A sociogrammatical analysis of linguistic gaps and transitional forms

Sjef Barbiers - Leiden University

Abstract

This chapter shows that integration of sociolinguistic and generative approaches is necessary to explain systematic linguistic gaps and transitional forms. It provides two case studies to illustrate what the geographic distribution of linguistic forms tells us about language as a cognitive system. The first case study discusses lexical variation in reflexives such as HIMSELF in dialects of Dutch. The wealth of forms of reflexives, with a clear geographical distribution, is constrained by a morphosyntactic agreement condition on their lexico-syntactic structure. The occasional attestation of a form that violates this is explained from its geographic distribution. It is a fudged form that only occurs in dialect contact zones. The second case study discusses word order variation in clause-final three-verb clusters. The wealth of word order variation is constrained, among others, by a principle that requires linearization to be harmonic. The occasional attestation of a word order that violates harmonic linearization is explained from its geographic distribution. It is a fudged order that only occurs in dialect contact zones. Combining formal theoretical and geographical analysis thus makes it possible to distinguish between lexical forms and word orders that are systematic and forms and word orders that are unstable violations of systematic constraints.

1. Introduction[1]

The question what sociolinguists and generative grammarians can learn from each other, whether they should cooperate and if so, how, has been on the table for quite some time now. It was the central question of the workshop *Syntactic Variation* (ICLaVE 2, Uppsala 2003), with Jenny Cheshire as one of the speakers (see her paper *Syntactic variation and spoken language* and the other papers in Cornips and Corrigan 2005). She took up the question again in a paper at the workshop *Is syntax special?* (ICLaVE 7, Malaga 2017), in which she discussed, among other things, the status quo of syntactic variation research, partly on the basis of the papers delivered at that workshop. She concluded there that syntactic variation research is in a very healthy state and that the future looks bright for an improved understanding of syntactic variation. However, she also argued that this improvement mainly comes from bringing together the results of the two approaches, not so much from an integration of the respective perspectives, questions, methods and analyses. For this, the goals of sociolinguistics and generative grammar are too different. Sociolinguistics analyzes the impact of the interaction of language, culture and society on language structure and language change, while generative grammar investigates syntactic variation to model syntax as a cognitive property of humans. Sociolinguists and generative linguists do complementary work that contributes to the big picture of syntactic variation, but they do not really need each other. Here Cheshire seems to reach a similar conclusion as Chomsky (2000). In this chapter I take issue with this conclusion, by showing that integration of sociolinguistic and generative approaches is useful and necessary to explain systematic linguistic gaps and transitional linguistic forms. More specifically, I discuss two case studies to illustrate what the geographic distribution of linguistic forms tells us about language as a cognitive system.

---

2. Systematic language gaps

This chapter takes systematic linguistic gaps, i.e., cases in which a conceivable linguistic pattern has not been attested, as its starting point. A systematic linguistic gap may be observed at the level of all languages (a macrogap), at the level of languages of one language family (a mesogap), or at the level of a set of closely related language varieties (a microgap), such as the dialects of Dutch in the DynaSAND database (Barbiers et al. 2006). The gap can occur at all levels of the mental grammar, including the Lexicon, Syntax, Logical Form (LF) and Phonological Form (PF). In this chapter I discuss a lexical gap and a word order gap. Since I need more than one level of the mental grammar for the analysis and sociolinguistics to explain these gaps, I use the term sociogrammar rather than sociosyntax. The hypothesis is that these linguistic gaps are not accidental but reveal the limits of possible variation – the "envelope of variation" in the Labovian (1972) sense of the term - and more broadly tell us something about the abstract building principles underlying natural language.

A good example of a linguistic gap is discussed in Hale and Keyser's seminal 1993 paper. They argue that verbs can be derived from nominals by incorporation of a noun into an abstract verb. For example, in (1a) the complement *calf* is able to incorporate into an abstract counterpart of *have*, yielding (1b). Hale and Keyser claim that derived verbs such as hypothetical *cow* in (1c) do not exist cross-linguistically, hence a macrogap.

(1) a. A cow had a calf.
   b.   A cow calved.
   c. *A calf cowed.

They explain this by an independently motivated syntactic principle, the ban on incorporation of an argument from a specifier position, in this case the subject *cow*, as opposed to incorporation of an argument from complement position, here the object *calf* (cf. Baker 1988 for discussion of the properties of incorporation). This ban on incorporation from a specfier position follows in turn from the so called Empty Category Principle (ECP), by then a commonly accepted syntactic principle regulating the distribution of empty positions in syntactic structures.[2] The structure corresponding to (1b) is given in (2a), and the illegitimate structure corresponding to (1c) is given in (2b). The positions marked with t_i indicate the base positions of the incorporated nouns.[3]

(2)

---
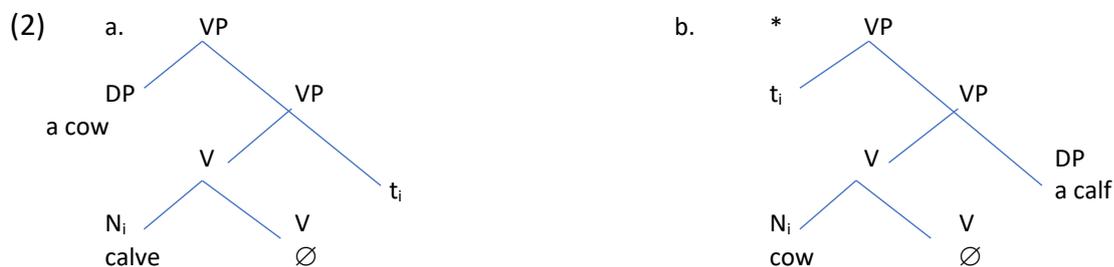
[2] The Empty Category Principle, part of Government and Binding theory (Chomsky 1981), requires that the base position of a moved constituent is properly governed. The idea behind this is that the base position of a moved constituent should be recoverable. The base position of the internal argument is properly governed because the verb V assigns a theta role to it under sisterhood. V does not assign a theta role to the external argument in its base position directly, therefore the external argument is not properly governed by V.
[3]These structures abstract away from surface linear order, which is not relevant given the possibility of *It calved a cow* but not *It cowed a calf*.

