

Ethnolect speakers and Dutch partitive adjectival inflection

A corpus analysis

Dirk Pijpops & Freek Van de Velde

TET 67 (2): 343–371

DOI: 10.5117/TET2015.2.PIJP

Abstract

This study applies the methodology described by Gries & Deshors (2014) within the framework of the Contrastive Interlanguage Analysis (Granger, 1996) to the partitive genitive inflection in post-quantifier adjectives in the Moroccan Dutch ethnolect. This implies fitting a logistic regression model on data from the complementary ConDiv and Moroccorp corpora to investigate the differences between the L1 variety and the (early L2/2L1) ethnolect variety. It was found that the Moroccan Dutch language users do not differ from 'ordinary' Dutch language users in the realisation of the partitive genitive *-s* suffix, neither through an outspoken preference for one of the inflectional variants, nor in the factors determining the alternation. This is considered a rather surprising result, as such differences do exist for a number of other grammatical phenomena (Cornips and Rooij, 2003; Van de Velde and Weerman, 2014). This finding can tell us something about the inflectional status of the partitive genitive. It appears that it is less non-transparent than other quirks in adjectival inflection.

Keywords: partitive, genitive, logistic regression, Contrastive Interlanguage Analysis, Moroccan Dutch, adjectival inflection, ethnolect

1 Introduction¹

1.1 The partitive genitive *-s* in Dutch

Over time, Dutch attributive adjectival inflection has been drastically reduced, as part of an overall deflexion tendency (van der Horst, 2008, p.143;

van der Horst, 2013), which has especially targeted the nominal domain (Schönfeld, 1970, p.117; Harbert, 2007, p.90). It seems the only vestige of the once rich adjectival case inflection system is an alternation between a schwa and $-\emptyset$ ending. The inflectional schwa is, however, not the only ending a Dutch adjective may receive. Hiding in an inconspicuous corner of Dutch grammar, an $-s$ ending has also survived the turmoils of deflexion. This $-s$ suffix can be attached to adjectives when they are postmodifying an indefinite pronoun or numeral (Haeseryn et al., 1997, p.412), as in (1)-(3).²

(1) *iets* *bijzonder-s*
 something special-GEN
 'something special'

(2) *wat* *zinnig-s*
 something sensible-GEN
 'something sensible'

(3) *veel* *goed-s*
 much good-GEN
 'a lot of good things'

The genitive case has proven to be quite resilient, surviving well into the twentieth and even twenty-first century (Weerman and De Wit, 1999; Scott, 2011, 2014). Perhaps its most well-known remnant is the prenominal genitive $-s$, which is used to mark possession, as in (4). However, in present-day Dutch, the possessive genitive can only be used on proper names and common names used as terms of address and can only take the form of an $-s$ suffix, indiscriminately applied to all genders (Haeseryn et al. 1997, p. 163). Even so, it faces competition, in particular by the so-called *z'n*-construction, as in (5) (also known as prenominal periphrastic possessive, resumptive possessive pronoun or possessor doubling construction, see Weerman and De Wit, 1999; van der Horst and van der Horst, 1999, pp. 164-165; Harbert, 2007, pp. 158-161; Allen, 2008, pp. 186-222; Hendriks, 2012).

(4) *Dirk-s* *boek*
 Dirk-GEN book
 'Dirk's book'

- (5) *Dirk z'n boek*
 Dirk his book
 'Dirk's book'

Like its more famous possessive sibling, the partitive genitive, which will be the subject of this article, has become more limited in use and form throughout its history, as well as subject to competition threatening its very existence. As its name implies, it could formerly express a much wider range of partitive meanings, and could appear on nouns following any kind of quantifier, as exemplified in (6)-(8). In present-day Dutch, however, the partitive suffix can only be used if the quantifier is an indefinite pronoun or numeral followed by an adjectival phrase. Meanwhile, formally, only the *-s* survives, with all other genitive endings disappearing from the language's history. Lastly, an alternative exists in the form of a construction without *-s* suffix (9), which is most popular in – but not limited to – informal language use in the South of the Dutch language area (Pijpops and Van de Velde, 2014).

- (6) *veel goed-er ghedachten*
 many good-GEN thoughts
 'many good thoughts'
 (Middle Dutch, van der Horst, 2008, p.575)

- (7) *een pont speck-s*
 a pound bacon-GEN
 'a pound of bacon'
 (Middle Dutch, van der Horst, 2008, p.575)

- (8) *een corste broot-s*
 a crust bread-GEN
 'a crust of bread'
 (Early Modern Dutch 16th century, van der Horst, 2008, p.1033)

- (9) *iets bijzonder*
 something special
 'something special'

All this leaves the partitive genitive *-s* contrastively, synchronically and diachronically in a peculiar place. Contrastively and synchronically, there seems to be no need to have the suffix, as is demonstrated by English, a

sister language of Dutch, and by Dutch itself, both of which make use of the alternative in (9). Also note that the *-s* in German *etwas Gutes* is not analogous to the Dutch partitive genitive *-s*, but is rather part of the regular and productive general German adjectival inflection. Here, it signifies the neuter singular nominative or accusative case of the entire phrase, as opposed to the Dutch *-s*, which historically signified the genitive case of only the post-modifier, as can be seen when comparing (10) to (11).

- (10) German: *etwas* *Gut-(e)s*
 [something good]-NOM/ACC-NTR-SG
 Dutch: *iets* *goed-s*
 [something good-GEN]
 ‘something good’

- (11) German: *zu etwas* *Gut-em*
 to [something good]-DAT-NTR-SG
 Dutch: *tot iets* *goed-s*
 to [something good-GEN]
 ‘to something good’

Another peculiarity of the partitive genitive is that the adjective follows the quantifier/numeral it is modifying instead of preceding it, as is usual in Dutch. This tendency for a modifier-head sequence is actually becoming ever stricter (van der Horst, 2008, pp. 1946-1961; Van de Velde, 2009, Ch. 3).³ Lastly, the use of genitival inflection on the adjective is quite uncommon, especially in light of the observation that nouns lost their genitive morphology completely in partitive constructions. The present-day counterparts of examples (6)-(8) all lack genitival inflection.

This peculiar situation in Dutch grammar has attracted a number of diverse theoretical analyses (Schultink, 1962, p. 62; Abney, 1987; Kester, 1996; van Marle, 1996; Broekhuis and Strang, 1996; Haeseryn et al., 1997, p. 356, 432; Hoeksema, 1998a; Booij, 2010a, pp. 223-228; Broekhuis, 2013, pp. 419-461). It is not the aim of this article to delve into this discussion. A large part of it revolves around the question which element is the head, and what bracketing structure should be assumed for this binominal construction, and this is not our central concern here. What is of importance to us is how ‘transparent’ the function of the *-s* suffix is.

