in the sense that both topical and focused NPs can be contrastive.

I also suggest that object fronting is used to introduce new topics into the discourse, that is, to mark a discourse *topic shift* (James 1995). In such cases, a discourse new object NP is positioned in the sentence-initial position in order to emphasise that it contrasts with the previous discourse topic. Object fronting is in such cases used to establish discourse coherence between the old and the new discourse topic (cf., Molnár & Winkler 2010). Apart from the “informationally heavy” sentence-initial object NP that is discourse new, such constructions tend to be “informationally light” in that they often contain a semantically weak main verb, such as a possessive verb, and a post-verbal subject that is given in the discourse and therefore high in discourse prominence. Further indirect evidence for this account comes from a study by Qian and Jaeger (2011). They found the information content of full sentences, that is, the in-context predictability of whole sentences, to be negatively correlated with the strength of the discourse topic shift in those sentences, as estimated on the basis of topic modelling. In other words, speakers appear to avoid unpredictable information in sentences with new discourse topics, in order to keep the level of information flow as constant as possible throughout the continuation of the sentence (Shannon 1948; Fenk-Oczlon 2001; Jaeger 2010). This account is also corroborated in the Grammar of the Swedish Academy. According to it, the meaning of an OVS sentence following a sentence-initial object that is discourse new and focused is likely to be predictable, so that the comprehender can infer it upon encountering the object (see Teleman et al. 1999 (4): 432).

5.2 Ambiguity avoidance in object-initial sentences

In this study, I investigated whether writers tend to avoid potentially ambiguous OVS sentences which either lack formal information about grammatical functions – case marked personal pronouns, auxiliary verbs, and verb particles – or in which those functions cannot be determined on the basis of animacy.

All three types of formal information regarding grammatical functions were found to occur most frequently in OVS sentences, followed by SVO sentences, and least frequently in passives. This is exactly the kind of pattern that can be expected if writers actively adapt their language in order to avoid potential ambiguities. Sentences with OVS word order almost always contain one or several formal markers that can be used to unambiguously assign grammatical functions. In OVS sentences, readers cannot rely on word order, and other formal means are often required. SVO-sentences less frequently occur with formal markers, although they quite commonly are used when the argument functions cannot be determined on the basis of animacy. In SVO sentences, readers can resort to word order when no other means are available. Formal markers are therefore not crucial for interpretation. Here, additional formal markers serve to assist the interpretation process, which is of particular importance when grammatical functions cannot be determined on the basis of animacy. In passives, on the other hand, additional formal markers are in fact quite rare. Even in passive sentences with two animate arguments, they occur in less than 50% of the occurrences. This is because the passive, which is syntactically intransitive, contains several unambiguous morphosyntactic markers to grammatical functions (i.e., the passive marker and the *av* preposition). Additional means will therefore most likely not facilitate interpretation further, independent of whether the argument functions can be determined on the basis of an argument animacy difference.

These results are in line with those of Rahkonen (2006) who found OVS sentences to occur more frequently with case marked personal pronoun subjects than both SVO sentences and OSV object relative clauses. Although Rahkonen's findings indicate that writers tend to avoid OVS sentences without a case marked subject, Rahkonen (2006) instead suggested that they stem from a dispreference for using passive sentences in which the Actor argument is realised as a personal pronoun (i.e., as the argument of a prepositional *av* ('by') phrase). Accordingly, writers tend to resort to the object-initial construction whenever the Actor needs to be realised as a personal pronoun, in order to avoid using a prepositional *av* ('by') phrase with a personal pronoun argument. The results of the present study speak against this interpretation, since they show that it is not only case marked personal pronoun Actors that occur less frequent in passive sentences, but also auxiliary verbs and verb particles. Since all of these information types serve as formal cues to grammatical functions in Swedish, the most plausible interpretation is instead that writers tend to resort to the passive construction – which is structurally unambiguous with respect to grammatical functions – whenever the target sentence at hand will lack any of these formal cues to grammatical functions (see Pitz 2006 for similar ideas regarding the use of the passive in Norwegian).

