When input fails to determine the grammar: on mid-distance binding in Swedish

Björn Lundquist

1. Introduction

The results from the Scandinavian dialect syntax survey (ScanDiaSyn) carried out in Scandinavia between 2005 and 2010 show that there is a lot of variation within and between the Scandinavian variants (see Lindstad et al. 2009 for information about the project and the available database). The variation is often determined by region, but not always. We often see that within the same measure point, the informants have different intuitions about the grammaticality of one and the same sentence, and in many of these cases, the variation does not seem to be determined by other demographic factors like age, gender or even social status. When we find linguistic variation, we seek the source of the variation in external factors (mainly input), and we call it dialectal or sociolectal variation. In doing so, we presuppose that a grammar is fully determined by input/primary linguistic data (in combination with Universal Grammar and socio-cultural factors). When we find variation that is not conditioned by region or age, we are likely to view it as "noise" in the sample, or attribute it to additional hitherto unknown socio-linguistic factors. The purpose of this article is to challenge the view that linguistic variation always has its source in external factors (i.e., variation in input and socio-cultural factors). The question that I will try to answer in this article is the following: Can more than one grammar arise from the same input in the same socio-cultural context? Or put slightly differently: can language learner A and language learner B arrive at two different grammars from the same input? The answer to these questions is, I believe, Yes, but of course we cannot be 100 per cent certain, since we don't know of any cases where two language learners are exposed to *exactly* the same input. I will argue that choice of learning strategy, or an individual bias for a certain type of linguistic rule system, can influence the grammar you end up with. Your choice of learning strategy might of course be influenced by external factors, though presumably not only linguistic input. I will in the concluding section of this paper also suggest that we cannot rule out that a certain bias for a learning strategy is partly genetically determined, following ideas in Ladd et al. (2008).

The case study of this paper is variation in so-called mid-distance binding in Swedish, i.e. binding into an infinitival clause by a main clause subject. In the literature on Swedish binding, mid-distance binding of a simplex reflexive is in general taken to be possible, as in (1-a) and (1-b) (examples from Platzack 1998):

(1) a. *Eva$_i$ bad honom$_j$ att PRO$_j$ kamma sig$_{i/j}$.*
Eva asked him to PRO comb RFLX
'Eva asked him to comb her/himself.'

b. *Eva$_i$ bad honom$_j$ att PRO$_j$ kyssa sig$_i$.*
Eva asked him to PRO kiss RFLX
'Eva asked him to kiss her.'

However, many speakers, including myself, strongly reject binding by the matrix subject in (1) (which makes (1-b) ungrammatical, since *kyssa* would take a complex reflexive when locally bound). However, there are speakers who accept mid-distance binding in both examples above. The acceptability of mid-distance binding was investigated in the Swedish part of the ScanDiaSyn-project. In the ScanDiaSyn project, four informants (old man, old woman, young man and young woman) from a total of over 200 measure points in Scandinavia were asked to give grammaticality judgments for 100-150 sentences. The sentences were judged using a five graded scale (1-5, 1 bad, 5 good). The informants heard recorded versions of the sentences, read in by a native speaker of the dialect. At the time of writing, data is available from 26 measure points in Sweden and Finland. The following sentence was tested in Sweden and Finland:

(2) *Hon$_i$ bad mig hjälpa sig$_i$.*
she asked me help RFLX
'She asked me to help her.'

The sentence was judged acceptable (4-5) for 21 of 102 (20.5 %) informants. The mean value was 2.3. This variation is not mainly determined by region, i.e. the 21 speakers accepting (2) are spread all over the country. It should be noted however that the sentence in general received higher scores from the older population, which might indicate that the this type of mid-distance binding is about to disappear in Sweden. However, in several locations the sentence gets high scores from one or two of the younger informants, sometimes even when the sentence gets a low score from the older informants.[1] In addition to mid-distance binding of simplex reflexives, mid-distance binding of possessive reflexives was tested as well:

(3) *Hon$_i$ bad mig passa sin$_i$ katt.*
she asked me watch RFLX.POSS cat
'She asked me to look after her cat.'

This sentence was found acceptable for **71.5**% of the informants, and the mean value it got was 4. The variation in judgments was not determined by region, age or gender. If we take the results from (2) and (3) at face value, we seem to have two or maybe three different mid-distance binding grammars in Sweden. I will in this talk focus on what I call Grammar A and Grammar B:

Grammar A: Rejecting binding of simple reflexive (2) but allowing binding of possessive reflexive (3) in infinitives by matrix subject. 55% of the informants.

Grammar B: Allowing binding of both simple reflexive (2) and possessive reflexive (3) in infinitives by matrix subject. 17% of the informants.

26% reject both (2) and (3). However, as was found out during field work in Norway (where only (2) was tested), some informants change their judgments from bad to good once an infinitival marker is present. This could be true for some Swedish informants as well, and in

[1] It should also be noted that the sentence gets higher score from the measure points in Finland compared to the measure points in Finland, indicating that region *is* a factor that determines the variation as well. I will return to the question why mid-distance binding is particularly bad in Sweden compared to its Scandinavian neighbors in section 4.

that case, it is possible that all the informants allow some mid-distance reflexives. However, we cannot rule out the existence of a Grammar C, where binding over an infinitival clause boundary is illicit. I will however not have anything to say about a possible Grammar C in this paper. 3 informants gave higher scores for (2) than (3), but no informant gave (2) a full score, while judging (3) as completely ungrammatical (1-2). I will for now simply assume that the logically possible Grammar D, where mid-distance binding of simplex reflexives but not possessive reflexives is licit, does not exist in Sweden.

The difference between Grammar A and Grammar B can not straightforwardly be analyzed as, say, a difference in size of binding domain. Rather, there seems to be a difference in binding properties between simplex reflexives and possessive reflexives. As we will see below however, mid-distance binding of simplex reflexives is possible in Grammar A as well, but only when the simplex reflexive does not immediately follow a verb. It should also be mentioned that the possessive reflexive *is* in general a local reflexive in Swedish, as can be seen from the fact that binding of *sin* in a finite complement clause is not possible. The following sentence was included in the ScanDiaSyn survey to test this:

(4) *Grannen$_i$ ville att vi skulle passa sin$_i$ katt.*
Neighbor wanted that I should watch RFLX.POSS cat
'The neighbor wanted that we should look after her cat'/'She wanted me to look after her cat.'