1.2 Non-transparent morphology and the partitive genitive

Some affixes transparently map onto a certain lexical meaning or grammatical function, whereas others have a meaning or function that is not clearly delineated or are non-transparently constrained in their use. Such non-transparent affixes are often obsolescent. An example are the thematic vowels of Indo-European nouns. They are assumed to originally have had derivational meanings, but the meanings are no longer reconstructable. One option for language users is to tolerate non-transparent morphology, and use it as superfluous or irregular inflection. But there is another option as well: quite frequently, language users can be seen to refunctionalise obsolescent morphology in a process of ‘exaptation’ (Lass, 1990; Van de Velde and Norde, 2016).⁴

The aversion to quirky morphology is arguably bigger in L2 speakers. The later they acquire the language, the more difficulty they have with morphology, especially with morphology that is constrained in a grammatically complex way (Kortmann and Szmrecsanyi, 2012). This is the reason that languages with a high proportion of L2 speakers tend to be morphologically less complex (Kusters, 2003; Lupyán and Dale, 2010; Trudgill, 2011; Bentz and Winter, 2013). Simplification of complex morphology can either be achieved by discarding the morpheme, as has been done in English plural verbs morphology, or by simplifying the constraints on the morpheme, as has been done in Dutch plural verbs, where the *-en* suffix extended to second person, which used to be expressed by *-t*, to give a simple example. This makes the *-en* more transparent: it marks PLURAL, rather than 1/3.PLURAL.⁵

(12) Middle Dutch (Loey, 1980, p. 55)	Present-day Dutch
<i>wi nem-en</i> (we take-1/3PL)	<i>wij nem-en</i> (we take-PL)
<i>gi neem-t</i> (you:PL take-2PL)	<i>jullie nem-en</i> (you:PL take-PL)
<i>si nem-en</i> (they take 1/3PL)	<i>zij nem-en</i> (they take-PL)

In the domain of Dutch adjectival inflection, the notion of transparency is demonstrably at play: Dutch adjectives alternate between an inflected form with *-ə* and a bare form. In their most basic form, ignoring numerous semantically, grammatically or phonologically conditioned exceptions, the rules can be outlined as in (13): the bare form is associated with predicative use (13a) and the inflected form with attributive use (13b), except in the condition where the adjective is part of a singular indefinite neuter NP (13c).⁶

- (13) a. Predicative use: ADJ-Ø
het boek is moeilijk
 the book is difficult:BARE
 'the book is difficult'
- b. Attributive use: ADJ-ə
het moeilijk-e boek
 the difficult-INFLECTED book
 'the difficult book'
- c. Except: [+sg -def +neutr] NPs: ADJ-Ø
een moeilijk boek
 a difficult:BARE book
 'a difficult book'

A radical simplification of the situation as outlined in (13a-c) would be to get rid of the -ə inflection. This is the option that English took in the course of its history. Another, less radical simplification would be to keep the -ə, but just ignore the constraint in (13c). That would amount to a refunctionalisation of the -ə suffix, which would then transparently mark attributive use, as opposed to predicative use. Van de Velde and Weerman (2014) show that this is indeed what is happening in Dutch, where all kinds of exceptional patterns are made to conform to a simpler system (13a-b), and that the ongoing changes are more outspoken in the Moroccan-Dutch ethnolect, underscoring the role of (early) L2 acquisition.⁷ We will return to the adjectival -ə below.

The partitive genitive -s in Dutch is also part of the adjectival inflection domain, and is, at first sight, rather quirky and non-transparent. The reason is that there is variation in the expression of the -s in the partitive construction of a quantifier + adjective. The -s ending can be dropped. The variation in the -s drop is mentioned in passing by Booij (2010a, p. 244) and Broekhuis (2013, p. 426), but it is dismissed as a regional feature of southern varieties. Now, it is true that large-scale -s omission is typical of the southern varieties, but as van der Horst (2008, pp. 1624-1625) points out, occasional omission is also attested in northern varieties, and such omissions are possibly even on the increase. Pijpops and Van de Velde (2014) show that the expression of -s is multifactorially determined, as summarised in (14).

- (14) More -s omission in:
- a. colour adjectives
 - b. adjectives *beter, fout, goed en verkeerd*

- c. informal registers
- d. low-frequency adjectives
- e. patterns with quantifiers *iets* and *niets*, as opposed to *wat en weinig* (quantifiers *veel* en *zoveel* give equivocal results) in Flanders. In the Netherlands, no difference can be detected between the quantifiers.

This looks like a strange assortment of constraints on the partitive *-s* suffix, but there is a good explanation for most of them.

Constraints (14a) and (14b) can be explained by the influence of superficially resembling constructions: colour adjectives are morphologically indistinguishable from colour nouns, so that *zoveel oranje* ‘so many orange things’ in (15), as opposed to *zoveel interessant* ‘so many interesting things’, is structurally ambiguous between Q + ADJ and Q + N (Van de Velde 2001, pp. 150-151). This structural ambiguity is then carried over to cases like *iets oranje* ‘something orange’, which can only be interpreted as Q + ADJ, as *iets* does not premodify nouns.

Similarly, *iets goed doen* as in (16) is structurally ambiguous between a reading where *goed* is in the partitive Q + ADJ construction: [_{VP} do [_{object} something good]] and a reading where *goed* is an adverb: [_{VP} do [_{Object} something] in a good way].⁸ Pijpops and Van de Velde (forthc.) were able to demonstrate that *-s* omission occurred even when the adverbial reading was semantically infelicitous, suggesting a similar ‘contamination effect’ as with the colour adjectives. To put it another way, in both cases, the *-s*-less variant is ‘primed’ by another, etymologically unrelated, but superficially analogous construction. This effect ties in with the idea that speakers often act on fortuitous similarities, creating ‘local generalisations’ (see Joseph, 1992; Enger, 2013; Van de Velde and van der Horst, 2013; Van de Velde and Weerman, 2014).

- (15) *Hey heeft nl vandaag ook gevoetbalt ofzo ?? <TiredV> ik zie*
 Hey has nl today also played_football or_so <TiredV> I see
zoveel oranje [Moroccorp]
 so_much orange
 ‘Hey, have the Netherlands also played football today? <TiredV> I see so many orange things.’ [Q + ADJ]
 ‘Hey, did the Netherlands also play football today? <TiredV> I see so much orange’ [Q + N]

- (16) <katertje> *ik heb toch nog iets goed gedaan vandaag*
 <katertje> I have still yet something good done today [ConDiv]

‘<katertje> at least, I’ve done something good today.’
 [VP do [object something good]]
 ‘<katertje> at least, I’ve done something correctly today.’
 [VP do [Object something] in a good way]

The constraint in (14c) is a more general effect that standard norms are more strictly adhered to in formal written registers and (14d) shows the well-known Conserving Effect (Bybee, 2006; Bybee and Beckner, 2010) that applies cross-linguistically and states that more frequent forms are better shielded against morphological change. If Van der Horst (2008, pp. 1624-1625) is right in assuming that the -s may be losing ground in the north, the constraints in (14c) and (14d) point to an ongoing change ‘from below’.⁹ Constraint (14e) is a so-called interaction effect: there is a difference between the quantifiers, but it only plays in Flanders.