It should be stressed, that the use of the passive is not only motivated by a need to topicalise or foreground the Undergoer. As shown in Section 4.5.2, the Actor argument of passives is much more often low in discourse prominence than in both SVO and OVS sentences. This is possibly because the passive also serves to background the Actor argument (Haspelmath 1990; Foley 2011). This is likely to more frequently be done with Actors that are low in discourse prominence. The passive construction have also been claimed to express inactivization of the event or situation, in the sense that the Actor no longer is responsible for the activity (Haspelmath 1990), as well as to modify the aspect of the event (Landén & Molnár 2003). Thus, the choice between the passive and the object-initial construction is obviously not *only* a choice that is driven by a motivation to avoid potential ambiguities. Also other information structural and semantic considerations come into play.

The present study also found that OVS sentences less frequently contain two animate NP arguments, than both SVO sentences and passives. This finding indicates that writers tend

to avoid using the OVS word order when the argument functions cannot be determined on the basis of animacy. It was further found that all three types of formal markers are used more frequently in sentences with two animate arguments than in sentences with at most one animate argument. This is not only the case in potentially ambiguous OVS sentences, but also in both SVO sentences as well as in passives. Writers therefore seem to be more inclined to provide formal markers to grammatical functions in cases where they might not be able to be determined on the basis of animacy.

Similar findings were done by Rahkonen (2006), who found case marked subjects in OVS sentences to be somewhat more frequent in sentences with two animate arguments in comparison to sentences with one animate argument. Rahkonen (2006) did not, however, interpret these findings as being driven by a motivation to avoid ambiguities. He argued that although OVS sentences with two animate arguments might be ambiguous with respect to their argument functions in isolation, this is very seldom the case in context. Semantically ambiguous sentences are therefore likely to be unproblematic to interpret in their discourse contexts, and there is no need to provide additional morpho syntactic information regarding the argument functions. However, Rahkonen (2006) failed to appreciate that language comprehension most likely involves interpretation at two or several different stages, and that the discourse context is taken into account only at the final stage of interpretation. More specifically, several models of language comprehension assume that interpretation of local structure on the sentence level – at which point participant role assignment occurs – temporarily precedes interpretation of global structure on the discourse level (e.g., Bornkessel-Schlesewsky & Schlewsky 2006; Friederici 2011). Although the argument functions might be able to be determined on the basis contextual information during the final stage of language comprehension, the comprehension process is still likely to be hampered at the earlier stage of local structure building. Contrary to Rahkonen's claims then, the use of formal markers in potentially ambiguous OVS sentences is therefore motivated by the need to facilitate language comprehension during local structure building, even if those sentences are unambiguous in their discourse context.

Overall then, the findings of the present study provides rather compelling evidence for the idea that writers are inclined to actively avoid using object-initial sentences when this might result in an ambiguity. As such, these findings provide additional evidence for the hypothesis that language producers – in this case writers – are inclined to actively avoid ambiguities. It should be noted that this tendency, which is probabilistic rather than absolute, most likely is driven by a trade-off between a motivation to avoid using redundant information, on the one hand, and a motivation to make the message informative enough for the reader, on the other. As mentioned in Section 2.2, Kurumada & Jaeger (2015) found Japanese speakers to be more prone to use overt object case marking in subject-initial transitive sentences where the function of the object argument was harder to infer on the basis of animacy or plausibility information. However, in all sentences in their study, the sentence-initial subject was unambiguously case marked, so that the argument functions could readily be assigned already upon encountering the subject. The results of their study therefore seem to reflect a tendency for speakers to balance their production efforts between avoiding redundant information, on the one hand, and providing enough information in order for the listener to effectively comprehend the message, on the other. There is no reason not to assume that the production of written language is influenced by similar principles. That is, writers of Swedish might be more inclined to avoid the object-initial construction when no other means for identifying grammatical functions – such as case marking or animacy information – will be available in the target sentence.

6 Conclusions

This study has investigated the conditions under which object-initial word order is used in written Swedish. It has also investigated whether the use of object-initial word order is dispreferred when additional information about grammatical functions, such as formal markers or animacy information, is unavailable. These issues were investigated on the basis of quantitative differences in written corpus distributions of NP prominence properties (e.g., givenness, and animacy) and other morphosyntactic cues (e.g., auxiliary verbs and case marking) in OVS sentences, SVO sentences, and passives. The study is the first to investigate distributional differences of these properties between these three sentence types. It therefore provides new and valuable insights about when Swedish OVS sentences are used and when they are dispreferred.