This sentence was rejected by basically all informants (4 speakers gave it 4, no 5 and the mean value was 1.4).

The question that I will try to answer in this paper is how Grammar A and Grammar B can arise in the same speech community. I will argue that in Grammar A, a simplex reflexive can only be interpreted as a de-transitivizing element when following a verb, while it can be interpreted as an argument in Grammar B in the same context. I will show how both Grammar A and B could logically arise from the same core input, and the grammar you end up with is determined by how the language learner analyses the input data. Grammar B will arise when the speaker analyses *sig* as the third person member of the object pronoun series. Grammar A will arise in the speakers that posit a special *sig*-rule (i.e., *sig* in the complement of a verb can only be interpreted as a de-transitivizing element).

I will also briefly discuss Scandinavian variants where all speakers seem to end up with a grammar where the simplex reflexive and the possessive reflexive don't differ in their binding properties, most notably Icelandic and Danish, and why a grammar like Grammar A is less likely to arise there.

The structure of the paper is the following: in section 2 I will discuss the notion of idiolectal variation, and also look at what type of linguistic phenomena are more likely to give rise to inter-speaker variation rather than inter-region variation. In section 3 I will go through the basics of the binding system in Swedish and show exactly how Grammar A and Grammar B differ from each other. I will also show how two different learning routes get you to Grammar A or Grammar B. In section 4 I will briefly discuss possible reasons for Grammar A in Danish and Icelandic. Section 5 concludes the paper.

2. Types of variation

As mentioned above, the grammaticality of around 100-150 sentences was tested in around 200 locations in mainland and insular Scandinavia. In total over 800 informants were consulted. Below I will focus on the result from Norway and Sweden. The acceptability of many of the sentences varied depending on the region or country they were tested in. For example, sentences that tested for particle placement showed a variation that was almost completely determined by the country borders: the test sentence with a verb-particle following a direct object was rejected in Sweden, while accepted in Norway. Similarly, past tense *-s*-passive sentences were rejected all over Norway, but accepted all over Sweden. Sentences that tested for the non-V2 word order in subject *wh*-questions were only found acceptable in certain regions of Norway (mainly northern Norway, Trøndelag, northern parts of Western Norway and Rogaland). Similarly, the absence of a *that*-trace effect was only found in a small region of Norway (Southeastern Norway).

However, many phenomena tested show variation that is not determined by region. Many sentences were accepted by only one or two informants at a measure point, but rejected by the other informants at the same measure point. It is an interesting question in itself what type of phenomena show variation that is determined by region, and what type of phenomena have variation that is not, but it is outside the scope of this article to fully investigate this issue. It is clear that factors like frequency, complexity of the grammatical rule, variation in the input and age of acquisition (which is presumably influenced or determined by aforementioned factors) influence what type of variation a certain phenomenon will show. Looking at the ScanDiaSyn results, it is striking that two phenomena show a high degree of non-regionally determined variation, or inter-speaker variation: non-local anaphora, and root phenomena in different types of embedded clauses (mainly the availability of V2 and topicalization in embedded clauses). These two phenomena have a number of shared characteristics. Most notably, they have alternative realizations that are less pragmatically marked. For non-local anaphors, a personal pronoun or a generic pronoun (*en*) can be used in most contexts where non-bound or long-distance anaphora are found. Choice of anaphora/pronoun is often determined by extremely subtle semantic or pragmatic factors, like point of view/logophoricity and topicality. When it comes to embedded V2, the same semantic and pragmatic force present in V2 subordinate clauses can be obtained in non-V2 subordinate clauses by adjusting the stress.

One could in principle argue that the type of inter-speaker variation discussed in this paper does not really exist. For example, one could argue that some informants are better at filling in the pragmatic information that is necessary to license e.g. non-local anaphora and embedded V2, or that some speakers judge certain sentences as ungrammatical when there is a less marked alternative available. However, even when the informants get help with setting up the right pragmatic context for a sentence, the difference in judgment between informants most often persists. Further, we know that even trained linguists disagree on some judgments, even when they are speakers of the same dialect. For the purpose of this article, I will assume that the inter-individual variation found in the ScanDiaSyn results is real, and that the different judgments given by different informants correspond to differences in the internal grammars of the informants, both for variation that is determined by region/place (i.e. dialectal/inter-regional variation) and that is not determined by region/place (i.e. inter-speaker variation). Below I will look at a phenomenon that show inter-speaker variation in Swedish, mid-distance binding, and argue that this variation can arise due to differences in individual biases for a certain type of rule systems.

3. Case study: mid-distance anaphora in Swedish

In this section I will go through the differences and commonalities between Grammar A and Grammar B, and give a plausible acquisitionist account of how the two grammars could arise from the same input. To remind the reader, Grammar A is the grammar that allows (5), but not (6), while Grammar B allows both (5) and (6).

(5) a. Binding of simple anaphor (*sig*) in the complement of embedded infinitive by matrix subject:
b. *Hon$_i$ bad mig hjälpa sig$_i$.*
she asked me help RFLX
'She asked me to help her.'

(6) a. Binding of possessive anaphor (*sin*) in the complement of embedded infinitive by matrix subject:
b. *Hon$_i$ bad mig passa sin$_i$ katt*
she asked me to watch RFLX.POSS$_i$ cat
'She asked me to look after her cat'

As already mentioned in the introduction, 55 % of the informants in the Swedish part of the ScanDyaSyn survey have Grammar A, while 17 % have Grammar B. We have further reasons to expect that Grammar A and B cover a larger part of the population than 72 % (55 + 17), since some speakers possibly require an overt infinitival marker in control infinitives. Below I will look at the shared properties of Grammar A and B. It should be noted that examples and judgments below do not originate from the ScanDiaSyn survey, but reflect my own intuitions and the intuitions from a handful of native informants.

3.1. Shared properties of Grammar A and Grammar B

It is highly likely that Grammar A and Grammar B only differ with respect to binding of simplex mid-distance reflexives. I will below look at three points where the two grammars converge:

1. Binding of both possessives and simplex reflexives in infinitival clauses by matrix subject is possible:

   (7) Hon$_i$ bad mig passa sin katt$_i$.
   she asked me to watch RFLX.POSS$_i$ cat
   'She asked me to look after her cat.'