1.3 L2 speakers, non-transparent morphology, and the Dutch partitive genitive

The question we address in this paper is what early L2 or 2L1 speakers of Dutch do with the morphological quirk of the Dutch partitive genitive.¹⁰ From earlier studies, we know that early L2 or 2L1 speakers speed up refunctionalisation in adjectival inflection, as argued above (see also Van de Velde and Weerman 2014 and references cited there).¹¹ Early L2 speakers increase the transparency of adjectival -ə by loosening the unmotivated constraint in (13c). But what happens with adjectival -s? Do early L2 speakers similarly experience difficulties with the constraints on the partitive construction, or is the suffix transparent enough to let early L2 speakers pick up on this morphological signal?

We will investigate to what extent early L2 speakers differ from L1 speakers with regard to the partitive genitive -s by applying the method described in Gries and Deshors (2014). To this end, we make use of two complementary corpora: the internet chat relay part of the ConDiv Corpus (Grondelaers et al., 2000) and Moroccorp, a corpus of the Moroccan ethnolect variety of Dutch, as used by adolescent Moroccan early L2 and 2L1 speakers of Dutch (Ruetten and Van de Velde, 2013).

This study has four foci, two specifically related to Dutch and two more general ones. On a more specific level, this study first fits in the research on the language use of multilingual youngsters, in particular street language and ethnolect and its (potential) influence on the future of Dutch, as discussed in Cornips and Rooij (2003). Second, it is an extension of a quantitative study of the partitive genitive in Dutch commenced in Pijpots and

Van de Velde (2014), and will further flesh out the description of the partitive genitive in present-day Dutch grammar. On a more general level, this article can be seen as a concrete study of deflexion and the role of second language learners or language contact in this process, tying in with recent findings on the relation between demography and language change (Lupyan and Dale, 2010; Bentz and Winter, 2013). Lastly, in comparing Moroccorp to a ‘standard’ in the form of a subcorpus of ConDiv, the corpus study reported here aspires to contribute to the methodology of Granger’s (1996) Contrastive Interlanguage Analysis (CIA). This framework aims to compare a native language to an ‘interlanguage’, by making use of quantitative analyses of comparable corpora. Its goal is to uncover “factors of foreign-soundingness” (Granger, 1996, p. 43) by investigating in what way the interlanguage corpus differs from the native language corpus, mostly with respect to a specific phenomenon, e.g. the use of adverbial and adnominal participles by learners of English with a French and Dutch background (Cosme, 2008; Gries and Deshors, 2014, p. 110). More specifically, this paper hopes to answer Gries and Deshors’ (2014) call for the use of more sophisticated analytical approaches in this research tradition, by applying the regression-based methodology described in their article. In fact, the first step in the methodological procedure proposed by Gries and Deshors (2014) has already been completed in Pijpops and Van de Velde (2014). This step is needed to assess which factors determine a particular linguistic choice by native speakers (Gries and Deshors, 2014, p. 111). As explained below, the Moroccorp and ConDiv corpora are ideally suited for such contrastive analyses; while the present investigation zooms in on the partitive genitive, these corpora can be used to apply this methodology to any type of linguistic alternation.

Our concrete research question is the following:

Do the language users in Moroccorp differ from the language users in the ConDiv chat corpus in their realisation of the partitive genitive alternation, either in absolute numbers (e.g. less partitives or more -s omission in Moroccorp) and/or in the number and/or choice of factors determining the alternation?

We have four competing scenarios:

- a. Moroccorp language users consistently use the [+s] variant, simplifying the rule by generalising the most frequent variant (‘the hyper-Netherlandic *overgeneralisation* option’). This would mean that adjectival -ə and -s inflection are treated similarly. Like the adjectival -ə ending, the

- s would be an integral part of the grammar of Moroccorp language users and would arguably have experienced an increase in transparency.
- b. Moroccorp language users consistently use the [-Ø] variant, simplifying the rule by deleting the superfluous morpheme in an unproductive case system ('the English *deflexion* option'). This would suggest that the partitive -s is a vulnerable victim in an ongoing deflexion trend, and does not carry a transparent function that can be easily picked up by early L2 language users of Dutch.
 - c. Moroccorp language users do not differ from their ConDiv L1 peers ('the constraint-sensitive *no-difference* option'). Firstly, this can be taken to indicate that Moroccorp language users are quite capable of picking up the exact construction in which to apply the -s ending. Secondly, it would suggest that unlike the factors determining the adjectival -Ø/-ə alternation in L1 Dutch, those determining -Ø/-s alternation operate at a level which is not readily bypassed by Moroccorp language users.
 - d. Moroccorp users differ from their ConDiv L1 peers, in using more or other constraints ('the hyper-sensitive *aemulatio* option'). They might for instance use a wider range of constraints on the realisation of the -s, as is done by southern L1 speakers, for whom there is a differentiation between the quantifiers (see 14e, above).

A difference in the number and/or choice of factors determining the alternation (i.e., scenario d) would be of particular interest to the CIA analysis of Moroccan Dutch. In which cases do these speakers have more trouble with the -s suffix and why? Is the multifactorial grammar behind the partitive genitive different in Moroccan Dutch? To answer this question, we will have to look into the interactions in the regression model (Gries and Deshors, 2014, pp. 120-126).

This article is structured as follows. Section 2 introduces the corpora used in this study (Moroccorp and ConDiv), explains why they are suited for the type of research envisaged here, and gives information on the extraction and analysis of the data. In Section 3, we carry out a regression analysis on these data and discuss the results in the perspective of the research question posed above. Section 4 summarizes the conclusions.

2 Corpora and data

2.1 Corpora

The data used in this corpus study have been extracted from Moroccorp and the subcorpus of ConDiv containing the Netherlandic chat material.¹² ConDiv is a lectally stratified corpus of Netherlandic and Flemish Dutch (Grondelaers et al., 2000). The subcorpus we have used here amounts to roughly 7 million tokens. Moroccorp (see Ruetten and Van de Velde, 2013) is a 10 million token corpus compiled to reflect Dutch as used by speakers of second or third generation Moroccan immigrants, who speak an ethnolectal variety of Dutch. Their language use exhibits a number of characteristics typically associated with early L2 or 2L1 varieties of Dutch, like the well-known *Ausgleich* of Dutch attributive adjectival inflection in indefinite singular neuter NPs (see 13c, above). The advantage of these corpora is that they are comparable in size and register,¹³ which is one of the most important requirements of the methodological approach followed here (Gries and Deshors, 2014, p. 110), but also that the results we get for the adjectival partitive *-s* can be compared to the findings on adjectival *-ə* in Van de Velde and Weerman (2014), who used the same two corpora.

The ConDiv corpus is regionally stratified, and contains both a Netherlandic chat component and a Flemish chat component. We only used the Netherlandic component, as Moroccorp only contains Netherlandic ethnolect speakers. Including the Flemish ConDiv chat component might therefore compromise the comparability of both corpora. Indeed, it has already been established that there exist important differences between the partitive genitive alternation in Flanders and in the Netherlands (Pijpops and Van de Velde, 2014).

As our dataset will thus be entirely comprised of chat material, this particular form of language use requires some further specifications. To use the words of Koch and Österreicher (2007, p. 359):

Der Chat ist sogar eines der schönsten Beispiele dafür, dass im graphischen Medium eine relative, natürlich immer limitierte Annäherung an dialogische, spontane Nähesprachlichkeit möglich ist.