This analysis has the important property that a systematic lexical gap is explained by a syntactic principle, the ECP. In this view, verbs are not lexical atoms but lexico-syntactic treelets that express the structural relations between the arguments and the internal aspect (initiation, process, end point) of the verb. Such lexico-syntactic treelets should obey general syntactic principles such as X-bar structure, the ECP or a modern version of these. Syntax is thus also a condition on lexical structures, not a completely independent module.

3. Gap 1: Impossible strong reflexives

3.1 Introduction
I will now discuss two microgaps in the form of strong reflexives in dialects of Dutch. Parallel to the impossible verbs discussed in the previous section, I show that these microgaps can be explained by a morphosyntactic principle, specifier-head agreement. I then proceed to show that apparent counter-examples to one of these gaps can be explained from the geographic distribution of the forms involved.

3.2 The impossibility of HIM OWN and SIG OWN as strong reflexives
I define strong reflexives here as reflexive anaphors that typically occur as the internal argument of transitive verbs. Examples for Dutch and English are given in (3) (strong reflexives in bold).

(3)   a.   Eduard kent **zichzelf** goed.
        b.   Eduard knows **himself** well.

We find the following forms of strong reflexives in the dialects of the Dutch language area (taken from Barbiers et al. 2005; Barbiers and Bennis 2004). Given that they were collected under methodologically identical conditions we can safely assume that these forms are semantically and functionally equivalent and true cases of variation (cf. Buchstaller 2009 and references cited there for discussion of these notions of equivalence). The glosses present the parts of the strong reflexives as syntactico-semantic atoms that are quite common in strong-reflexives cross-linguistically (cf. Safir 1996).

(4)   zich-zelf   SIG-SELF[4]
        hem-zelf   HIM-SELF
        zijn-zelf   HIS-SELF
        zijn-eigen   HIS-OWN
        zijn-eigen-zelf   HIS-OWN-SELF

The geographic distribition of these forms is given in map 1. Some background knowledge of the methodology is necessary for the proper interpretation of the dialect maps in this chapter, as they are an idealization of linguistic reality. The Dutch dialect data in DynaSAND (Barbiers et al. 2006) were collected between 2000 and 2004 with oral interviews among dialect

---

[4] I use SIG to indicate weak reflexive morphemes. One of the contexts in which these occur are inherently reflexive verbs in Dutch, as in *zich herinneren* 'remember'. English does not have a morpheme for SIG, but many other languages do.

speakers between 55 and 75 years old that were born and raised in the locations of the interviews, just like their parents. Research into the full linguistic complexity, variation and change in the locations on the map is not possible with these data, as there are no data from people who do not meet these selection criteria, e.g., from younger people, people originating elsewhere, etc. Also, investigation of intra-speaker variation and change is not possible with these data, as they are based on judgement and translation tasks, not on spontaneous conversations. For extensive descriptions of the methodology, see Barbiers and Bennis (2007) and Barbiers, Cornips and Kunst (2007).

We find in map 1 an eastern and north-western (red) area with the standard Dutch form *zich-zelf* SIG-OWN, a north-eastern (blue) area with the (Frisian) form *hem-zelf* HIM-SELF, a central (yellow) area with the form *zijn-eigen* HIS-OWN, and a south-western (brown) area with the (Flemish) form *zijn-zelve* HIS-SELF. There are also some mixed areas, see below for further discussion. A few southern dialects have the complex form *zijn-eigen-zelve* HIS-OWN-SELF (green). This seems to be a secondary form, strengthening *zijn-eigen*, as all of the dialects that have this complex form also have the simpler *zijn-eigen* HIS OWN, and the addition of *zelf* SELF as a more general strenghtening device. I will not discuss this complex form in this chapter.
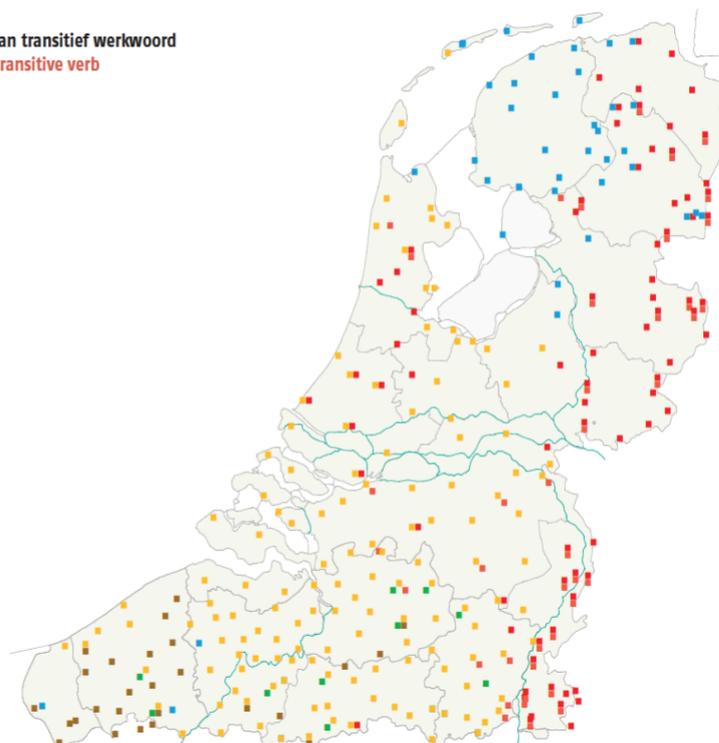
Map 1: The geographic distribution of strong reflexive forms in the Dutch language area
from Barbiers et al 2005: Syntactic Atlas of Dutch Dialects Volume 1. Amsterdam: AUP.