The results of the study show that the object-initial word order not only is used when the object is topical and discourse prominent. Object fronting can also be used either to indicate that the object is contrastive, or in order to introduce new topic NPs into the discourse. These two strategies are related in that the new topic NP is positioned in the sentence-initial position in order to emphasise its contrast with the previous discourse topic. Such topic-introducing constructions tend to be “informationally light” in that they often contain a semantically weak main verb and a highly discourse prominent post-verbal subject. In line with earlier accounts (Shannon 1948; Fenk-Oczlon 2001; Jaeger 2010), I suggest that this is a reflex of a general preference for keeping the information flow as constant as possible throughout the discourse. An initial NP that introduces a new discourse topic is “informationally heavy”. There is therefore a preference for keeping the rest of the sentence informationally low and predictable in order to keep the discourse information flow balanced.

The results also show that the object-initial word order is used less frequently when the target sentence either will lack formal information about grammatical functions, or when those functions cannot be determined on the basis of animacy. In such situations, writers instead most frequently resort to a passive sentence. Whereas OVS sentences are potentially ambiguous with respect to grammatical functions in that those functions cannot be assigned on the basis of an SVO-word order preference, passive sentences are always unambiguous since they contain additional morphosyntactic information about grammatical functions. These findings therefore provide additional support for the hypothesis that language producers – in the present case writers – are inclined to actively avoid potentially ambiguous sentences.

Abbreviations

1 = first person, 2 = second person, 3 = third person, ACC = accusative, NOM = nominative, OVS = object-verb-subject, PASS = passive, REFL = reflexive, SG = singular, SVO = subject-verb-object.

Additional Files

The additional files for this article can be found as follows:

- **Appendix 1.** Corpus searches and exclusions. DOI: <https://doi.org/10.5334/gjgl.502.s1>
- **Appendix 2.** Search strings. DOI: <https://doi.org/10.5334/gjgl.502.s2>

Acknowledgements

I am grateful to Östen Dahl as well as to five anonymous reviewers for comments and suggestions on previous versions of this manuscript.

Competing Interests

The author has no competing interests to declare.

References

- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, Jennifer E., Thomas Wasow, Ash Asudeh & Peter Alrenga. 2004. Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language* 51(1). 55–70. DOI: <https://doi.org/10.1016/j.jml.2004.03.006>
- Bader, Markus & Jana Häussler. 2010. Word order in German: A corpus study. *Lingua* 120(3). 717–762. DOI: <https://doi.org/10.1016/j.lingua.2009.05.007>
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bickel, Balthasar. 2010. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199281251.013.0020>
- Bohnacker, Ute. 2010. The clause-initial position in L2 Swedish declaratives: Word order variation and discourse pragmatics. *Nordic Journal of Linguistics* 33(2). 105–143. DOI: <https://doi.org/10.1017/S033258651000017X>
- 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. DOI: <https://doi.org/10.1017/S0272263108080741>
- Bohnacker, Ute & Christina Rosén. 2009. Fundamentet i svenskan och tyskan – Syntax och informationsstruktur: Ett problemområde för språkinläring och undervisning [The fundament in Swedish and German – Syntax and information structure: A problem area for language acquisition and education]. *Språk & Stil* [Language & style] 19. 142–171.
- Bornkessel, Ina & Matthias Schlesewsky. 2006. The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review* 113(4). 787–821. DOI: <https://doi.org/10.1037/0033-295X.113.4.787>
- Bornkessel-Schlesewsky, Ina & Matthias Schlesewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3(1). 19–58. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00099.x>
- Bouma, Gerlof. 2008. *Starting a sentence in Dutch* (Groningen Dissertations in Linguistics 66). Groningen: Center for Language and Cognition, University of Groningen dissertation.
- Bouma, Gerlof. 2011. Production and comprehension in context: The case of word order freezing. In Anton Benz & Jason Mattausch (eds.), *Bidirectional optimality theory*, 169–190. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.180.07bou>
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.
- Dahl, Östen. 2000. Egophoricity in discourse and syntax. *Functions of Language* 7(1). 37–77. DOI: <https://doi.org/10.1075/fof.7.1.03dah>
- Dahl, Östen. 2008. Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua* 118. 141–150. DOI: <https://doi.org/10.1016/j.lingua.2007.02.008>
- Dahl, Östen & Kari Fraurud. 1996. Animacy in grammar and discourse. In Thorsten Fretheim & Jeanette K. Gundel (eds.), *Reference and referent accessibility* (Pragmatics & Beyond New Series), 47–87. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/pbns.38.04dah>
- Du Bois, John. 2003. Argument structure: Grammar in use. In John Du Bois, Lorraine Kumpf & William Ashby (eds.), *Preferred argument structure: Grammar as architecture for function*, 11–60. Amsterdam: John Benjamins.