While on the medial dichotomy, chat is of course part of *graphic*, i.e. written, language use, its place on the conceptual continuum from written to spoken is more on the side of spoken language (Koch and Österreicher, 2007; Söll, 1974; Grondelaers et al., 2000, p. 358). Still, Koch and Österreicher (2007, p. 359), as well as Ágel and Hennig (2007, pp. 202, 206-214) point

to a possible inhibiting influence of the graphic medium on language use. As for manners, however, its written and physically distant nature may also generate a disinhibiting effect (Suler 2004). That is, as can be inferred from a quick glance at the comments section of a newspaper's website, people may feel freer to speak their mind or to use ruder language from behind a computer screen – especially if anonymity is guaranteed – than would be considered acceptable in face-to-face conversations.

Not all types of chat conversations can be seen as part of the same discourse tradition, or as positioned on the same place in the conceptual continuum from spoken to written language. A chat conversation between a software technician and a customer experiencing a programme malfunction will be of a different nature than one between brother and sister. The chatlogs gathered in our corpora contain language from youngsters who typically only know one another from the chat box. To further characterise the conversational setting, we make use of Koch and Österreicher (2007, p. 351) communicative parameters:

- **Privatheit/Öffentlichkeit:** Anyone can join, store and even publish the chat conversations – as has been done in the compilation of the corpora – and the chatters are sometimes explicitly aware of this (Ruetten and Van de Velde, 2013, p. 461, 464). Still, the chatters mostly assume they are amongst like-minded peers – they can see who is 'listening in' and, after all, their identity is protected by anonymous nicknames. As such, their conversations are often of a private nature.
- **Vertrautheit/Fremdheit der Kommunikationspartner:** While the chatters mostly do not know one another in real life, this is not to say that they are necessarily strangers. Some are regular visitors of the chat box, and recognise each other by their nicknames.
- **Starke/geringe emotionale Beteiligung:** personal problems and (strong) personal opinions are often discussed. Emotional participation is rather strong (cf. Suler's 2004 Online Disinhibition Effect).
- **Situations- und Handlungseinbindung/-entbindung:** a brief look at the corpus material shows that the discussed subjects are mostly situational and act-oriented matters, rather than abstract ideas.
- **Referenzielle Nähe/Distanz:** Referential distance.
- **Raum-zeitliche Nähe/Distanz:** Spatio-temporal distance.
- **Kommunikative Kooperation/keine komm. Koop.:** Communicative cooperation.
- **Dialogizität/Monologizität:** Multiple dialogs are often held simultaneously.
- **Spontaneität/Reflektiertheit:** Spontaneity.

- **Freie Themenentwicklung/Themenfixierung:** Free subject development.

These parameters accumulate to a conversational setting of considerable communicative proximity (Koch and Österreicher, 2007, p. 351). As for language material that can easily be gathered in digital corpora on a large scale, this kind of data approximates ‘natural’ spoken language use fairly well. This is important, as normative works explicitly repudiate the variant without *-s* ending (Taaladvies.net, Taaltelefoon.be, Taalnet) and we know from earlier work that this norm affects formal language use – although even there, it is not strictly adhered to (Pijpops and Van de Velde, 2014).

2.2 Data

The data from the Netherlandic chat corpus could be reused from an earlier study of the partitive genitive in Pijpops and Van de Velde (2014). The Moroccorp data, however, had to be extracted from scratch, and because Moroccorp is not syntactically annotated they had to be manually checked in order to ensure that the occurrences gathered were in fact genuine partitive genitives.¹⁴

All instances in which one of the following quantifiers preceded one of the following adjectives, with or without *-s* suffix, were extracted from Moroccorp and manually checked. The extraction was done using the freely available *AntConc* software (Anthony, 2011).

- Quantifiers: *iets* (‘something’), *niets* (‘nothing’), *wat* (‘something’), *veel* (‘a lot’), *zoveel* (‘so much’)
- Adjectives: *aardig* (‘nice’), *apart* (‘apart’), *belangrijk* (‘important’), *beter* (‘better’), *bijzonder* (‘particular’), *blauw* (‘blue’), *concreet* (‘concrete’), *deftig* (‘decent’), *dergelijk* (‘similar’), *erg* (‘awful’), *geel* (‘yellow’), *gek* (‘crazy’), *goed* (‘good’), *groen* (‘green’), *interessant* (‘interesting’), *klein* (‘small’), *lekker* (‘tasty’), *leuk* (‘fun’), *mooi* (‘beautiful’), *nieuw* (‘new’), *nuttig* (‘useful’), *oranje* (‘orange’), *positief* (‘positive’), *purper* (‘purple’), *raar* (‘weird’), *rood* (‘red’), *spannend* (‘exciting’), *speciaal* (‘special’), *verkeerd* (‘wrong’), *verschrikkelijk* (‘horrible’), *vreemd* (‘weird’), *warm* (‘warm’), *wit* (‘white’), *zinnig* (‘sensible’), *zwart* (‘black’)

For the selection of the adjectives, we used three criteria:

- The adjective had to occur at least seven times in post-quantifier position in the Corpus of Spoken Dutch (CGN) (van Eerten, 2007). The reason why we resorted to another corpus for this selection criterion

is that we needed a PoS-tagged corpus to ensure we included all the relevant attestations. To the resulting set, we added the major colour adjectives, if they did not violate the next criterion, because we have a special interest in them (see 14a).

- Adjectives ending in a (post-)alveolar fricative were excluded, because the presence of a partitive *-s* is phonologically indiscernible. This is the reason why the list above includes *purper* ('purple'), but not the more common colour term *paars* ('purple'), for instance.
- Adjectives that are homonymous with plurals of nouns were excluded as well. Cases in point are *ouders* ('elder-GEN' or 'parents') and *extra's* ('extra-GEN' or 'bonuses').

The selection of the quantifiers is based on the lists in Haeseryn et al. (1997, p. 356, 432), applying a threshold of at least one occurrence in both corpora of the combination of the quantifier and one of the selected adjectives. Note that our dataset does not contain the quantifier *weinig* ('few'). The hits of *weinig* were originally extracted from Moroccorp, but it was later found that the entire Moroccorp and ConDiv dataset contained no occurrences of *weinig* without *-s* suffix. This leads to problems with the estimates in the regression model. A total of 11 hits of *weinig* were removed from the dataset, leaving us with a dataset of 2378 partitive genitives occurrences in total.

These datapoints are used to carry out a logistic regression analysis. This statistical technique measures the effect of one or more explanatory variables, or 'predictors', on a binary response variable (see Baayen, 2008; Gries, 2013; Speelman, 2014). In the case at hand, we want to predict under what circumstances the partitive *-s* suffix is realised or not. Instead of testing each of the relevant predictors (to be introduced below) bivariate, by means of an association test like a Chi-Square or Fisher's Exact test, a multiple logistic regression allows one to get an idea of the impact of a predictor while controlling for all other predictors. Such an approach is statistically superior, as separate bivariate testing can yield problematic results, for instance by ignoring 'interaction effects', when the effect of one explanatory variable differs depending on the levels of another explanatory variable.