The atoms that the strong reflexives in (4) are made of include: the possessive pronoun *zijn* or *z'n* HIS, the personal pronoun *hem* HIM, the reflexive pronoun *zich* SIG, the possessive word *eigen* OWN and the intensifier *zelf* SELF. Since all Dutch dialects have a form for the personal pronoun *hem* HIM and for the possessive word *eigen* OWN, it is striking that there is no dialect of Dutch that has *hem eigen* HIM OWN or *zich eigen* SIG OWN as a strong reflexive. I hypothesize that these are not accidental microgaps but strong reflexive forms that are

4

impossible due to a general syntactic principle. Barbiers and Bennis (2004) and Barbiers (2009) provide an explanation for these microgaps which I will slightly adapt below.

These explanations are based on work by Helke (1970), Pica (1987), Safir (1996) and Postma (1997), who maintain that strong reflexives cross-linguistically are possessive nominal groups with a pronoun as the possessor and a noun derived from a body part as the nominal head.[5] Postma (1997) proposes that Dutch *zelf* SELF consists of the possessive head *ze* HIS (common in colloquial Dutch as a possessive pronoun) and the body part *lf* (Dutch *lijf* BODY). If we assume that strong reflexives are full nominal groups (DPs), a reasonable assumption given that they have the syntactic distribution of full DPs, their structure would be as in (5).[6]

(5)	[DP	Spec	[D	Possessive Head	[NP	Body Part Head]]]
a.		hem		ze		lf
		him		his		body
b.		zich		ze		lf
		SIG		his		body
c.		z'n		eigen		Ø[7]
		his		own		
d.		z'n		ze		lf
		his		his		body
e.		*hem		eigen		Ø
		him		own		
f.		*zich		eigen		Ø
		SIG		own		

The claim is now that the impossibility of (5e) is due to the lack of agreement between *hem* HIM and *eigen* OWN. This claim is supported by the observation that there is a general gender and number agreement requirement on non-reflexive possessive DPs in colloquial Dutch, as the data in (6) show.

(6)	a.	Jan	ze/*d'r	lijf
		Jan	his/her	body
		'John's body'		
	b.	Marie	d'r/*ze	lijf
		Marie	her/his	body

---

[5] I will not go into the question here as to why natural language needs these complex forms to express reflexivity. See Barbiers (2000) for a proposal. The core idea there is that identity of arguments of a binary predicate reduces the two arguments to one. To avoid that and retain binarity the two arguments must be formally distinct. This is why the personal pronoun must be embedded in a complex DP.

[6] The analysis of the secondary form *z'n eigen zelf* 'his own self' mentioned above as an intensified form of the primary form *z'n eigen* cannot be the same as (5), as *eigen* and the *ze-* part of *zelf* would compete for the same possessive head position. However, it is clear from intensified non-reflexive noun phrases such as *de jongen zelf* 'the boy himself' that *zelf* should be able to occur in other structural positions, possibly adjunct positions, anyway, as in this noun phrase D is already filled by *de* 'the' and N is filled by *jongen* 'boy'.

[7] Interestingly, the N head needs to be empty in the dialects of Dutch when the possessive head is filled with *eigen* OWN. It is clear from the analysis of *zelf* SELF in (5) that *zelf* cannot be in N, as it consists of a possessive head and a nominal body part head. It is apparently impossible to use a different body part noun to fill N and to form a strong reflexive. This suggests that reflexive formation is not a productive process. This justifies the assumption that reflexives are represented in the Lexicon as lexico-syntactic units.

5

c.	de man	z'n/*hun	lijf	
	the man	his/their	body	
d.	de mannen	hun/*z'n	lijf	
	the men	their/his	body	
e.	hem	ze/*d'r/*hun	lijf	
	him	his/her/their	body	

The feature bundles of the elements involved in (5) are given in (7).

(7)	hem	HIM	[3, masculine, singular]
	ze	HIS	[3, masculine, possessive]
	z'n	HIS	[3, masculine, singular, possessive]
	zich	SIG	[3, reflexive]
	eigen	OWN	[possessive]

It is clear that *eigen* OWN does not have person, number or gender features itself, as it is compatible with all members of the paradigm (*m'n-eigen* MY OWN, *je-eigen* YOUR OWN, *z'n eigen* HIS OWN *d'r eigen* HER OWN *hun-eigen* THEIR OWN, etc.). The reflexive *zich* SIG only occurs with third person antecedents, both masculine, feminine, singular and plural ones and I therefore take it to be underspecified for these features and only have the feature bundle [3, reflexive]

If the specifier-head agreement requirement on possessive DPs is more precisely formulated as *the pronoun in the specifier position SpecDP should agree with the possessive head in D on at least one morphosyntactic feature*, then it is immediately clear why *hem eigen* HIM OWN is an impossible strong reflexive. There is not any feature that these two elements share, as opposed to, for example, *z'n* HIS and *eigen* OWN that share the feature [possessive].[8]

A similar reasoning holds for the other form that is unattested according to map 1, *zich eigen* SIG OWN (5f). The pronoun *zich* SIG has the feature bundle [3, reflexive] and does not share any feature with *eigen* OWN, which only has the feature [possessive]. This analysis is simpler and more elegant than the one provided in Barbiers and Bennis (2004), where we ruled out *zich eigen* SIG OWN by assuming that *-ich* in *zich* and *eigen* are both possessive heads competing for the same D-position. One of the problems with that analysis it that it requires additional assumptions (and structure) to account for the existence of the strong reflexive *zich-zelf* SIG SELF, which contains both possessive heads.

