Appendix 1. Corpus searches and exclusions

Corpus searches

Searches were conducted with TIGER search 2.1. (König et al. 2003) with search strings in the TIGER search query language (König & Lezius 2003). Three classes of search patterns were used to find the three general syntactic patterns of subject-initial, object-initial and passive sentences. These syntactic patterns are shown and exemplified in Table A.1. The actual search strings are provided in Appendix 2. The patterns contain noun phrases (NP), adverb phrases (AdvP), verbs (V1, V2, V3, V4) and verb particles (PT). The initial verbs (V1) were always finite, the second verb (V2) was either an infinite verb or an infinitival marker, and the final verbs (V3 and V4) were infinite. In passives, the final verb of the sentence at hand was also required to inflected with the passive marker. Passives were also required to end with a prepositional *av* ('by') phrase. In the table, optional elements are shown in parentheses.

Table A.1. The three general syntactic patterns corresponding to the search patterns searching for SVO, OVS and passive sentences. Optional constituents / words are shown in parentheses. NP = Nominal Phrase; AdvP = Adverbial Phrase; V = Verb (V_1 = finite; V_2 = infinite or infinitival marker *att*; V_3 + V_4 = Infinite); PT = Verb particle. In passive sentences, the final verb was required to be inflected with a passive marker.

	Search pattern
SVO	[NP] ([ADVP]) ([ADVP]) [V_1] ([ADVP]) ([ADVP]) ([V_2]) ([V_3]) ([V_4]) ([PT]) [NP]
	[Du som har examen enligt den tidigare studieordningen och uppfyller de nya kraven för en magisterexamen$_{NP}$] [kommer$_{V1}$] [att$_{V2}$] [kunna$_{V3}$] [ta$_{V4}$] [ut$_{PT}$] [en sådan$_{NP}$]
	'You who have a degree according to the previous curriculum and is eligible for a magister's degree will be able to acquire such'
OVS	[NP] [V_1] ([ADVP]) ([ADVP]) [NP] ([ADVP]) ([ADVP]) [V_2] ([V_3]) ([V_4]) ([PT])
	[Några egentliga genombrott från de informella överläggningarna ute på Bushs lantställe Camp David$_{NP}$] [kunde$_{V1}$] [de$_{NP}$] [inte$_{ADVP}$] [redovisa$_{V2}$]
	'Any serious breakthroughs from de informal negotiations at Bush's country house Camp David they were unable to report'
Passive	[NP] ([ADVP]) ([ADVP]) [V_1] ([ADVP]) ([ADVP]) [V_2] ([V_3]) ([V_4]) ([PT]) [av [NP]]
	[En gemensam styrka med ett 70-tal frivilliga poliser$_{NP}$] [har$_{V1}$] [satts$_{V2}$] [upp$_{PT}$] [av [alla stadens polisdistrikt$_{NP}$].
	'A joint force of about 70 volunteering polices has been commissioned by all of the city's police districts'

Due to inconsistencies and errors in the syntactic annotation of the corpus, search strings relied on both hierarchical dependencies and linear order. Multi-word constituents were defined with reference to the directly dominating node as well as to the left- and right-most word nodes dominated by the constituent node at hand. Consecutive search queries that defined syntactic relationships between constituents would then make reference to precedence relations between either the right- or the left-most word node of the constituent node at hand and some other node. In order to allow for the structural variation shown in Table A.1, the search queries contained disjunctions of alternative search patterns that were organized in a nested fashion. For example, search patterns for subject-initial sentences without adverbial phrases had the following simplified structure:

[NP1] [V1] & (([V1] [NP2] | [V1] [PL] [NP2])
| [V1] [V2] & (([V2] [NP2] | [V2] [PL] [NP2])
| [V2] [V3] & ([V3] [NP2] | [V3] [PL] [NP2])
| [V3] [V4] & ([V4] [NP2] | [V4] [PL] [NP2]))))

Which is equivalent to the following pattern, in which parentheses represent optional elements:

[NP1] [V1] ([V2]) ([V3]) ([V4]) ([PL]) [NP2]

In order to limit search times, four different search strings were used for SVO sentences and passives, respectively. The first pair of search strings (i.e., one for SVO sentences and a corresponding for passives) searched for sentences with no adverbial phrase, the second pair for sentences with a preverbal adverbial phrase, the third for sentences with a post-verbal adverbial, and the fourth for sentences with both a pre- and post-verbal adverbial phrase. For OVS sentences, only one search string was required.