Technically, what regression does is minimizing the distances between observed values in a Cartesian *n*-dimensional space and a so-called 'hyperplane'. The geometric properties of this hyperplane are 'fitted' to the data. A crucial step in this fitting process is to decide on the number of predictors to be included in the regression model. The more predictors we add to

the model, the more accurately we can predict the values of our observations, but regressing means finding a balance between having enough predictors to make reasonable predictions on the one hand, and ‘overfitting’ the model, with a concomitant lack in the summarising power. To find the optimal balance, we take two measures.

The first measure is to make use of a so-called ‘mixed-effect’ logistic regression model. Such models make a distinction between ‘fixed effects’ – the predictors you are actually interested in – and ‘random effects’ – predictors that can be assumed to have an actual effect, but are not of interest, and the levels of which vary randomly if you were to redo the analysis on a different dataset. Technically, random effects assume different slopes and/or intercepts of each of the 1-dimensional regression lines that together make up the hyperplane. In our study, we use random effects with different intercepts.

The second measure for finding the optimal balance is using a bidirectional step-wise procedure for variable selection. This procedure involves introducing and dropping variables in the model and seeing how this affects the Akaike Information Criterion (AIC), a measure allowing for quality comparison between models. The lower the AIC, the better the model fits the data.

The analyses have been carried out with the aid of the open-source software R (R Core Team, 2013).¹⁵ The variables included in this procedure are listed below.

Response variable:

– -s: *with, without*

Explanatory variables:

- Corpus: *ConDiv, Moroccorp*
- Quantifier: *iets* (‘something’), *niets* (‘nothing’), *wat* (‘something’), *veel* (‘a lot’), *zoveel* (‘so much’)
- Type-Adjective: *other, deviant, colour*
- Length-Adjective: *1, 2, 3, 4*
- Number-of-words-AP: *1, 2*
- Frequency: log-transformed frequency of the phrase
- Phrase: *iets leuk(s)* (‘something fun’), *niets zinnig(s)* (‘nothing sensible’), *weinig concreet(s)* (‘few concrete things’),...

The set-up of the present study requires a slightly different set of predictor variables than the one in Pijpops and Van de Velde (2014). Four adjustments were necessary. First, the variable *Corpus* has of course been added, which distinguishes between the ConDiv subcorpus of Netherlandic chat language and Moroccorp material. In the present study, this explanatory variable carries major theoretical weight. Second, the variable *Variety*, distinguishing between the Netherlands and Flanders was removed, for reasons discussed above. Third, the variable *Register* was dropped. This variable exploited the register stratification of ConDiv, but could obviously not be retained in a model that exclusively looks at chat data. In Moroccorp we have no counterpart for the more formal registers in ConDiv. Fourth, we have used base 10 for the logarithmic transformation of frequency, as it is more easily interpretable than the base we used in Pijpops and Van de Velde (2014). This does not make a real difference for the model as such, as different bases only differ by multiplication by a constant (Fox and Weisberg, 2011, p. 127).

The rest of the variables are copied from the Pijpops and Van de Velde (2014) study:

- *Type-Adjective* makes a distinction between (i) colour adjectives, (ii) the adjectives *beter* ('better'), *goed* ('good'), *fout* ('incorrect') and *verkeerd* ('wrong'), which are here called 'deviant adjectives' for terminological convenience, and (iii) all the other adjectives. The idea is that (i) and (ii) were assumed to display higher rates of -s drop, for reasons stated above, see (14a-b).
- The variable *Length-Adjective* indicates the phonological weight of the adjective, expressed as the number of syllables. Note that the adjective *interessant* ('interesting') is counted as a three-syllable word, as the first schwa is often syncopated.
- *Number-of-words-AP* distinguishes between single-word adjectives and adjectives that are premodified by a degree adverb, e.g. (17) and (18).
- *Frequency* is the logarithmically transformed frequency of the phrase type – phrase type being the unique combination of a Q + ADJ pair. The uniqueness disregards the presence of the -s suffix, so that *iets moois* ('something beautiful-GEN) and *iets mooi* ('something beautiful-Ø') belong to the same phrase type. It does however take into account whether or not the adjective is premodified by a degree adverb, so that *iets beter(s)* ('something better(-GEN)') and *iets veel beter(s)* ('something much better(-GEN)') do belong to different phrase types.
- The variable *Phrase* has all these distinct phrase types as individual levels.

- (17) <engeltje> dag: *liever iets lekker warm*:) <dag> ok,
 <engeltje> dag preferably something tasty warm <dag> ok
er is melk, koffie, nesquick... [Condiv]
 there is milk coffee nesquick
 ‘<engeltje> dag: I’d prefer something nice and hot. <dag> ok, there is milk,
 coffee, hot chocolate milk ...’
- (18) *ik heb die van Utrecht laten lopen, maar geloof me ik heb*
 I have that_one of Utrecht let run but believe me I have
iets veel beter-s nu [Moroccorp]
 something much better-GEN now
 ‘I let the one from Utrecht go, but believe me, I’ve got something much better
 now.’

The variable *Phrase* will be implemented as a random effect in the regression models, whereas all other explanatory variables will be entered as fixed effects into the variable selection procedure. *Corpus*, *Quantifier* and *Type-Adjective* are categorical variables, *Length-Adjective*, *Number-of-words-AP* and *Frequency* numeric ones.

3 Results and discussion

3.1 Results

As said in the previous section, we analysed the dataset by means of mixed-effects (multiple) logistic regression modelling, more specifically, by using the procedure proposed by Gries and Deshors (2014, pp. 122-136). Our method does, however, differ from their example in one aspect. Whereas Gries and Deshors (2014) only entered interactions between the fixed effect *Corpus* and the other fixed effects in the variable selection procedure, we included all possible two-way interactions between any of the fixed effects under scrutiny.

The model that we arrived at was subjected to a number of additional diagnostics (see Speelman 2014). First, all predictors which did not significantly improve the model were dropped. Second, we tested whether the residual deviance was not much larger than the degrees of freedom. Large residual deviance signals overdispersion, suggesting that the data behave too heterogeneously to have confidence in the model. Third, we carried out a Hosmer-Lemeshow-Cessie goodness-of-fit test. If the test result is significant, the model does not fit well, for instance because there are important

predictors missing from it (Speelman, 2014). Fifth, we checked whether all Variance Inflation Factors were below 4, to make sure the model does not run into the problem of multicollinearity, meaning that several variables measure the same thing – obviously something one wants to avoid in a maximally parsimonious regression model. None of the diagnostics yielded problematic results, so, in a final step, we added the random effect *Phrase* and once again removed all predictors which no longer made a significant contribution to the model's quality.

This left us with the model presented in Table 1. As can be appreciated, the number of parameters is well below the number of observations of the response variable's least frequent level divided by twenty, which is another rule-of-thumb in logistic regression modelling. The C-index gives an idea of the overall quality of the model. Values above 0.80 are considered as satisfactory. With 0.8420, we have a powerful regression model. The predictors are ordered from most to least important.

Note that the categorical variables were implemented using dummy coding. This means that one level is taken as the reference level, and the others are used as separate regressors. The numeric variables could be implemented directly. All estimates and confidence intervals of the estimates are rounded off to 2 decimals, the p-values are rounded off to 4 decimals.