3.3 Unexpected cases of *zich eigen* SIG OWN
Van der Feest (2007) observed that the strong reflexive *zich-eigen* SIG OWN occasionally turned up in texts on the internet. She located these hits geographically where possible and then went to the locations identified in this way to check with oral interviews whether these forms were real. The results of this research are given on map 2.[9]

---

[8] This implies that in forms such as *me-eigen* MY OWN, *je-eigen* YOUR OWN, *me* and *je* should have a possessive feature to be able to agree with *eigen*. This is correct, given that *me* and *je* can be used as possessive pronouns in noun phrases such as *me boek* 'my boek' and *je boek* 'your book'.
[9] The symbol that combines a circle and a cross is assigned to locations where the form was found both on the internet and in the interview.
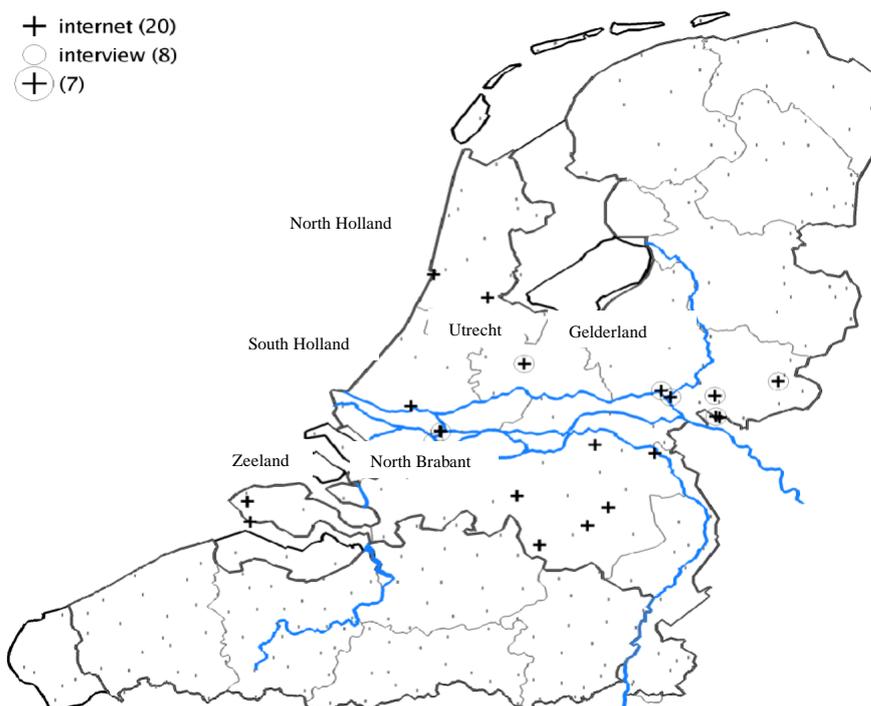
Map 2: Attestations of the form *zich eigen* SIG OWN in the Dutch language area
(based on van der Feest 2007)

If we compare maps 1 and 2, we see that the locations in the east (eastern North Brabant and eastern Gelderland) that have *zich-eigen* SIG OWN are all locations that are close to the border between the eastern *zich-zelf* SIG SELF zone and the central *z'n-eigen* HIS OWN zone. The locations in the west that have *zich-eigen* SIG OWN are in an area with dialects in which both *zich-zelf* SIG SELF and *zijn-eigen* HIS OWN are possible (South Holland) or at the border between the *zijn-eigen* HIS OWN area and the *zich-zelf* SIG SELF area (North Holland and Utrecht). The two dialects of Zeeuws on the peninsula of Walcheren in Zeeland in the south-west seem to be an exception to the observation that dialects with *zich-eigen* are in contact zones.

Putting the two dialects of Zeeuws aside for the moment, this suggests that the form *zich-eigen* SIG OWN is the result of dialect contact, combining the first part of *zich-zelf* and the second part of *z'n-eigen* into the form *zich-eigen*. This fudged form (cf. Chambers and Trudgill 1998) arises in areas with two or more competing grammars (cf. Kroch 1989 and the Introduction in Cornips and Corrigan 2005 for discussion). The fact that it was not found in the fieldwork on which map 1 is based and that it does not have a geographic distribution corresponding to any known dialect area suggests that it is not a stable form, i.e., that it does not belong to the grammatical system of any dialect speaker or group.

That is different for the other logical possibility, the form *z'n-zelf* HIS OWN (5d), which combines the first part of *z'n eigen* HIS OWN and the second part of *zich-zelf* SIG SELF. Map 1 shows that this form is typical for West Flanders, a well established dialect area with many distinctive dialect properties. Put differently, as opposed to *zich-eigen* the form *z'n-zelf* seems to be a grammatically stable option. The explanation for this contrast has already been given

above. While *z'n* HIS and the *ze*-part of *zelf* SELF share the features [3, possessive], *zich* SIG and *eigen* OWN do not share any grammatical feature and therefore do not agree with each other.

If this analysis is on the right track, it illustrates how research into syntactic variation requires a sociogrammatical approach, a combination of generative and sociolinguistic methods and perspectives (cf. Adger and Smith 2010 for another illustration). The sociolinguistic (or dialectological) method in this case study is investigating the geographic distribution of the various forms of strong reflexives. The generative method is to analyze the lexico-syntactic forms attested and not-attested as hierarchical structures, DPs, that must meet a specifier-head agreement requirement, and to show that they are different realizations of one abstract form. It is really the interaction between sociolinguistic factors, in this case geographical distribution and ensuing dialect contact, and systematic formal requirements, in this case specifier-head agreement in a hierarchial DP structure representation in the Lexicon, that sheds light on the status of contact forms such as *zich-eigen* SIG OWN and provides a fuller explanation of the properties of the microgaps.