- Einarsson, Jan. 1976a. Talbankens skriftspråkskonkordans [The speech bank's concordance of written language]. Unpublished Work.
- Einarsson, Jan. 1976b. Talbankens talspråkskonkordans [The speech bank's concordance of spoken language]. Unpublished Work.
- Engdahl, Elisabet. 2001. Valet av passivform i modern svenska [The choice of passive form in modern Swedish]. In *Svenskans Beskrivning 24. Förhandlingar vid Tjugofjärde Sammankomsten för Svenskans Beskrivning* [The Description of Swedish 24. Negotiations at the twenty-fourth union for the description of Swedish], 81–90. Linköping: Linköping University Electronic Press.
- Engdahl, Elisabet. 2006. Semantic and syntactic patterns in Swedish passives. In Benjamin Lyngfelt & Torgrim Solstad (eds.), *Demoting the agent: Passive, middle and other voice phenomena*, 21–45. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.96.04eng>
- Engdahl, Elisabet & Anu Laanemets. 2015. Prepositional passives in Danish, Norwegian and Swedish: A corpus study. *Nordic Journal of Linguistics* 38(3). 285–337. DOI: <https://doi.org/10.1017/S0332586515000232>
- Engdahl, Elisabet & Filippa Lindahl. 2014. Preposed object pronouns in mainland Scandinavian. *Working Papers in Scandinavian Syntax* 92. 1–32.
- Erteschik-Shir, Nomi. 2007. *Information structure* (Oxford Surveys in Syntax and Morphology). Oxford: Oxford University Press.
- Fenk-Oczlon, Gertrud. 2001. Familiarity, information flow, and linguistic form. In Joan Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 431–448. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.45.22fen>
- Ferreira, Victor S. 2008. Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation* 49. 209–246. DOI: [https://doi.org/10.1016/S0079-7421\(08\)00006-6](https://doi.org/10.1016/S0079-7421(08)00006-6)
- Ferreira, Victor S. & Gary S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40(4). 296–340. DOI: <https://doi.org/10.1006/cogp.1999.0730>
- Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–25. London: Holt, Rinehart and Winston.
- Foley, William A. 2011. A typology of information packaging in the clause. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. 1: Clause structure*, 362–446. Cambridge: Cambridge University Press.
- Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Friederici, Angela D. 2011. The brain basis of language processing: From structure to function. *Physiological Reviews* 91(4). 1357–1392. DOI: <https://doi.org/10.1152/physrev.00006.2011>
- Gelman, Andrew & Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511790942>
- Givón, Talmy. 1979. *On understanding grammar*. New York, NY: Academic Press.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307. DOI: <https://doi.org/10.2307/416535>
- Gundel, Jeanette K. & Thorstein Fretheim. 2004. Topic and focus. In Laurence Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 174–196. Oxford/Cambridge, MA: Blackwell.