The estimates give an indication of the effect size. Because the variant without -s is the success level of the response variable, a positive value for the estimate of a numeric predictor means the probability of -s drop rises as the value of the variable increases. A positive value for the estimate of one of the levels of a categorical variable means that there is more -s drop than in the reference level. Conversely, a negative sign means that the level in question favours -s retention. The higher the absolute value, the more severe the impact is. A value of 4.34 for colour adjectives, for instance, means that the logit of the -s drop is 1.28 (4.34 plus the intercept -3.06).

The logit is a (double-)transformed measure of the probability, and the corresponding probability can be computed by taking inverse logit function of the estimates. For the adjectives, this means that colour adjectives have a 'partial' estimated probability of 0.78.

A more human-friendly visualisation of the results is given in the 'effect plots' (Fox 2003) in Figure 1, which directly give the estimated probability of -s drop and a corresponding confidence interval.¹⁶ The 'bar code' on the x-axis on the right side is a so-called 'rug', indicating the marginal distribution of the Frequency observations. Each bar stands for one phrase type.

— AIC:	1240
— C-index:	0.842
— Number of phrases:	96
— Total number of hits:	2378
— Hits with -s:	2143
— Hits without -s:	235

Table 1 Regression model predicting -s drop

Predictors	Levels	Estimates	Confidence intervals		P-values
			2,5%	97,5%	
	intercept	-3.06	-4.32	-1.80	< 0.0001
Type-Adjective	<i>other</i>	Reference level			
	<i>deviant</i>	1.78	1.07	2.48	< 0.0001
	<i>colour</i>	4.34	2.71	5.97	< 0.0001
Frequency		0.14	-0.61	0.89	0.7136
Corpus	<i>ConDiv</i>	Reference level			
	<i>Moroccorp</i>	0.98	-0.33	2.28	0.1416
Interaction Frequency - Corpus	<i>ConDiv</i> <i>Moroccorp</i>	Reference level -0.87	-1.59	-0.15	0.0175

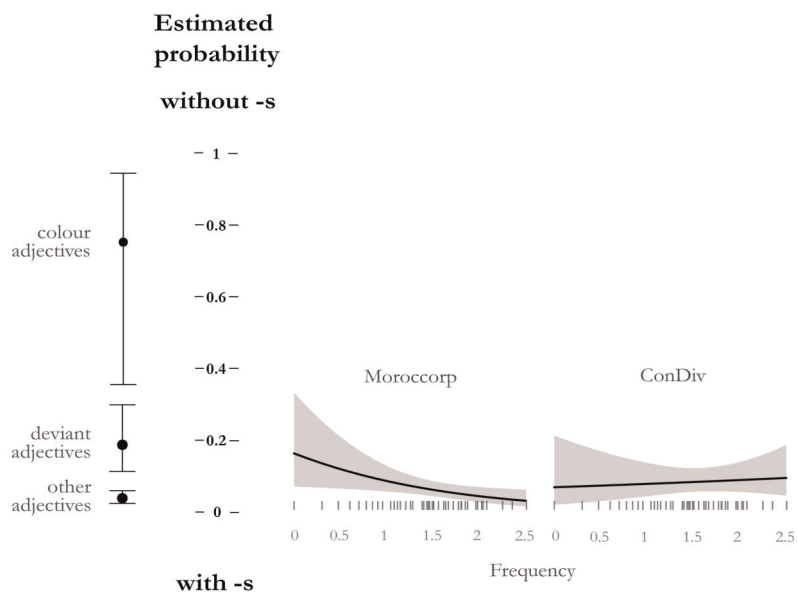


Figure 1 Probabilities, estimated by the regression model. The influence of Type-Adjective is strong and stable across both corpora. The influence of Frequency seems to differ across the corpora, at first sight.

Our most important predictor is *Type-Adjective*, followed by *Frequency* and *Corpus*. The model also contains an interaction between *Frequency* and *Corpus*. This interaction seems to indicate that the Moroccorp chatters tend more towards -s omission in the low frequency phrases than the Netherlandic chatters of ConDiv. However, it would be inadvisable to make sweeping conclusions on the basis of this effect. First, exactly because the infrequent phrases are infrequent, there is a lot of uncertainty about their behaviour, as can be seen in the large confidence intervals in the low frequencies in Figure 1. Second, as can be seen in Figure 2, over two thirds of our dataset is made up of Moroccorp material. This means that the calculated frequencies are more strongly influenced by the Moroccorp material than the ConDiv material, which may explain why we only find the expected frequency effect in the Moroccorp data. Finally, as explained below, neither *Frequency* nor *Corpus* seems vital to the overall model.

	Moroccorp	ConDiv
without -s	149	86
with -s	1464	679

Figure 2 Partitive genitives in the Moroccorp and ConDiv corpora

As can be seen in Figure 1 and Figure 2, there is hardly any difference between the Moroccorp and ConDiv material regarding a general preference for -s omission. In fact, -s retention is even slightly more frequent in Moroccorp, though this difference is not significant, neither in the regression model (see Table 1) nor in a bivariate chi-squared test ($p = 0.126$). Also, with regard to the factors determining the alternation, Moroccorp and ConDiv are very similar. The regression model contains only a single interaction with *Corpus*, which, as stated above, does not seem to be of crucial importance. If we were to remove not only this interaction from the model, but also the main effects of *Corpus* and *Frequency*, retaining only the random effect *Phrase* and the fixed effect *Type-Adjective*, the model's predic-

tive quality is even slightly raised, to a new C-index of 0.844.¹⁷ The high predictive quality of this last model at the cost of so few predictors is perhaps what is most surprising about our analysis, and another testimony to the power of the *Type-Adjective* predictor (cf. Pijpops and Van de Velde, 2014, p. 18).

It appears then, that Moroccorp users are equally sensitive to the *Type-Adjective* constraint as the ConDiv users. This is confirmed if we measure how well a ConDiv-only model can predict the Moroccorp observations. The reasoning behind this procedure is as follows: let's first build a bivariate logistic regression model with *Type-Adjective* as the only fixed effect based on the ConDiv data only. Then we fit this ConDiv model on the Moroccorp data. If the Moroccorp data are highly comparable to the ConDiv data, the model should yield a good fit, and the differences between the estimated values of the model and the observed values should be small. This is indeed what we find. The mean estimated difference is only off by 0.04.

In theory, there is a possibility that the similarity between ConDiv and Moroccorp is overestimated: Moroccorp users may avoid the partitive genitive construction when in doubt, and only use it in straightforward cases.¹⁸ We tested this hypothesis in two ways. First, we checked whether Moroccorp has fewer instances of the partitive genitive per 10,000 tokens. This is emphatically not the case: ConDiv has 1.10 instances per 10,000 tokens, while Moroccorp has 1.57 instances per 10,000 tokens. So if anything, Moroccorp chatters use the construction *more* than ConDiv chatters. Second, we looked at whether there is more variation in the phrase types used in either of the corpora. Moroccorp has more different phrase types than ConDiv (86 vs. 67), but of course, we have to correct for the fact that Moroccorp is 47% larger than ConDiv. Correcting is not so easy in this case. It does not make sense to just calculate the number of phrase types per 10,000 tokens, because type accumulation peters out as the corpus size grows, due to the Zipfian distribution of phrase types (see Bentz et al. 2014). To correct for this, we divided the number of phrase types by the logarithm of the corpus size. Again, Moroccorp users outperform the ConDiv chatters in using the construction with more different phrase types.