Needless to say, further investigation of the contact dialects and speakers that have *zich-eigen* SIG OWN is necessary to test the correctness of this analysis and to gain further insights. For example, we expect to find in these dialects contact forms for the other members of the reflexive paradigm, as well as beyond the pronominal system. More complex questions for future research about such dialects include: Do these speakers have two distinct dialect systems with occasional contact phenomena, or are the contact phenomena part of a separate, fused dialect system? This question can not be answered on the basis of the data that are currently available. It requires, among others, a sufficiently large corpus of spontaneous speech to find out whether and how often speakers that use *zich-eigen* also use the strong reflexives from the neighbouring dialects within the same register. Another question is whether contact forms such as *zich-eigen* SIG OWN can be reanalyzed in a later stage of the dialects such that they no longer violate grammatical principles and become stable, and if so, what such a reanalysis would look like? More generally, it is an important question whether there are any constraints on such fudged forms themselves.[10] We can also ask why contact between the *standard* language, which has *zich-zelf* SIG SELF, and dialects that have *zijn-eigen* HIS OWN usually does not seem to lead to the contact forms described above, while contact between dialects does. If contact between the standard language and the dialects would easily yield forms such as *zich-eigen* SIG OWN, we would expect a much larger number of dialects to have it than the ones we find on map 2. The only candidates are the two dialects of Zeeuws mentioned above. Finally, it is an open question whether the West-Flemish form *z'n zelf* HIS SELF is due to contact between dialectal *z'n eigen* and standard *zich-zelf*, or just an independent realization of one the possible structures.

4. Gap 2: Impossible word orders in verb clusters

It has been observed many times (cf. Wurmbrand 2006, Zwart 1996 that there is one word order in three-verb clusters in Dutch and German that seems to be systematically excluded, the so called 2-1-3 order (8e). The numbers in this notation correspond to the hierarchical positions of the verbs involved: 1 (*moet* 'must' in 8a) for the highest verb in the syntactic tree, that selects 2 (*kunnen* 'can.inf' in 8a), and 3 for the lowest verb in the syntactic tree (*zwemmen*

---

[10] I thank an anonymous reviewer for pointing this out.

'swim.inf' in 8a), that is selected by 2. A sequence of the type 2-1-3 thus indicates how the original hierarchical order (1 higher than 2, 2 higher than 3) is linearized. The observation that the linear order 2-1-3 does not exist has been replicated in the large scale dialect syntax survey *Syntactic Atlas of the Dutch Dialects* (cf. Barbiers et al. 2006). This linguistic gap is illustrated in (8) for a Dutch verb cluster with two modal auxiliaries. The order 2-3-1 in (8f) is excluded as well when the two auxiliaries are modals, but not when the first auxiliary is a perfective auxiliary and the second a modal. I will not discuss this 2-3-1 order any further in this paper (cf. Barbiers, Bennis and Dros-Hendriks 2018; henceforth BBD).

(8) a. Ik vind dat iedereen goed 1-moet 2-kunnen 3-zwemmen.
         I find that everyone well must.INF can.INF swim.INF
         'I think that everyone should be able to swim.'
     b. Ik vind dat iedereen goed 1-moet 3-zwemmen 2-kunnen.
     c. Ik vind dat iedereen goed 3-zwemmen 2-kunnen 1-moet.
     d. Ik vind dat iedereen goed 3-zwemmen 1-moet 2-kunnnen.
     e. *Ik vind dat iedereen goed 2-kunnen 1-moet 3-zwemmen.
     f. *Ik vind dat iedereen goed 2-kunnen 3-zwemmen 1-moet.

Map 3 shows the geographic distribution of the attested orders in (8).
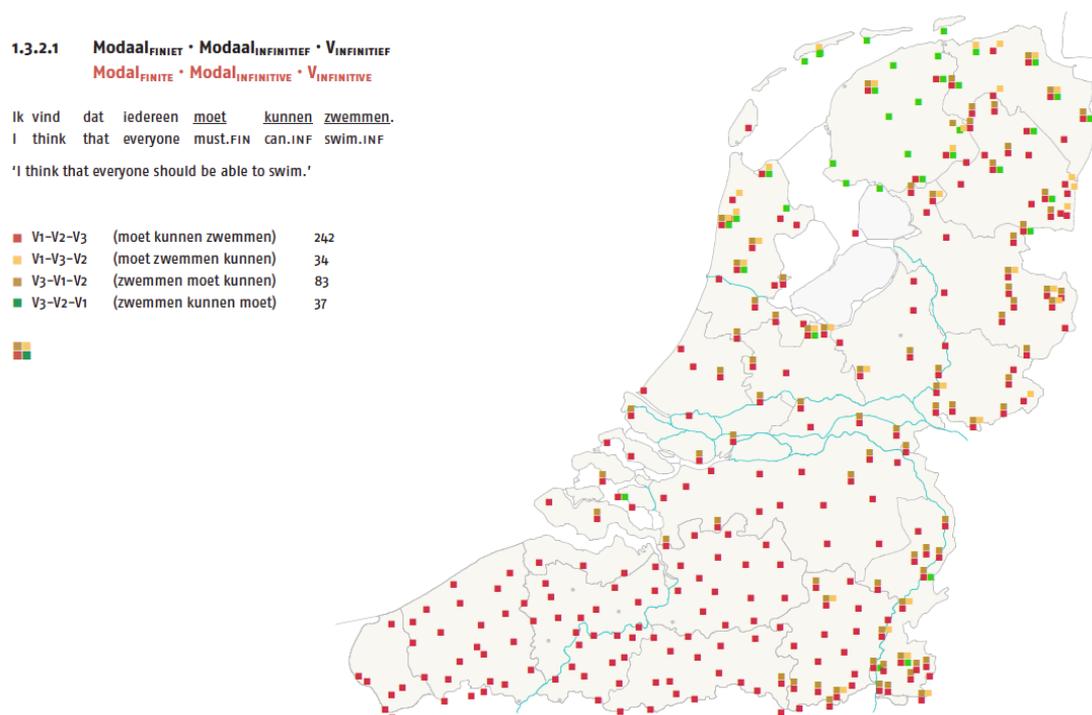
Map 3: The geographic distribution of word orders in three-verb clusters containing two modals in the Dutch language area from Barbiers et al 2008. Syntactic Atlas of the Dutch Dialects, Volume II. Amsterdam: AUP.