- Gustafson-Capková, Sofia & Brita Hartmann. 2006. Manual of the Stockholm Umeå Corpus version 2.0. Unpublished Work.
- Haspelmath, Martin. 1990. The grammaticalization of passive morphology. *Studies in Language* 14(1). 25–72. DOI: <https://doi.org/10.1075/sl.14.1.03has>
- Haywood, Sarah L., Martin J. Pickering & Holly P. Branigan. 2005. Do speakers avoid ambiguities during dialogue? *Psychological Science* 16(5). 362–366. DOI: <https://doi.org/10.1111/j.0956-7976.2005.01541.x>
- Holm, Sture. 1979. A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics* 6. 65–70.
- Hörberg, Thomas. 2016. *Probabilistic and prominence-driven incremental argument interpretation in Swedish*. Stockholm: Department of Linguistics, Stockholm University dissertation.
- Hörberg, Thomas, Maria Koptjevskaja-Tamm & Petter Kallioinen. 2013. The neurophysiological correlate to grammatical function reanalysis in Swedish. *Language and Cognitive Processes* 28(3). 388–416. DOI: <https://doi.org/10.1080/01690965.2011.651345>
- Howell, David. 2010. *Statistical methods for psychology*. 7th edition; Belmont, CA: Cengage Wadsworth.
- Jaeger, T. Florian. 2006. *Redundancy and reduction in spontaneous speech*. Stanford, CA: Stanford University dissertation.
- Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: <https://doi.org/10.1016/j.jml.2007.11.007>
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: <https://doi.org/10.1016/j.cogpsych.2010.02.002>
- Jaeger, T. Florian. 2011. Corpus-based research on language production: Information density and Reducible Subject Relatives. In Emily M. Bender & Jennifer E. Arnold (eds.), *Language from a cognitive perspective: grammar, usage and processing*. *Studies in honor of Tom Wasow*, 161–198. Stanford, CA: CSLI Publication.
- James, Angela D. 1995. Topic shift in casual conversation. *Totem: The University of Western Ontario Journal of Anthropology* 2(1). 33–36.
- Jørgensen, Nils. 1976. *Meningsbyggnaden i talad svenska* [The sentence structure in spoken Swedish]. Lund: Department of Scandinavian Languages, Lund University dissertation.
- Josefsson, Gunlög. 2012. What's first? On the left periphery in Swedish declaratives. In Johan Brandtler, David Håkansson, Stefan Huber & Eva Klingvall (eds.), *Discourse & grammar: A festschrift in honor of Valéria Molnár*, 359–372. Lund: Centre for Languages and Literature, Lund University.
- Kempen, Gerard & Karin Harbusch. 2004. A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. *Trends in linguistics studies and monographs* 157. 173–182. DOI: <https://doi.org/10.1515/9783110894028.173>
- Kirri, Arto. 1975. Studier över passivkonstruktioner i nysvenskt skriftspråk [Studies about passive constructions in new-swedish writing]. *Meddelanden från institutionen för nordisk filologi vid Helsingfors universitet. Serie A, nr. 2* [Messages from the department of nordic philology at the University of Helsinki. Series A, no. 2]. Helsinki: University of Helsinki.
- König, Ester & Wolfgang Lezius. 2003. The TIGER language – A description language for syntax graphs. *Formal definition*. Unpublished Work.