3.2 Discussion

We have found there to be hardly any difference between the Moroccorp and ConDiv corpora concerning the partitive genitive alternation. Also, both corpora exhibit a variance-suffocating dominance of nearly 90% of the -s retaining variant. Because of this, we believe there is little reason to

continue along the path sketched out by Gries and Deshors (2014, pp. 126-131) by applying what they call the ‘MuPDAR’ approach to the partitive genitive alternation in Moroccorp and ConDiv. This MuPDAR analysis (‘Multifactorial Prediction and Deviation Analysis with Regressions’) is a statistical analysis of the deviation in the choices the two varieties make for a certain linguistic response variable. Instead, we can immediately turn to answering the research question.

The partitive *-s* realisation of the Moroccorp language users does not differ from that of the Netherlandic native speakers of the ConDiv chat corpus, or even from the Dutch written standard language (if we take into account the results of Pijpops and Van de Velde 2014, pp. 19-20, 23), neither in absolute numbers (e.g. more *-s* omission in Moroccorp), nor in the number and/or choice of factors determining the alternation. Moreover, Moroccorp chatters do not use the construction less frequently, or in a lexically less varied way. Especially striking is the strong influence of *Type-Adjective* in both ConDiv and Moroccorp. With regard to the four scenarios set out in Section 1.3, our data suggest that the third scenario, the so-called ‘constraint-sensitive *no-difference* option’, is the correct one.

This entails that the Moroccorp language users do not simplify the adjectival $-\emptyset/-s$ alternation, by overgeneralising the *-s* inflection, i.e. Scenario (a) in Section 1.3., in contrast to what is the case with the $-\emptyset/-\emptyset$ alternation (see Van de Velde and Weerman, 2014 for several case studies of adjectival inflection patterns, where Morrocorp and ConDiv do differ). This suggests that the factors governing the $-\emptyset/-s$ alternation, most notably *Type-Adjective*, are of a different nature than those governing the $-\emptyset/-\emptyset$ alternation. (Early) L2 language users appear equally responsive as L1 users to links between superficially resembling, yet structurally unrelated constructions (see the explanation of constraints (14a) and (14b) in Section 1.2).¹⁹

Furthermore, contrary to what could have been expected on the basis of morphological difficulties in L2 language use, Moroccorp language users do not wash away the partitive *-s* in the waves of ongoing deflexion either, i.e. Scenario (b) in Section 1.3. If Moroccorp chatters are statistically more likely to jettison non-transparent morphology, then our results could be taken as an indication that the *-s* suffix is not so ‘odd and quirky’ after all. In our view, the transparency of the partitive *-s* derives from the fact that it has a construction-marking function.

The partitive genitive is best seen as a construction, in the sense of Construction Morphology (see Booij, 2010a,b for an extensive motivation for the need of a constructional approach to morphology), and the *-s* is

interpreted as a partitive ending only when it occurs in the construction at issue. In other words: it derives its function from the construction it occurs in. This is why Booij (2010a, pp. 211-236) speaks of ‘construction-dependent morphology’. In itself, the partitive genitive *-s* is ill-motivated, as it does not form part of a case paradigm. Dutch has largely lost its case system, and the refunctionalised *-s* of the partitive genitive is found on adjectives, not on nouns. In isolation the inflectional *-s* does not signify partitivity; it needs the constructional template to be interpreted felicitously. Following Booij (2010a, p. 227), we can use the formalisation in (19) to capture the constructional meaning, in which the formal part is specified on the left-hand of the double arrow, and the function part is specified on the right-hand side.

(19) $[NP_i [\dots [X-s]_A]_{APj}]_{NPK} \leftrightarrow [Quantity_i \text{ with Property}]_k$

In a construction grammar approach, the *-s* is not necessarily quirky. If it surfaces across the board as soon as the constructional template occurs, it is more transparent than the rule in (13), with its intricate grammatical condition in (13c). Admittedly, the regression model shows that there are conditions on (19) as well, but as explained in Section 1.2, these conditions are not grammatical, but semantic and pragmatic in nature, and derive from contamination from superficially resembling constructions.²⁰

4 Conclusions

We conclude this article by summarizing what the results presented here mean for the various ‘tracks’ on which this study is situated. As for the research into the Dutch language use of early L2 ethnolects, it is remarkable how good the ethnolect language users are at adopting what at first sight can only be considered – in comparison to related languages – a weird quirk in Dutch grammar. It may prove to be a fruitful undertaking to further investigate exactly for which linguistic phenomena these language users do diverge from ‘mainstream’ language use, and for which phenomena they do not. For instance, do they prove to be more creative with contextual-inflectional, inherent-inflectional, derivational or syntactic processes (see Booij, 1996 for the first two terms)?

Focusing on the partitive genitive, we think this study can be taken as evidence that the partitive genitive has good prospects for survival, at least in the Netherlands. Grand-scale *-s* omission seems at present limited to Belgium. In the Netherlands, though it is not supported by a productive

case system, the suffix survives in a specific constructional niche. This has been observed for other remnants of the genitive case as well by Hoeksema (1998b). In fact, it is not unusual for old constructions to revive in specific corners of the grammar (Van de Velde 2015).

Finally, we hold the present study to be of relevance to CIA-research. Although the findings can be considered null results in terms of a CIA-analysis, our study still shows the applicability and feasibility of the method of Gries and Deshors (2014) on morphological alternations, just as they have already shown it for lexico-syntactic alternations. In this respect, the present study contributes to the methodology of the field of Dutch language variation.

We hope that this study will add to the understanding of the fascinating inflectional quirk of the Dutch language that is the partitive genitive. Its future might turn out to be somewhat brighter than is sometimes assumed (van der Horst, 2008, pp. 1624-1625), and its quiriness seems to be less of a problem for Moroccorp chatters than for Dutch linguists.

Notes

1. This article has profited from the comments by two reviewers, as well as by Timothy Coleman, whom we would all like to thank. The research carried out was supported by a BOF research grant from the University of Leuven and a fellowship from the Research Foundation Flanders (FWO).
2. Although historically, the partitive genitive *-s* is clearly part of the Dutch case inflection system, it may be synchronically more cautious to consider it an isolated irregular suffix, as the case inflection system has long collapsed and its debris is scattered around in Dutch grammar. In the glosses, the *-s* has been marked *-GEN*.
3. This is sometimes seen as an argument to assume that the quantifier is no longer the head of the construction, but the adjective is (Van Marle 1996, pp. 73, 80).
4. To illustrate the notion of exaptation, consider the fate of negation in some dialects of Dutch. Where an erstwhile negation marker, *en* was reanalyzed into a marker of subordination. The clitic *en* lost its transparency when negation *niet* took over in the Dutch Jespersen Cycle. In the gradual decline of *en*, relic attestations were more frequent in subordinate clauses, and this created the conditions for a reanalysis in which *en* was seen as a subordinator (see Van der Auwera 2012, pp. 413, with reference to earlier work by Overdiep and Neuckermans). In example (i), the *en* historically derives from a negation particle, but it occurs in subordinate clauses with positive polarity. (i)

Toen we bij de poort en kwamme ...
 when we at the gate SUBORDINATOR came
 'when we arrived at the gate'

5. The situation is more complex, as the *-en* affix is also used for the infinitive and many varieties of Dutch have apocope of the *n* in speech. The old 2PL *-t* ending is attested until the 20th century (b.v. *jullie gokt* 'you gamble', WNT s.v. wereld) and is still present

in some dialects. For a more in-depth theory of transparency in morphology, see among others Leufkens (2015).