From north to south roughly three dialect areas can be distinguished: a relatively homogeneous northern (green) area in Frisia and surroundings with dialects that only have the 3-2-1 order, a southern (red) area with only the 1-2-3 order and an area in between where many dialects have both the 1-2-3 order and the 3-1-2 order. The distribution of the 1-3-2

9

order does not correspond to any known dialect area. We see that it occurs particularly along the border with Germany, both in the Low-Saxon area (north-east) and the Franconian area (south-east). In addition, we find this order in the north-western transition zone between the Hollandic dialects and the Frisian dialects, and we find it in the north-eastern transition zone between the Frisian and the Low-Saxon dialect areas. Since 3-2-1 is a possible order in German too, the generalization seems to be that the order 1-3-2 occurs in transitional areas between dialect areas that have 1-2-3 and dialect areas that have 3-2-1. It is furthermore striking that there in not any dialect in which this 1-3-2 order is the only possible order.

This geographic distribution of the 1-3-2 order is reminiscent of the distribution of the strong reflexive pronoun *zich-eigen* SIG-OWN discussed in the previous section. Like *zich-eigen*, the 1-3-2 order does not occur in any known dialect area, but only in transitional areas where 1-2-3 orders and 3-2-1 orders meet. Like *zich-eigen*, the 1-3-2 order can be seen as a property of fudged varieties in the sense of Chambers and Trudgill (1998), Britain and Trudgill (2005) as it combines grammatical properties of the two surrounding orders: 1-3-2 is like 1-2-3 in that 1 precedes 2 and 3, and it is like 3-2-1 in that 3 precedes 2. In the words of BBD (2018): dialects of Dutch either have ascending (1-2-3) word order in verb clusters, or they have descending (3-2-1) order. Word orders in transitional dialects are ascending and descending at the same time.

In the case of *zich-eigen* SIG-OWN, I have claimed above that this is an unstable form in current dialects of Dutch, as there is no grammatical agreement between the two parts of the form. Therefore, *zich-eigen* is either bound to show up occasionally in space and time, to be reanalyzed or to disappear. The question is now whether 1-3-2 also violates a grammatical principle and is therefore unstable.

According to BBD (2018), the grammar of verb clusters works as follows. In the syntactic module of the mental grammar the hierarchy of all three-verb clusters is identical, regardless of the linear surface word order. This hierarchy looks as in (9).

(9)

```
                    VP1
              ╱          ╲
           V1             VP2
                       ╱       ╲
                     V2          V3
```
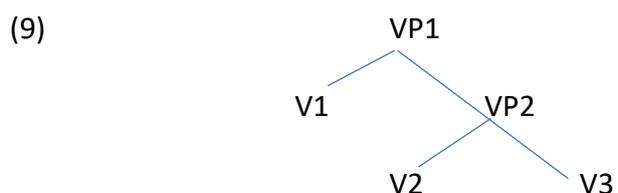
The PF-module of the mental grammar is the module where hierarchical structures such as (9) get linearized.[11] BBD (2018) claim that there are two linearization options which are both harmonic: ascending 1-2-3 or descending 3-2-1. All other orders attested in the Dutch language area are exceptional in that they have an additional property that seemingly blurs this harmonic picture. We can therefore clean up maps such as map 3 by removing all orders

---

[11] PF (Phonological Form) is the module of the mental grammar that is the interface between the hierarchical structure built in Syntax and the sensorimotor system that takes care of the externalization of linguistic structures as sounds or signs. PF maps a hierarchical structure to a linear structure which is read off by the sensorimotor system. Other modules of the mental grammar are the Lexicon and LF (Logical Form). LF is the interface between the hierarchical structure built in Syntax and the conceptual-intentional cognitive system. Cf. Chomsky (2000).

that are exceptional. Map 4 shows the resulting map, for three different types of three-verb clusters, clusters with two modal auxiliaries and a main verb (*moet kunnen zwemmen* 'must.fin can.inf swim.inf'), three-verb clusters with a modal auxiliary, a perfective auxiliary and a main verb (*moet hebben gemaakt* 'must.fin have.inf made.pcp') and verb clusters with a perfective auxiliary, an aspectual auxiliary and a main verb (*is gaan zwemmen* 'is.fin go.inf. swim.inf').
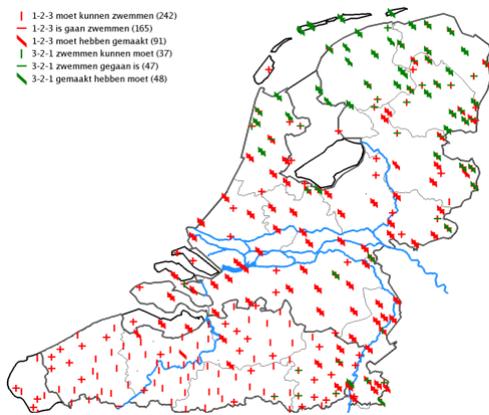
Map 4: Dialects of Dutch that have at least one 3-2-1 order (green) and dialects that have at least one 1-2-3 order (red). from BBD (2018)

What about the exceptional orders? BBD (2018) provide arguments showing that 3-1-2 orders are not real verb clusters, as 3 should be analyzed as an adjective when it is a participle, and as a noun when it is an infinitive. Similarly, 1-3-2 orders are not real verb clusters when 3 is a participle, as in such cases the participle is an adjective as well. This is supported by the fact that the 1-3-2 order with 3 a participle occurs particularly in the Belgian part of the Dutch language area and in this part many dialects allow verb clusters to be interrupted by other types of syntactic constituents as well. Maps 5 and 6 illustrate this correlation between verb cluster interruption and 1-3-2 order with 3=participle.[12]