   All speakers that I have consulted (of both Grammar A and B) also allow mid-distance binding of *sig* only when *sig* is in the complement of a preposition:

   (8) *Hon$_i$ bad mig$_j$ PRO$_j$ stanna hos sig$_i$/henne$_{i,k}$ över natten.*
   she asked me PRO stay at RFLX/her over night.DEF
   'She asked me to stay with her overnight.'

2. *Sig* can be used as a "de-transitivizing" element (as a "lexical" reflexive, see e.g. Reinhart and Reuland 1993, see also Kayne 1975, Alsina 1996 and Medová 2007 for other

accounts). Here *sig* presumably does not have the status of an argument, since these predicates behave like intransitive predicates: only mono-argumental predicates can appear in existential clauses such as (9):

(9) a. *Det öppnade sig en dörr framför honom.*
it opened RFLX a door front.for him
'A door opened in front him.'
b. *Det satte sig en man på den andra stolen.*
it sat RFLX a man on the other chair
'A man sat down on the other chair.'
c. *Det tvättade sig några män nere vid stranden.*
it washed RFLX some men down at beach.DEF
'Some men washed (themselves) down at the beach.'

3. *Sig* can be used as an argument, which most clearly can be seen in PP-complements where *sig* easily can be conjoined with a NP:

(10) a. *Hon ordnade ett möte hemma hos sig och sin sambo.*
she arranged a meeting home at RFLX and RFLX.POSS partner
'She arranged a meeting at her and her partner's place.'
b. *Hon lade sonen i sängen mellan sig och sin make.*
she put son.DEF in bedDEF between RFLX and RFLX.POSS husband
'She put the son in the bed between herself and her husband.'

Further, PP's do not become "intransitive" when they have reflexive complement:

(11) *Han ställde glaset framför sig.*
he put glass.DEF front.for sig
'He put the glass in front of himself.'

In (11), we see a transitive relation between a figure (*glaset*) and a ground (*sig*)

Since *sig* can be an argument, and mid-distance binding is OK for speakers of Grammars A and B, we expect both grammars to accept mid-distance binding of simplex reflexives in the complement of verbs. However, as we have seen, this is rejected by speakers with Grammar A. The three points above clearly show that the difference in between Grammar A and Grammar B cannot be stated in terms of "size of binding domains", since both grammars allow mid-distance binding. Neither can it be described as a difference in binding properties between the simplex reflexive and the possessive reflexive, since both grammars allow mid-distance binding of both *sig* and *sin*. Rather, mid-distance binding of *sig* is only ruled out when *sig* directly follows a verb. I will claim that the speakers of Grammar A allow only a de-transitivizing *sig* in the complement of a verb, while the speakers of Grammar B also allow an argumental *sig* following a verb. The two grammars are summarized in the table below:

	A	B
1. MID-DISTANCE BINDING	✓	✓
2. DE-TRANSITIVIZING *sig*	✓	✓
3. PRONOMINAL / ARGUMENTAL *sig*	✓	✓
4. *ARGUMENT *sig* AFTER VERB	✓	

Grammar B is simpler than Grammar A, at least in that in contains fewer rules, but still Grammar A is more common than Grammar B. Why do most speakers add the fourth rule ("*argument *sig* after verb")? It should be a marked option, since no other element can be an argument of P but not V, as far I'm aware.[2] Below I will give a tentative explanation to why most Swedish speakers build up a grammar that does not generate a sentence like (2).

3.2. Arriving at Grammar A: The sig-specific rule

It is well known that a language often has more than one way of marking reflexivity (see Haspelmath 2008 for an overview). In many of the Germanic languages (and other languages as well) we find a difference between complex and simplex reflexive pronouns, where the complex reflexives consist of the simplex reflexives and a *self*-element (see e.g. Hellan 1988 and Reinhart and Reuland 1993 for discussion). Complex reflexives can normally only surface when the antecedent is local, while simplex reflexives sometimes can be found in long- and mid-distance contexts. Many verbs can only take a complex reflexive when the internal and external argument are co-referent (12-a), while other verbs can (or sometimes have to) take a simplex reflexive (12-b):[3]

(12) a. *Han älskar sig *(själv).*
he loves RFLX self
'He loves himself.'
b. *Han tvättade sig (själv).*
He washed RFLX (self)
'He washed (himself).'

As was seen in (9) above, the simplex reflexive can be used as a detransitivizing element. The complex reflexive can however not be used this way. We will thus not find any complex reflexives in existential clauses:

[2]It should however be noted that we don't have a any good theory around for determining what counts as "a simpler Grammar".

[3]It is not obvious that the optional *self*-part in the reflexive following a verb like *tvätta* ('wash') should be analyzed on par with the *self*-part in the reflexive in the complement of *älska* ('love'). *Sig själv* in the complement of *älska* is clearly a constituent, as can be seen in the fact that you cannot split *sig* and *själv*, while this is possible for *tvätta*:

(i) a. **Han älskade sig inte själv.*
He loved RFLX not self
int. 'He didn't love himself.'
b. *Han tvättade sig inte själv.*
he washed RFLX not self
'He didn't wash himself.'

It is possible that *self* following a verb like *tvätta* always is an emphatic marker, though this I will in this article assume that *sig själv* following a verb like *tvätta* at least sometimes is a true complex reflexive.

(13) a. *Det tvättade sig (*själva) några män nere vid stranden.*
it washed RFLX SELF.PL some men down at beach.DEF
'Some men washed (themselves) down at the beach.'
b. **Det älskar sig själva många män nuförtiden*
it loves RFLX SELF many men nowadays
Int. 'Many men love themselves nowadays.'

The simplex reflexive has a different distribution compared to regular complex reflexives, pronouns and other nominal elements. Most notably this can be seen in verb-particle constructions, where simplex reflexives sometimes have to precede a verb-particle, as shown in (14) and (15) below:

(14) a. *Han kastade sig ner på golvet.*
he threw RFLX down on floor.DEF
'He threw himself down on the floor.'
b. *Han kastade ner honom på golvet.*
he threw down him on floor.DEF
'He threw him down on the floor.'