- König, Ester, Wolfgang Lezius & Holger Voorman. 2003. TIGERSearch 2.1. Unpublished Work.
- Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Fery, Gisbert Fanselow & Manfred Krifka (eds.), *Interdisciplinary studies of information structure* 6. 13–55. Potsdam: Universitätsverlag Potsdam.
- Kurumada, Chigusa & T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152–178. DOI: <https://doi.org/10.1016/j.jml.2015.03.003>
- Laanemets, Anu. 2012. *Passiv i moderne dansk, norsk og svensk, Et korpusbaseret studie af tale- og skriftsprog*. [Passive in modern Danish, Norwegian and Swedish. A corpus-based study of spoken and written language.] Tartu: University of Tartu dissertation.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511620607>
- Landén, Barbro & Valéria Molnár. 2003. Passive as activity aspect. On the relation between aspectuality and passive: A comparative study of German and Swedish. In Jorunn Hetland & Valéria Molnár (eds.), *Structures of focus and grammatical relations*, 187–230. Berlin: De Gruyter.
- Lee, Hanjung. 2006. Parallel optimization in case systems: Evidence from case ellipsis in Korean. *Journal of East Asian Linguistics* 15(1). 69–96. DOI: <https://doi.org/10.1007/s10831-005-3004-1>
- Levinson, Stephen C. 2004. Deixis. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 97–121. Oxford/Cambridge, MA: Blackwell Publishing.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511605789>
- Mahowald, Kyle. 2011. An LGF approach to word order freezing. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG11 Conference*. Stanford, CA: CSLI Publications.
- Molnár, Valéria. 2002. Contrast – from a contrastive perspective. *Language and Computers* 39. 148–161. DOI: https://doi.org/10.1163/9789004334250_010
- Molnár, Valéria. 2006. On different kinds of contrast. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 197–233. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110922011.197>
- Molnár, Valéria & Susanne Winkler. 2010. Edges and gaps: Contrast at the interfaces. *Lingua* 120(6). 1392–1415. DOI: <https://doi.org/10.1016/j.lingua.2008.08.010>
- Nordman, Marianne. 1992. *Svenskt fackspråk* [Swedish technical language]. Lund: Studentlitteratur.
- Øvrelid, Lilja. 2004. Disambiguation of syntactic functions in Norwegian: Modeling variation in word order interpretations conditioned by animacy and definiteness. In *Proceedings of the 20th Scandinavian Conference of Linguistics*, 1–17. Helsinki: University of Helsinki.
- Piantadosi, Steven T., Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122(3). 280–291. DOI: <https://doi.org/10.1016/j.cognition.2011.10.004>
- Pitz, Annelise P. 2006. The relation between information structure, syntactic structure and passive. In Benjamin Lyngfelt & Torgrim Solstad (eds.), *Demoting the agent: Passive, middle and other voice phenomena*, 21–45. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.96.11pit>
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York, NY: Academic Press.

- Prince, Ellen F. 1984. Topicalization and left-dislocation: A functional analysis. *Annals of the New York Academy of Sciences* 433(1). 213–225. DOI: <https://doi.org/10.1111/j.1749-6632.1984.tb14769.x>
- Qian, Ting & T. Florian Jaeger. 2011. Topic shift in efficient discourse production. In *Proceedings of the 33rd annual conference of the Cognitive Science Society*, 3313–3318.
- Rahkonen, Matti. 2006. Some aspects of topicalization in Swedish declaratives. *Linguistics* 44(1). 23–55. DOI: <https://doi.org/10.1515/LING.2006.002>
- R Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94.
- Roland, Douglas, Jeffrey L. Elman & Victor S. Ferreira. 2006. Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98(3). 245–272. DOI: <https://doi.org/10.1016/j.cognition.2004.11.008>
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1. 75–116. DOI: <https://doi.org/10.1007/BF02342617>
- Shannon, Claude. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. DOI: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Silén, Beatrice. 1997. Agentadverbialet i modern svenska [Agent adverbial in modern Swedish]. *Studier i nordisk filologi* [Studies in Nordic philology] 76. Helsinki: Svenska litteratursällskapet i Finland [Swedish literature society in Finland].
- Sköld, Tryggve. 1970. The Object in Pre-predicative Position in Swedish. In *Acta Societatis Linguisticae Upsaliensis* (Nova Series 3, vol. 2), 57–98. Stockholm: Almqvist & Wiksell.
- Snider, Neal & Annie Zaenen. 2006. Animacy and syntactic structure: Fronted NPs in English. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 323–339. Stanford, CA: CSLI Publications.
- Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik* [The grammar of the Swedish Academy]. 1st edition; Stockholm: Svenska Akademien.
- Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79(3). 464–484. DOI: <https://doi.org/10.1353/lan.2003.0189>
- Van Valin, Robert D. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511610578>
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Weber, Andrea & Karin Müller. 2004. Word order variation in German main clauses: A corpus analysis. In *Proceedings of the 20th International Conference on Computational Linguistics*, 71–77.
- Westman, Margareta. 1974. *Bruksprosa* [Use prose]. Lund: Liber Läromedel.

How to cite this article: Hörberg, Thomas. 2018. Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account. *Glossa: a journal of general linguistics* 3(1): 81. 1–36, DOI: <https://doi.org/10.5334/gjgl.502>

Submitted: 06 August 2017 **Accepted:** 25 April 2018 **Published:** 20 July 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 