---

[12] Particles, as part of so called separable compound verbs (e.g., *op-geven* lit. up give 'give up') are deliberately left out of consideration. They can interrupt verb clusters in the entire language area and seem to have a syntactic status that is different from the interrupting categories depicted in map 5.
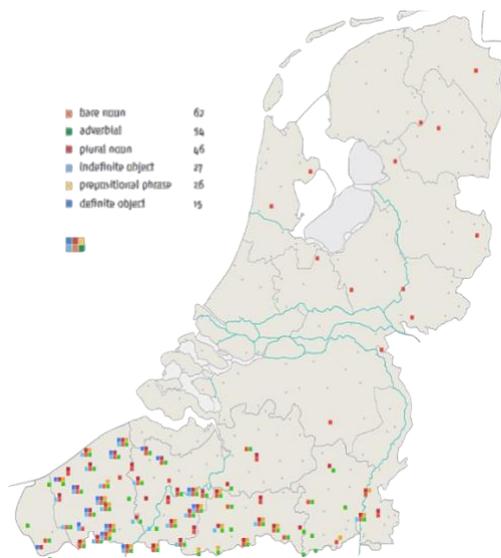
Map 5 Geographic distribution of verb cluster interruption

Map 6 Geographic distribution of 1-3-2 with 3 = participle (yellow)

both maps from Barbiers et al 2008. Syntactic Atlas of the Dutch Dialects, Volume II. Amsterdam: AUP.

BBD (2018) explain the possibility of the 2-3-1 order for the cluster type *is gaan zwemmen* 'is.fin go.inf. swim.inf' in a similar way. The impossibility of the order 2-1-3, the systematic linguistic gap, is explained by its disharmonic nature (ascending and descending at the same time) and the fact that this order cannot be derived even if 2 is taken to be adjectival or nominal. Since there is a direct (selectional) dependency between 2 and 3, the verb in position 1 cannot occur in between.

With this background we can answer the question whether the verb cluster order 1-3-2 with 3 an infinitive is an ungrammatical order (cf. BBD 2018 for more extensive discussion). The only possibility to get this order is if 3 can be analyzed as a nominalized verb, parallel to the analysis of southern Dutch 1-3-2 orders in which the participle has an adjectival status. However, whereas the latter correlates with other cases of interruption of the verb cluster by non-verbal material, the 34 dialects that have 1-3-2 orders with 3 an infinitive do not allow any of the cluster interruptions that are allowed in southern Dutch (cf. map 5). We conclude, then, that the order 1-3-2 with 3 an infinitive as it occurs in the transitional zone along the border with Germany is an ungrammatical order as it is disharmonic and verb 3 between 1 and 2 cannot be analyzed as a nominalization. Like *zich-eigen* SIG OWN, this order is bound to show up occasionally in space and time, to be reanalyzed or to disappear.

An obvious direction of reanalysis would be that 3 is reanalyzed as a nominal. As already noted, that would still not make this order a grammatically well-formed order in the eastern transitional dialects, as these do not allow verb cluster interruption. The order is also not expected to get grammaticalized easily in the southern dialects that have verb cluster interruption, as these dialects for some unknown reason resist nominalization of the main verb in the cluster. This can be seen from another contrast between the southern dialects and the central Dutch dialects. The latter allow the order 3-1-2 with the cluster type *moet kunnen zwemmen* 'must.fin can.inf swim.inf, the former don't. If BBD (2018) are right that 3 is a

nominal infinitive in such as case, then the absence of 3-1-2 in the southern Dutch dialects shows that they resist nominalization of infinitives in verb clusters.

5. Conclusion

The two case studies of this paper show that explaining the full complexity of linguistic microvariation requires integration of the sociolinguistic and the generative perspectives. The case studies included lexical variation in the expression of reflexive anaphors and word order variation in verb clusters. In both cases the goal was to explain systematic linguistic gaps and the occurrence of exceptional forms. The starting point of each of the case studies was the geographic distribution of the various forms and word orders, a traditional dialectological and sociolinguistic method. This geographic distribution immediately showed that certain logically possible forms and orders are absent, the systematic linguistic gaps. It also showed that there are unexpected lexical forms and word orders that occur in areas that do not correspond to traditional dialect areas.

More specifically, I identified two systematic linguistic gaps, the reflexive anaphors *hem-eigen* SIG OWN and the word order 2-1-3 as in *kunnen moet zwemmen* 'can.inf must swim.inf', and two exceptional forms, i.e. forms that should not exist, the reflexive anaphor *zich-eigen* SIG OWN and the word order 1-3-2, as in *moet zwemmen kunnen* 'must swim.inf can.inf'. The systematic gaps were explained by formal principles from generative grammar: a violation of the specifier-head agreement condition on possessive nominal groups in the case of *hem-eigen* HIM OWN and a violation of harmonic linearization (ascending or descending) in the case of the 2-1-3 order.

Strictly speaking, the geographic distribution of the various forms and orders would not have been necessary for this explanation of the linguistic gaps, although it certainly helps that we have a dataset at our disposal that covers the entire language area and that was collected systematically with sociolinguistic methods. This makes the chance that the observed gaps are accidental considerably lower.

Taking the geographic distribution into account is crucial, however, for the explanation of forms and orders that are unexpected given the proposed formal explanations of the linguistic gaps. We have seen that the unexpected form *zich-eigen* SIG OWN that violates the specifier-head agreement condition is a fudged form in the sense of Chambers and Trudgill (1998) that shows up occasionally in areas with contact between dialects that have *zich-zelf* SIG SELF and dialects that have *zijn-eigen* HIS OWN. We have also seen that the word order 1-3-2 (*moet zwemmen kunnen* 'must swim.inf can.inf') that violates harmonic linearization is a fudged word order that shows up in areas with contact between 1-2-3 orders and 3-2-1 orders. These insights could only be gained by integrating the formal analysis that is common in generative grammar and the distributional analysis that is common in dialectology and sociolinguistics.