(15) a. *Han la sig ner i sängen.*
he lay.PAST RFLX down in bed.DEF
'He laid down in the bed.'
b. *Han la ner henne i sängen.*
he lay.PAST down her in bed.DEF
'He laid her down in the bed.'

My proposal is simply that speakers of Grammar A always treat the simplex reflexive in the complement of a verb as a de-transitivizing, verbal element rather than a direct object. It is admittedly hard to prove this point. First, not all verbs with simplex reflexives can occur in existential constructions.[4] However, this is true for regular intransitive verbs as well. In general, verbs that denote actions/events that are easily observable, or that have results that are easily observable, are more suitable in existential constructions than verbs that denote "internal " events, like thinking and feeling. As shown below, verbs of cognition/psych-verbs are odd in existential construction, both when they are reflexive marked or unmarked:

(16) a. ??*Det oroade sig många efter katastrofen.*
it worried RFLX many efter catastrophe.DEF
Int. 'Many people worried after the catastrophe.'
b. ??*Det sörjde många efter katastrofen.*
it mourn.DEF many after catastrophe.DEF
Int. 'Many people mourned after the catastrophe.'

Secondly, not all simplex reflexives can precede verb particles, and for some verb-particle construction, the reflexive can either precede or follow the particle, giving rise to different interpretations:

[4]Some speakers even find an existential construction with *tvätta sig* ('wash') marked. Speakers simply seem to differ in to what extent they allow clearly agentive verbs to occur in existential constructions.

(17) a. *Han tog ut sig.*
he took out RFLX
'He exhausted himself.'
b. *Han tog sig ut.*
he took RFLX out
'He got (himself) out.'

However, as shown in (18), a verb with a simplex reflexive following a particle can still occur in an existential construction, which indicates that a verb phrase of the shape verb-particle-reflexive still can be interpreted as an intransitive predicate. Note that the complex reflexive is not licit here (even though this verb allows both a simplex and a complex reflexive in its complement).

(18) *Det har låst in sig (*själv) någon på toaletten.*
it has locked in RFLX (self) someone on bathroom
'Someone has locked himself in the bathroom.'

The language learner will thus have plenty of evidence that the simplex reflexive does not have the status of a direct object, and that the simplex reflexive is different from both pronouns and complex reflexives, both from existential clauses and particle constructions. There are other pieces of evidence for this as well. First, simplex reflexives cannot be topicalized, at least not for speakers of Grammar A, as opposed to complex reflexives and pronouns (as will be mentioned below, there seems to be a split within Grammar A with respect to the acceptability of topicalized *sig*):

(19) *Honom/sig själv/*sig skulle han däremot aldrig låsa in.*
Him/RFLX self/RFLX would he on.other.hand never lock in
'Him/himself, he would on the other hand never lock up.'

Secondly, simplex reflexives can undergo what is sometimes called "long object shift" that is, it can precede the subject. This is only (if ever) allowed for pronouns under very special contexts (and complex reflexives do not undergo object shift at all):

(20) *Därefter tvättade sig/*henne mannen noggrant.*
there.after washed RFLX/her man.DEF carefully
'Afterwards, the man carefully washed himself/her.'

The only type of input that would clearly show that a simplex reflexive following a verb was a true argument would come from mid-distance (and long-distance) binding. A mid-distance reflexive clearly could not be a de-transitivizing element, since the reflexive marked predicate in this context clearly is not intransitive. The embedded verb-phrase in (21) is clearly transitive, and so is the verb-phrase in (22) when the reflexive is not interpreted as a local reflexive:

(21) a. *Eva$_i$ bad honom$_j$ att PRO$_j$ kyssa sig$_i$.*
Eva asked him to PRO kiss RFLX
'Eva asked him to kiss her.'
b. *Eva$_i$ bad honom$_j$ att PRO$_j$ kamma sig$_{i/j}$.*
Eva asked him to PRO comb RFLX

'Eva asked him to comb her/himself.'

However, this type of mid-distance reflexive is presumably very rare. Even if the language learner gets plenty of input from both speakers of Grammar A and B, it is not obvious that sentences like the ones in (21) will ever appear in the input. As has been shown by Jakubowicz 1994 for Danish, adult speakers who always accept and correctly interpret sentences with mid-distance reflexives, almost never produce them, but choose to use a non-reflexive pronoun or a full NP instead of the simplex reflexive. As I will argue in the next subsection, a language learner may arrive at Grammar B even in the absence of sentences like (21-a) and (21-b) in the input. I will hypothesize that the language learner that arrives at Grammar A at some point posits the following rule, based on the input:

(22) The *sig*-specific rule: a simplex reflexive in the complement of a verb is a de-transitivizing element.

Since there is plenty of evidence in the input that *sig* is a de-transitivizing element when following a verb, and possibly no evidence that it is not, the *sig*-specific rule should arise naturally. One important trigger for this rule is presumably the opposition between *sig* and *sig själv*, where *sig själv* is clearly an argument.[5]

As has been shown by Jakubowicz (1994), Danish children learn very early the simplex–complex distinction. At 3.6 years, children seem to know that some verbs take simple reflexives while others take complex reflexives. It is presumably safe to assume that Swedish language learners acquire this difference early as well. Let us assume that children simply learn that some verbs require a full reflexive argument (i.e., *sig själv*) in order to get a reflexive interpretation, while it is possible to just insert an element that alters the thematic structure (i.e., *sig*) for other verbs. I will not here discuss why certain verbs require full reflexive argument, while other suffice with a valency changing *sig*, or what the exact structure of verbs with simplex reflexives is.

However, the size of the binding domain is something that the children learn extremely late. More specifically, children take a long time to acquire non-local anaphora, as has been shown by Jakubowicz (1994) for mid-distance binding in Danish, and Sigurjónsdóttir and Hyams (1992) for long-distance binding in Icelandic. in comprehension studies carried out by Jakubowicz (1994), children of the age 7-8 years only managed to interpret mid-distance anaphors correctly at 35-59 % per cent of the time (depending on experiment). Not until the age of 10 have the children approached a target like grammar, getting it right in 70-78 % of the times). Let us assume that mid-distance binding is acquired equally late in Swedish, i.e., mid-distance binding of a possessive reflexive (Grammar A and B), a reflexive in the complement of a preposition (Grammar A and B) and a reflexive following a verb (Grammar A) is acquired very late, much later than the so-called *sig*-specific rule is learned (for Grammar A).