Exclusions

The initial data set with false inclusions consisted of a total of 32,406 sentences. Some of these were duplicates that had to be removed. For example, in some cases the sentence final NP dominates one or multiple NP nodes due to errors in the corpus annotation (such as in, e.g, [[*en tredje klassens*$_{NP}$] *medborgare*$_{NP}$] – 'a third grade citizen'). The search patterns therefore find the "correct" version of the sentence with the dominating NP, together with the "incorrect" versions with the dominated NPs. All other exclusion types are listed in Table B2 and discussed in the following.

Direct questions and *non-transitive sentences* were incorrectly included in the initial data set due to annotation errors, or due to containing an indirect object or a non-argument NP, and were excluded. Sentences with *non-argument NPs* contained embedded adverbial phrases that in many cases were mistaken for the direct object NP in both intransitive and transitive clauses (i.e, when consisting of an NP adjunct, see Teleman et al. 1999 (3): 64). All sentences with embedded adverbial phrases were therefore excluded.

In *non-NP object sentences*, the direct object either consists of an embedded clause (Teleman et al. 1999 (3): 293) or an infinitival complement phrase (Teleman et al. 1999 (3): 292), rather than an NP. These were also excluded. However, sentences with the future construction *kommer att* (e.g., *Era barn kommer att tacka er* – 'Your children will thank you') were kept, although from a purely syntactic perspective, they do contain an infinitival complement phrase. The reason for keeping them is that the *kommer att* future construction is highly grammaticalized in Swedish, and therefore should be considered a complex future tense construction that, together with the lexical verb (e.g., *tacka* in the example), functions as the sentence predicate. Ditransitive sentences with an infinitival phrase argument (e.g., *Någon rekommenderade mig att pröva komage* – 'Someone advised me to try cow stomach') and

Table A.2. Excluded clause types and example sentences of each type.

Clause type		**Example**
Direct question		Var har [magin$_{NP}$] tagit [vägen$_{NP}$] , Morrissey? Where has the magic gone, Morrissey?
Intransitive sentence		[Sommartid$_{NP}$] går [en buss$_{NP}$] härifrån till Bourg St Maurice In the summer a buss departs from here to Bourg St Maurice
Incorrectly annotated sentence		Eller [rättare sagt$_{*NP}$] återupplivar [en gammal$_{NP}$] Or rather revive an old one
Non-NP object	clause	[En sådan människa$_{NPi}$] trodde [jag$_{NP}$] [att jag var [__$_{NPi}$]$_S$] Such a person I thought I was
	infinitival phrase	När [Aktuellt$_{NP}$] väljer [att följa den linjen$_{InfP}$] When Aktuellt choose to take that approach
Dummy argument	Subject	[det$_{NP}$] står [20 studenter$_{NP}$] i korridorerna 20 students stand in the hallways
	object	Tillsammans ska [vi$_{NP}$] bona om [det$_{NP}$] där hemma Together we will make it more habitable at home
Embedded adverbial phrase		när [han$_{NP}$] [den där höstdagen 1977$_{NP}$] skjutsade [henne$_{NP}$] when he in the autumn of 1977 drove her
Ditransitive sentence	NP	Och [bättre betyg$_{NP}$] kan [jag$_{NP}$] inte ge [mig själv$_{NP}$] And I can´t give my self a better score
	Clause	[Jag$_{NP}$] frågar [Jane Landeryou$_{NP}$] [varför hon är emot tjurfäktningar$_S$] I ask Jane Landeryou why she is against bullfights
Predicative sentence	subject predicative	[Var tredje människa$_{NP}$] är [analfabet$_{NP}$] Every third person is illiterate
	object predicative	[Hummern$_{NP}$] kallar [Per Moksnes$_{NP}$] för Rocky The lobster Per Moksnes calls Rocky
	object with infinitive	[Jag$_{NP}$] såg [världen$_{NP}$] gå under I saw the world perish
	causative sentence	[En hastig ingivelse$_{NP}$] får [mig$_{NP}$] att härma tjuren A sudden impulse makes me imitate the bull
Idiomatic expression	idiom	Även [fack och forskare$_{NP}$] får [sin släng av sleven$_{NP}$] Also the unions and scientists get theirs (lit: Also the unions and scientists get their throw of the trowel)
	lexicalisation	[alla invånare$_{NP}$] [skulle ha [råd$_{NP}$] med [bil$_{NP}$]$_{VP}$] all citizens would afford a car
Light constituent	light verb	[Demodokos$_{NP}$] gör [ett besök$_{NP}$] i underjorden Demodokos makes a visit to the underworld
	light object	[det$_{NP}$] kostar [pengar$_{NP}$] It costs money
		[han$_{NP}$] målar [massor$_{NP}$] He paints a lot