References

Adger, D and Smith, J. 2010. Variation in Agreement: a lexical feature-based approach, *Lingua* 120.5, 1109-1134.

Baker, M. 1988. A Theory of Grammatical Function Changing. Chicago: University of Chicago Press.

Barbiers, S. 2000. On the interpretation of movement and agreement: PPs and Binding. In H. Bennis, M. Everaert, & E. Reuland (Eds.), *Interface Strategies.* (pp. 21-36). Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

Barbiers, S. et al 2006. *Dynamische Syntactische Atlas van de Nederlandse Dialecten (DynaSAND).* Amsterdam, Meertens Instituut. URL: http://www.meertens.knaw.nl/sand/.

Barbiers, S. and H. Bennis 2004. Reflexives in dialects of Dutch. In J. Koster and H. van Riemsdijk (eds). *Germania et alia. A Linguistic Webschrift for Hans den Besten*. Electronic publication, University of Groningen, Groningen. http://odur. let. rug. nl/~ koster/DenBesten/contents. htm

Barbiers, S., H. Bennis, G. De Vogelaer, M. Devos, and M. van der Ham. 2005. *Syntactic Atlas of the Dutch Dialects*, Volume 1. Amsterdam: Amsterdam University Press.

Barbiers, S. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics, in L. Cornips and K. Corrigan (eds.) *Syntax and Variation: Reconciling the Biological and the Social*. John Benjamins, Amsterdam, 233-264.

Barbiers, S. J. van der Auwera, H. Bennis, E. Boef, G. De Vogelaer, M. van der Ham 2008. Syntactische Atlas van de Nederlandse Dialecten / Syntactic Atlas of the Dutch Dialects.- Volume 2. Amsterdam: Amsterdam University Press.

Barbiers, S. and H. Bennis 2007. The Syntactic Atlas of the Dutch Dialects. A discussion of choices in the SAND-project. *Nordlyd* 34, 53-72.

Barbiers, S. L. Cornips and J.P. Kunst 2007. The Syntactic Atlas of the Dutch Dialects: A corpus of elicited speech and text as an on-line dynamic atlas. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora. volume 1: Synchronic databases.* (pp. 54-90). Hampshire: Palgrave Macmillan.

Barbiers, S. 2009. Locus and Limits of Syntactic Variation. *Lingua* 119 (11), 1607-1623

Barbiers, S., H. Bennis and L. Dros-Hendriks 2018. Merging verb clusters. *Linguistic Variation* 18 (1), 144-196.

Britain, D. and P. Trudgill 2005. New dialect formation and contact-induced reallocation: Three case studies from the English fens. *International Journal of English Studies*, *5*(1), 183-209.

Buchstaller, I. 2009. The Quantitative Analysis of Morphosyntactic Variation: Constructing and Quantifying the Denominator. Language and Linguistics Compass 3/4), 1010–1033.

Chambers, J. K. and P. Trudgill 1998. *Dialectology*, second edition. Cambridge: Cambridge University Press.

Cheshire, J. 2003. Syntactic variation and spoken language. Paper presented at the workshop *Syntactic Variation,* ICLaVE 2, Uppsala.

Cheshire, J. 2005. Syntactic variation and spoken language. In L. Cornips and K. Corrigan (eds.) *Syntax and Variation. Reconciling the Biological and Social* [Current Issues in Linguistic Theory 225]. Amsterdam/Philadelphia: John Benjamins.

Cheshire, J. 2017. Is syntax special? Paper presented at ICLAVE 7, Malaga 2017.

Chomsky, N. 1981. *Lectures on Government and Binding: The Pisa Lectures.* Berlin: Mouton de Gruyter.

Chomsky, N. 2000. New Horizons in the Study of Language and Mind. Cambridge University Press.

Cornips, L. and K. Corrigan (eds.) *Syntax and Variation. Reconciling the Biological and Social* [Current Issues in Linguistic Theory 225]. Amsterdam/Philadelphia: John Benjamins.

Feest, J. van der 2007. Een ezel stoot zich eigen in 't gemeen niet twee keer aan dezelfde steen. Unpublished bachelor thesis, Utrecht University.

Hale, K., and J. Keyser 1993. On argument structure and the lexical expression of syntactic relations, in Hale and Keyser (eds.) *The View from Building 20: A Festschrift for Sylvain Bromberger,* MIT Press, Cambridge, MA. 53-108.

Helke, M. 1971. *The Grammar of English Reflexives*, Ph.D. dissertation, MIT.

Kroch, A. 1989. Reflexes of Grammar in Patterns of Language Change. *Language Variation and Change 1,* 199-244.

Labov, W. 1972. Some principles of linguistic methodology, *Language in Society* 1: 97–120.

Pica P. 1987. On the nature of the reflexivization cycle, in J. Mcdonough and B. Plunkett (eds.), *Proceedings of NELS 17*, GLSA, Amherst.

Postma G. 1997, Logical Entailment and the Possessive Nature of Reflexive Pronouns, in H. Bennis, P. Pica and J. Rooryck (eds.), *Atomism and Binding*, Foris Publications, Dordrecht, 295–322.

Safir, K. 1996. Semantic Atoms of Anaphora. *Natural Language & Linguistic Theory*, Volume 14 (3), 545–589.

Wurmbrand, S. 2006. Verb clusters, verb raising, and restructuring. In M. Everaert and H. van Riemsdijk (eds.) *The Blackwell Companion to syntax*, Volume V, Article 75, 229-343. Oxford: Blackwell.

Zwart, J.-W. 1996. Verb clusters in Continental West Germanic dialects. In James Black & Virginia Motapanyane (eds.), *Microparametric syntax and dialect variation*. 229-258. Amsterdam: John Benjamins