[5] A separate question is why complex anaphors in general do not allow mid-distance binding. Since mid-distance binding is allowed in Swedish, and *sig själv* always is an argument, we will expect mid-distance binding of *sig själv*. We will just have to state that *själv* has different binding properties than *sig* and *sin*. It could be related to the adjectival status of *själv*, or simply the fact that *själv* shows Gender and Number agreement with its antecedent, just like other adjectives. As far as I am aware, there are no long- or mid-distance adjectives in Swedish. There is always a strictly local relation between the an adjective and its subject.

In other words, you first learn the rule (1) that simple reflexives have a de-transitivizing function when they follow a verb, (and in these cases, co-reference with nearest subject is always required). When you later learn the right size of the binding domain (including control/raising infinitive), rule 1 is already deeply entrenched in your grammar. You do not "unlearn" the rule that *sig* in the complement of a verb is a de-transitivizing element. A simplex reflexive in a control infinitive can thus only be interpreted as a de-transitivizing element, even though it is within the binding domain of the matrix subject.

The question that remains to be answered is thus why not all speakers end up with Grammar A, and also why speakers of both Danish and Icelandic end up with a grammar that looks more like Grammar B. There are at least two possible answers to this question: either, (1) speakers of Grammar B never posit a *sig*-specific rule, or (2) they posit a *sig*-specific rule which they later override. I will argue that the first answer is the correct one, and show why a language learner might have good reasons not to posit a *sig*-specific rule.

3.3. Arriving at Grammar B: sig as an object pronoun

Instead of positing a *sig*-specific rule, as the speakers of Grammar A, I will argue below that some speakers arrive at Grammar B by positing a more general rule, applying to a whole series of object pronoun, namely the following:

(23) A locally bound simplex object pronoun can only have a valency changing function when following a verb.[6]

The learner should presumably notice that *sig* is a part of the paradigm of object pronouns, including first person singular (*mig* 'me') and plural (*oss* 'us') pronouns, and second person singular (*dig* 'you'), and plurar *er* 'you') (and excluding *honom* 'him', *henne* 'her', *den/det* 'it' and *dem* 'them'). I will call this class of object pronoun "bound object pronouns", since they either get there reference determined by a clause-mate NP or pronoun, or a speech act participant (speaker or hearer). They all have special possessive forms, that unlike NP's and other pronoun agree with the head noun in gender and number, like modifying adjectives, and lack the typical possessive *-s*:

(24) a. $\text{min}_{1st,sg,common}$ – $\text{mitt}_{1st,sg,neuter}$ – $\text{mina}_{1st,plural}$
b. $\text{sin}_{3rd,sg,common}$ – $\text{sitt}_{3rd,sg,neuter}$ – $\text{sina}_{3rd,plural}$

The first and second person object pronouns, just like third person *sig* have a de-transitivizing function when they are locally bound, as can be seen in (25), where a first person singular object pronoun precedes a verb-particle:

(25) *Jag kastade mig ner på golvet.*
he threw RFLX down on floor.DEF
'I threw myself down on the floor.'

[6]There might be speakers of Grammar A who have a different version of this rule, stating only that locally bound simplex object pronouns *can* have a valency changing function, but still allowing a locally bound simplex pronoun to be interpreted as a true argument when following a verb. These speakers are predicated to allow e.g. topicalization of locally bound simplex object pronouns. For these speakers, the difference between post-verbal *sig* and *sig själv* must be of a different nature than for the other speakers. See Kiparsky (2002) for alternative ways of stating the difference between simplex and complex reflexives (and note that nothing in my account explains the choice of reflexive in other contexts than after a verb, for example after prepositions.

The first and second person object pronouns can however not appear in existential constructions, since their antecedents necessarily are definite (it could only be a first or second person subject pronoun), and there is a definiteness restriction on the subject in existential constructions. It is however obvious that simplex first and second person pronouns create predicates that are clearly intransitive in meaning, like (26):

(26) *Jag ändrade mig – han ändrade sig*
I changed me – he changed RFLX
'I changed (my mind) – he changed (his mind)'

Just like *sig*, the first and second person bound pronouns have complex forms, that have to be used in cases of co-reference between subject and object:

(27) *Jag älskar mig ??(själv)*
I love me (self)
'I love myself'

The simplex locally bound first and second person pronouns cannot really be topicalized, and instead the complex version has to be used:[7]

(28) *Mig själv/??mig skulle jag däremot aldrig låsa in.*
My self/me would he on.other.hand never lock in
'Me/myself, I would on the other hand never lock up'

In short, the first and second person bound object pronouns have the same quirks as *sig*.[8], [9] However, as soon as the first and second person bound object pronouns are not bound by the closest subject, they have a distribution similar to the non-bound pronouns *honom* ('him')

[7]The simplex form is alright in a context where the subject and the anaphor clearly are not co-referent, as in the following context: 'If I were you, I might lock up John, but ME I would on the other hand not lock up'. Note that this effect is not seen if the simplex first person object pronoun is not topicalized, or when the complex first person is topicalized. As mentioned above, there might be a sub-group of Grammar A speakers who allow argumental locally bound simplex object pronouns, for whom topicalization would be felicitous without a forced disjoint interpretation.

[8]There are some differences that are worth pointing out. First, the third person *sig*lacks number marking, as opposed to the other pronouns. Even though linguists have emphasized the importance of the underspecification of third person reflexives, I am not convinced that the singular-plural syncretism makes *sig* different from other pronouns. Syncretism in number and case in pronoun paradigms is extremely common, and I doubt that the number syncretism of *sig* is any more special than the number (and case) syncretism of English *you*, for example. Another difference is that we never see long object shift of a co-referent first or second locally bound pronoun, but this is only due to the fact that the antecedent of a first or second person locally bound object pronoun is a pronoun, and we never see long object shift over a pronoun, not even in the third person.