sentences with an adverbial complement that is required by the verb (Teleman et al. 1999 (3): 435) (e.g., *Vi lämnade bilen under lindarna i Berlin* – 'we left the car under the lindens in Berlin') were also kept, although the infinitival and adverbial phrases of such sentences can be considered to be additional arguments of the verb.

Dummy argument sentences with either a dummy subject or direct object with a purely grammatical function were also excluded. This was also the case for all types of *predicative sentences* and related clause types (see Table A.2 for examples), elaborated upon in the following. In predicative sentences, a predicative expression is used to assign a property to either the subject (subject predicative) or the object (object predicative, more commonly referred to as *secondary predication*, Monica-Alexandrina 2005; Teleman et al. 1999 (3): 325). Subject predicatives contain a copula verb and a nominal predicative that does not function as an argument, but rather is (part of) the sentence predicate. Object predicatives, on

the other hand, contain a complement or an adjunct predicative expression that assigns a property to the object NP, and thereby functions as a secondary predicate of the object. Object with infinitive constructions, often referred to as *raising* constructions, are functionally similar to object predicatives, in that they contain an infinitival complement phrase that functions as a predicate of the object (Brolin 2006; Teleman et al. 1999 (3): 575). In causative clauses, the subject argument denotes *the cause* of the event expressed by an additional complement phrase, but the actual Actor of that event is denoted by the object. The complement phrase of the causative clause therefore functions as a secondary predicate of the object, as in object predicative and object with infinitive constructions.

Sentences with *idiomatic expressions* and lexical predicate constructions were also excluded. The former clause type consists of idioms *per se*, that is, sentences with a non-compositional meaning such as the example in Table A.2. In the latter sentence type, the verb(s), the initial NP and either a verb particle and/or an additional adverbial phrase as a whole functions as the predicate of the sentence at hand (Teleman et al. 1999 (3): 311), for example, the construction *ha råd med* in the example in Table A.2. The NPs in such constructions in most cases consist of an uninflected single word noun (i.e., such as *råd* in the example).

Sentences with *light constituents*, finally, were also excluded. These contain traditional light or function verb constructions, in which it is the NP rather than the verb(s) that provides the primary semantics of the predicate (Teleman et al. 1999 (3): 262). They also include sentences with "light object" constructions, containing a semantically "rich" main verb, but in which the syntactic direct object is semantically impoverished and has to be considered either part of the predicate (Teleman et al. 1999 (3): 179) or to function as an adverbial rather than a direct object proper (see examples in Table A.2) (see Teleman et al. 1999 (3): 222, see also Teleman et al. 1999 (2): 682). It was in many cases hard to draw a clear line between lexicalized predicate constructions with a semantically weak NP, on the one hand, and light object constructions, on the other.

All exclusions were done with reference to Teleman et al. (1999) as well as on the basis of my intuitions as a native speaker. The identification of excluded sentences was done automatically to the extent that it was possible. For example, subject predicatives were identified on the basis of the sentence head verbs. It was however in many cases impossible to automatically identify all sentences of a specific exclusion type. Light constituent constructions, for instance, were in most cases identified by searching for verbs and nouns that commonly are used in such constructions (e.g., the verb *ta* and the noun *beslut*), but this obviously does not entail that all light constituent constructions in the data are identified. Since I never conducted a complete manual search of the full data set, the final data still contains a small number of false inclusions. A random sample of 166 cases, constituting approximately 1% of the total data, was shown to contain 15 false cases. This indicates that only about 8.5% of the total data consist of false inclusions.

References

Brolin, Charlotte. 2006. Perceptionsverb+ objekt med infinitiv [Verbs of perception+ object with infinitive]. Department of Swedish Language, University of Gothenburg, Gothenburg, Sweden. Unpublished work.

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.

Monica-Alexandrina Irimia. 2005. Types of secondary predication. *Toronto Working Papers in Linguistics* 25. 20–29.

Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. Svenska Akademiens grammatik [The grammar of the Swedish Academy]. 1st ed. Stockholm: Svenska Akademien.