[9]Another obvious similarity between first and second and person bound object pronouns and *sig* can be interpreted as bound variables, i.e. you can get sloppy readings under ellipsis. This is true for both simplex and complex anaphors, and I think that a sloppy interpretation is obligatory for all simplex anaphors bound by the nearest subject, and also complex anaphors in most contexts:

(i) a. *Jag ändrade mig. – Det gjorde jag också.*
I changed me – It did I too
'I changed my mind – So did I'
b. *Jag hatar mig själv – det gör jag också.*
I hate me self – it do I too
'I hate myself – so do I'

or *henne* ('her'). Most importantly, they can of course appear in the complement of a verb in their simple forms, without changing the valency of the predicate:

(29) a. *Hon älskar dig (*själv).*
she love.PRES you (self)
'She loves you(*rself)'
b. *Hon tvättade mig (*själv)*
she wash.PAST me (self)
'She washed me/(*myself)'

Further, non-locally bound first and second object pronouns have the same distribution as regular pronominal arguments. They follow verb-particles (30-a) and can be topicalized (30-b):

(30) a. *Han slängde ner mig på golvet*
he throw.PAST down me on floor.DEF
'He threw me down on the floow'
b. *Mig skulle han aldrig låsa in*
me would he never lock up
'He would never lock me up.'

The language learner simply needs to learn that the bound pronouns have different distribution when they are bound by the nearest subject compared to when it is bound by either a speech act participant or a non-local antecedent. This is of course true for both speakers of Grammar A and Grammar B. However, if we assume that the learners that end up with Grammar B only focus on the shared characteristics of all the members of the paradigm, they would never postulate a special rule for third person reflexive *sig*. *Sig* in the complement of a verb, just like all the other pronouns have a special distribution when bound by the local subject. The difference between *sig* and the first and second person bound pronouns is that *sig* will basically always be bound by the nearest subject when following a verb (and in other cases too), while first and second person bound pronoun very often have are "free" (or only

---

Sloppy readings are expected either if the object pronoun is a bound variable or if it is a valency changing verbal element. There are some contexts where complex anaphors seem to allow strict interpretations, probably only in contexts where the missing VP is not replaced by a pronoun (*det* 'it'). Below I give an example with a first person subject, but the effect would be the same with a third person subject.

(ii) *Jag tvättar mig (själv) noggrannare än vad du gör.*
I wash me (self) careful.COMP than what you do
'I wash myself more careful than what you do.'

The strict reading is only available when *själv* is present.

When a first or second person object pronoun is bound by a possessor inside a subject, both strict and sloppy readings are available, as in (iii-b):

(iii) a. A: Min bror hjälper mig med allting
Speaker A: my brother helps me with everything
b. Det gör min bror också
Speaker B: So does my brother.

In short we can conclude that a 'sloppy' reading is only available when the anaphor is a de-transitivizer, whereas a strict reading can be obtained in other contexts.

bound by the speaker or hearer).

As soon as the learner of Grammar B knows that the binding domain includes control and ECM infinitives, third person *sig* in the complement of an embedded infinitival verb will be generated/accepted by their grammar. Even if the sentences of the type (2) are absent in the input, they will still be as natural as a sentence with a mid-distance bound first or second person object pronoun:[10]

(31) *Jag bad honom hjälpa mig.*
I asked him help.INF me
'I asked him to help me.'

Summarizing Grammar B: the learner will assume that *sig* has the same properties as the first and second person members of the "bound" object pronouns. That is, when the bound object pronoun directly follows the verb and is co-referent with the nearest subject, it is only interpreted as a de-transitivizing element, though when it is by something else (i.e, a speech act participant, or a non-local subject), it behaves like an argument. *Sig* can of course not be bound by a speech act participant, since a third person referent is not a speaker or a hearer, so *sig* will basically only surface when it is bound by the local subject, and possibly by a non-local subject in control and ECM contexts, though this may not be in the input for the learner.

The best way to capture the difference between learners that end up with Grammar A and learners that end up with Grammar B is the following: learners of Grammar A pay attention to the fact that *sig* in contrast to the first and second person object pronoun *always* has a de-transitivizing function when following a verb. In the Grammar they end up with *sig* following a verb can only be a valency changing element. The learners of Grammar B ignore the fact that *sig* following a verb always has a de-transitivizing function. Instead they focus on the properties of the whole paradigm, and will not bother to postulate a *sig*-specific rule. The variation can thus be said to have its sole origin in the choice of learning strategy of the learner: children focusing on individual lexical items will end up with Grammar A, while children paying attention to whole paradigms end up with Grammar B.

4. A note on cross-linguistic variation

I have argued above that Grammar A and Grammar B can arise side by side in the same speech community. We thus see variation that is not determined by region (or location). However, within Scandinavia we seem to find places where a grammar looking more like

[10]If the embedded verb is a particle verb, the anaphor will follow the particle, i.e., it will have the syntax of a non-locally bound pronoun:

(i) *Jag bad honom att slänga ut mig genom fönstret*
I asked him to throw.INF out me through window.DEF
'I asked him to throw me out the window.'

For speakers of Grammar B, a mid-distance *sig* would also follow the particle. The following sentence would thus have different interpretation depending on the placement of *sig* with respect to the pronoun:

(ii) *Han$_i$ bad henne$_j$ att slänga {sig$_j$} ut {sig$_i$} genom fönstret.*
he asked her to throw RFLX out RFLX through window.DEF
'He asked her to throw him out the window' or 'He asked her to throw herself out the window'

Grammar B is the only one available, i.e. a grammar where simple reflexives following a verb can follow a verb without de-transitivizing the verb (or put slightly differently, where possessive reflexives and simple, post-verbal, reflexives both appear in Control or ECM infinitives). I will below say couple of words about why Grammar B is more pervasive in Icelandic and Danish than in Swedish.

First, in Icelandic, there seem to be no difference between the binding properties of simple reflexives following a verb and possessive reflexives.[11] There are three strong reasons why a Grammar A would not arise in Iceland:

1. The strong case-paradigm: the language learner will see that the simple third person reflexive not only has a possessive paradigm that identical to the first and second person bound object pronoun (agreeing with the head noun), but also dative and genitive forms matching the first and second person paradigm.

2. The logophoric *sig*: Binding into finite subjunctive clauses is possible in Icelandic. In these cases, the anaphor finds it antecedent in the "Author" of the sub-ordinate clause, i.e. a speech act participant. This of course strengthens the ties between *sig* and its first and second person cousins, and presumably it also gives rise to a lot of input that contains *sig* that is not co-referent with the closest subject.

3. No *sig*-specific syntax: There is no special syntax associated with "de-transitivizing" anaphors in Icelandic, which possibly may suggest that anaphors in Icelandic never have a valency changing function (as argued by Jónsson 2011). There are no real tests for transitivity in Icelandic, since transitive sentences can occur in existential clauses as well as intransitive clauses. Further, not only "de-transitivizing" anaphors but all object pronouns precede verb particles in Icelandic. Finally, as emphasized by Jónsson 2011, in verb-anaphor combinations that are semantically "intransitive", the anaphor can carry quirky case (i.e. dative or genitive), and if we assume that a verb only can assign quirky case to an argument, we have to conclude that "de-transitivizing" anaphors in Icelandic really are arguments.

From the three points above, we can conclude that the Icelandic language learner has no reasons to postulate a *sig*-specific rule. It is possible that *sig* following a verb always is a true argument in Icelandic, even for so-called inherently reflexive verbs (i.e., verbs that can only take an anaphor as its complement), as argued by Jónsson (2011), but it is outside the scope of this article to verify or refute this claim.

In Danish, all speakers seem to accept an anaphor in the complement of an infinitival to be mid-distance bound.[12] Danish does not have logophoric reflexives (with the exception of the Western Jutland dialect), and there are no strong case-paradigms either. All arguments precede verb-particles in Danish, so there is no input indicating that local anaphors have a special placement with respect to particles. *Sig* can however appear in existential clauses,

[11] Not all Icelanders accept binding into control infinities though. as reported by Sigurjónsdóttir and Hyams 1992, only 50 % of the adult informants allowed mid-distance binding, while 90 % allowed long-distance binding into a finite subjunctive clause.

[12] At least as far as I am aware, though we do not have data from this in the Nordic Dialect Database. All of the adults in the test group used in the acquisition test on binding in Danish in Jakubowicz (1994) accepted mid-distance binding of a simple reflexive in the complement of a verb.

and there *is* a strict restriction on transitive expletive constructions. We thus have to conclude that locally bound *sig* can have a valency changing function in Danish just like in Swedish. It is thus slightly surprising that Grammar A is absent in Danish (unless we believe that it is the order between the particle and *sig* in Swedish that is solely responsible for the *sig*-specific rule of Grammar A in Swedish). I will however propose that the mid-distance simplex reflexives are licit in Danish due to the structural size of Control and ECM-infinitives in Danish. More specifically, it is likely that Control and ECM infinitives lack a structural subject in Danish. The proposal is based on the fact that most speakers accept a pronominal form of a possessive inside the object of a control infinitive to be co-referent with the "PRO" subject. In the Danish part of the Survey, 42 of 55 informants found the following sentence fully acceptable (and note that the corresponding Swedish sentence would be marked or ungrammatical for, I think, all Swedish speakers):

(32) *Hun bad ham$_i$ hente hans$_i$ barn.*
she asked hum fetch.INF his child
'She asked him to pick up his child.'

The acceptance rate for (32) is similar to that of a sentence with co-refererence between an object and a pronominal possessor inside a PP, as in (33) (which is acceptable all over Scandinavia):

(33) Vi så ham$_i$ i hans$_i$ have.
'we saw him in his garden'

Let us assume that (33) does not give rise to a principle B-violation due to fact that the object is not a subject of the PP (or the clause). The absence of a Principle B violation in (32) and (33) presumably has the same explanation, i.e. the absence of a subject (in the PP in (33) and in the VP in (32)). If the "Control infinitive" in (32) really lacks a structural subject (and thus not being a control infinitive), we may have an explanation of the general acceptance of sentences like (2) in Danish. Valency changing anaphors may simply be illicit in contexts where no subject is present. As discussed in Lundquist (2011), valency changing anaphors are not licensed in certain contexts that seem to lack a structural subject, such as nominalizations and certain participial phrases in Swedish. This is presumably true in other languages as well, at least in languages where the valency changing anaphor agrees in person with the structural subject. Let us just assume that (32) and (33) have the same structure (the only difference being the label of the complement of the main verb):[13]

(34) a. (33): Hun bad ham [$_{VP}$ hente [$_{DP}$ hans barn]]
b. (34): Vi så ham [$_{PP}$ i [$_{DP}$ hans have]]

Let us assume that a valency changing anaphor is only licensed in a verb phrase containing an external argument, or, let's say, a Voice head, following Kratzer (1996). An infinitival verb without an external argument will in all relevant aspects behave like a preposition, and a an anaphor bound by the subject will be licensed in both contexts (and a valency changing anaphor will not). Thus it is important to note that Danish is not really like either Grammar

[13]We have to assume however that the complement of control verbs can be of a larger size, since sentences like (1-a) are ambiguous for Danish speakers, i.e. the anaphor can be bound by either the main clause subject or the object/embedded subject.

A or Grammar B in Swedish.[14] Binding of an anaphor in an infinitival clause in Swedish involves mid-distance binding, i.e. binding by a non-local subject. In a Danish sentence like (35) (from Vikner 1985) the anaphor is still bound by the nearest subject:

(35) *Peter$_i$ hørte Anne omtale sig$_i$*
Peter hear.PAST Anne mention.INF RFLX
'Peter$_i$ heard Anne mention him$_i$'

It is possible that mid-distance binding is not allowed at all in Danish, and that all anaphors are bound by the nearest subject. The difference between Swedish and Danish is would thus only be in the size of the infinitival complement of a Control or ECM verb: Swedish Control and ECM infinitives are always at least VoiceP's (but presumably they are even bigger), while Danish Control and ECM infinitives can be as small as VP's.[15] The question that remains to be answered is why the Swedish Control/ECM infinitives have to be structurally "big" while Danish infinitival complements can be small. It is however outside the scope of this paper to discuss this issue, but let us for now assume that there are enough cues in the Danish input that will lead the language learner to assume that infinitives can be structurally small in Danish.[16]

5. Conclusion: variation and the third factor

I have argued that Grammar A and Grammar B can both arise from the same input. Grammar A arises in speaker who are more likely to focus on quirks on individual lexical items, while Grammar B arises in speakers who focus on general patterns shared by all members of a paradigm. The variation does thus not arise from differences in the input, but by choice of learning strategy, or individual biases for certain rule systems. Following Chomsky (2005) and Yang (2010), we can isolate three factors in language acquisition and/or language design:

1. Genetic endowment, "which interprets part of the environment as linguistic experience . . . and which determines the general course of the development of the language faculty"

2. Experience, "which leads to variation, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally".

3. Principles not specific to the faculty of language: "(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural

[14] It is possible that some Danish language learners also posit also posit a *sig*-specific rule like the following: *sig* inside a VoiceP is always a de-transitivizing element.

[15] I refer the reader to Vikner (1985) for a more thorough description of the Danish binding system.

[16] One noticeable property of Swedish is that small verbal structures have some striking morphosyntactic properties that big verbal structures lack. Most notably, as argued in Lundquist (2008) verb-particles always have to incorporate/prefix to small verbal structures in Swedish, as most clearly seen in nominalizations and participle phrases. Infinitives never allow incorporated particles in Swedish, and the absence of infinitives with incorporated particles in the input may be enough for the language learner to assume that infinitives always are structurally big. In Danish however, this does not happen in general, making big verbal structures and small verbal structures look virtually identical. See also the discussion in Wurmbrand (2001) on structurally small VP's in German.

architecture and developmental constraints. . . including principles of efficient computation"

In the list above, variation is taken to solely be triggered by the second factor, i.e. Experience, which in this context is equal to linguistic input. What I have argued for above is that the variation discussed in this paper does not have its root in the second factor, but rather the third factor, that is "principles of data analysis ... principles of structural architecture and developmental constraints." More specifically, there can be variation in how different people analyze linguistic input. Of course, this variation may have its root in experience, but presumably not linguistic experience.

I would like to end the paper with opening up the idea that variation like the Grammar A/Grammar B split we see in Swedish could be partly genetically determined, or more specifically, that an individual bias for a certain rule system may be genetically determined, as has been already proposed by Ladd et al. (2008). We know that people are different: some are introvert, some are extrovert, some are good at math, some are good at remembering names, some people focus on details and some people see "the whole picture". We also know that our personality is shaped both by our environment and our genes (see e.g. Pinker (2002) for reports on studies on identical twins separated at birth and raised in different environments). It wouldn't be surprising if some language learners paid more attention to regularities of paradigms (giving rise to Grammar B), while other pay extra attention to specific words. Whether you end up being a "paradigm person" ("a chunker") or an "individual item person" ("a splitter") might very well be partly genetically determined. Given enough input though, the grammars will presumably end up the same, and only if the relevant input is scarce and/or contradictory, will more than one grammar arise. As has been argued in this article, at least in a language like Icelandic, the input is clear enough not to give rise to two different grammars (at least not for the phenomenon discussed in this paper).

References

Alsina, Alex. 1996. *The Role of Argument Structure in Grammar: Evidence from Romance*. CSLI lecture notes 64. CSLI Publications, Stanford.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36 1: 1–22.

Haspelmath, Martin. 2008. A frequentist explanation of some universals of reflexive marking. *Linguistic Discovery* 6 1: 40–63.

Hellan, Lars. 1988. *Anaphora in Norwegian and the theory of grammar*. Studies in generative grammar; 32. Foris, Dordrecht.

Jakubowicz, Celia. 1994. Redlexives in French and Danish. In *Binding, Dependencies and Learnability*, edited by Barbara Lust, Gabriella Hermon, and Jaklin Kornfilt, pp. 113–144. Lawrence Erlbaum, Hillsdale, NJ.

Jónsson, Jóhannes Gísli. 2011. Reflexive *sig* is an argument. In *Nordlyd 37: Relating to Reflexives*, edited by Tania Strahan, pp. 99–118. University of Tromsø, Tromsø.

Kayne, Richard S. 1975. *French Syntax: The Transformational Cycle*. MIT Press, Cambridge, Ma.

Kiparsky, Paul. 2002. Disjoint reference and the typology of pronouns. In *More than Words*, edited by Ingrid Kaufmann and Barbara Stiebels, Studia Grammatica 53, pp. 176–226. Akademie Verlag, Berlin.

Kratzer, Angelika. 1996. Severing the external argument from the verb. In *Phrase Structure and the Lexicon*, edited by Johann Rooryck and Laurie Zaring, pp. 109–137. Kluwer, Dordrecht.

Ladd, D. Robert, Dan Dediu, and Anna R. Kinsella. 2008. Languages and Genes: Reflections in Biolinguistics and the Nature-Nurture Question. *Biolinguistics* 2 1: 114–126.

Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen, and Øystein Alexander Vangsnes. 2009. The nordic dialect database: Mapping microsyntactic variation in the scandinavian languages. In *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*, edited by Kristiina Jokinen and Eckhard Bick, pp. 283–286. NEALT Proceedings Series Volume 4, Odense.

Lundquist, Björn. 2008. *Nominalizations and Participles in Swedish*. Ph.D. thesis, University of Tromsø.

Lundquist, Björn. 2011. Restrictions on reflexive and anti-causative readings in nominalizations and participles. In *Nordlyd 37: Relating to Reflexives*, edited by Tania Strahan. University of Tromsø, Tromsø.

Medová, Lucie. 2007. *Reflexive clitics in the Slavic and Romance languages: A comparative view from an antipassive perspective*. Ph.D. thesis, Lund University, Lund.

Pinker, Steven. 2002. *The blank slate*. Penguin Books, London.

Platzack, Christer. 1998. *Svenskans inre grammatik – det minimalistiska programmet*. Studentlitteratur, Lund.

Reinhart, Tanya and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24 4: 657–720.

Sigurjónsdóttir, Sigriður and Nina Hyams. 1992. Reflexivization and logophoricity: Evidence from the acquisition of Icelandic. *Language Acquisition* 2 4: 359–413.

Vikner, Sten. 1985. Parameters of binder and of binding category in danish. *Working Papers in Scandinavian Syntax* 23: 1–61.

Wurmbrand, Susanne. 2001. *Infinitives: Restructuring and Clause Structure*. Mouton de Gruyter, Berlin.

Yang, Charles. 2010. Three factors in language variation. *Lingua* 120: 1160–1177.