6. In the southern spoken Dutch variety of Flanders, the condition in (13c) is slightly different: definiteness does not play a role.
7. In actual fact, the account in Van de Velde and Weerman (2014) is more complicated. They show that in premodifying position the *-ə* suffix is used to demarcate the determiner from the adjective. The gist of the account is the same as what is reported here, however: the non-transparent condition in (13c) is replaced by a more transparent one, which amounts to refunctionalisation / exaptation.
8. The bracketed representation is meant as an approximate formulation. No theoretical significance should be attributed to it.
9. The actual history is more complicated. The *-s* was not consistently expressed in earlier stages of Dutch, so it is possible that it was introduced in the north by a change 'from above' first, in a more general tendency of reviving the case system in Early Modern Dutch, by copying the revered Latin language, and then later suffered from change from below, reintroducing the *-s-less* variant.
10. We are looking into informal Dutch produced by language users of Moroccan ethnicity. For convenience sake, we will refer to the variety under study as an '(early L2) ethnolect', though it may be more accurate to see it as a 2L1 variety.
11. Late L2 speakers are more likely to 'fossilise' the construction (see Matras 2009, pp. 75 for this term).
12. To avoid misunderstandings, we will use Dutch to refer to the Dutch language, including the Flemish variety, and Netherlandic to refer to Dutch spoken in the Netherlands, as others have done (cf. Geeraerts 2010).
13. There is a possible confound to the comparability, as Moroccorp was compiled about 15 years after ConDiv.
14. We are not aware of the existence of reliable taggers for Dutch chat material, which is exceptionally hard to annotate because of its fragmented nature. Here, the biggest advantage of chat data for linguists, i.e. its proximity to natural, spoken language, becomes a practical disadvantage.
15. For the analysis and visualisation, we made use of the MASS (Venables and Ripley, 2002), rms (Harrell, 2013), lme4 (Bates et al., 2013) and effects (Fox, 2003) packages.
16. Note that the fitted probabilities in the effect plots may slightly differ from the values based on the estimates in the model summary in Table 1 because the other variables are left 'as is' in the model, while the values in the effect plots adjust the value of the other variables to the mean.
17. Of course the AIC is raised as well, to a value of 1252, which is why this model was not selected by the variable selection procedure.
18. We would like to thank one of the anonymous reviewers for pointing out this possibility.
19. Or they are exceedingly good at picking up and reproducing lexical preference patterns of L1 speakers. A vast body of CIA-research, however, shows that this is exactly where L2 language users struggle most (see Granger 2004, pp. 132; Cosme 2008; Gries and Deshors 2014, among others).
20. An alternative explanation for the fact that Moroccorp differs from ConDiv in the expression of the adjectival schwa inflection, but does not differ in the expression of the partitive *-s* genitive, could be sought in the 'stylisation' function of these grammatical markers. Overgeneralisation of the adjectival schwa is a known shibboleth of Dutch ethnolects, and can accordingly be used to (semi-)consciously index the ingroup

speech, whereas partitive genitive -s is not. We think this is unlikely, as the findings in Van de Velde and Weerman (2014) show that there are very subtle effects in the refunctionalisation of the schwa in adnominal elements that go well beyond the (13c) condition on which the shibboleth hinges.

References

- Steven Paul Abney, *The English noun phrase in its sentential aspect*, Cambridge: Dissertation Massachusetts Institute of Technology, 1987.
- Vilmos Ágel and Mathilde Hennig, 'Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens [Considerations on the theory and practice of near and distant speaking]', In: Vilmos Ágel and Mathilde Hennig (red.), *Zugänge zur Grammatik der gesprochenen Sprache*. Tübingen: Niemeyer, 2007, pp. 179-214.
- Cynthia Allen, *Genitives in early English: typology and evidence*, Oxford: Oxford University Press, 2008.
- Laurence Anthony, *AntConc (Computer Software, version 3.3.3)*, Tokyo: Waseda University, 2011, <http://www.antlab.sci.waseda.ac.jp/>.
- Johan van der Auwera, 'Wat het Nederlands ons kan leren over de Jespersencyclus [What Dutch can teach us about the Jespersen's cycle]', *Nederlandse taalkunde* 17 (2012), pp. 403-413.
- Rolf Harald Baayen, *Analyzing linguistic data: a practical introduction to statistics using R*, Cambridge University Press, 2008.
- Douglas Bates, Martin Maechler, Ben Bolker and Steven Walker, *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4*, 2013, <http://cran.r-project.org/package=lme4>.
- Christian Bentz, Douwe Kiela, Feli Hill and Paula Buttery, 'Zipf's law and the grammar of languages: A quantitative study of Old and Modern English parallel texts', *Corpus Linguistics And Linguistic Theory* 10, 2 (2014), pp. 175-211.
- Christian Bentz and Bodo Winter, 'Languages with More Second Language Learners Tend to Lose Nominal Case', *Language Dynamics and Change* 3, 1 (2013), pp. 1-27.
- Geert Booij, 'Inherent versus contextual inflection and the split morphology hypothesis', In: Geert Booij and Jaap van Marle (red.), *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1996, pp. 1-16.
- Geert Booij, *Construction morphology*, Oxford: Oxford University Press, 2010a.
- Geert Booij, 'Construction morphology', *Language and Linguistics Compass* 4, 7 (2010b), pp. 543-555.
- Hans Broekhuis, *Syntax of Dutch: Adjectives and Adjective Phrases*, Amsterdam: Amsterdam University Press, 2013.
- Hans Broekhuis and Anke Strang, 'De partitieve genitiefconstructie [The partitive genitive construction]', *Nederlandse taalkunde* 1, 3 (1996), pp. 221-238.
- Joan Bybee, 'From Usage to Grammar: The Mind's Response to Repetition', *Language* 82, 4 (2006), pp. 711-733.
- Joan Bybee and Clay Beckner, 'Usage-based theory', In: Bernd Heine and Heiko Narrog (red.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2010, pp. 827-855.
- Leonie Cornips and Vincent de Rooij, 'Kijk, Levi's is een goeie merk: maar toch hadden ze 'm gedist van je schoenen doen 'm niet', In: Jan Stroop (red.), *Waar gaat het Nederlands naartoe? Panorama van een taal*. Amsterdam: Bert Bakker, 2003, pp. 131-142.
- Christelle Cosme, 'Participle clauses in learner English: the role of transfer', In: Gaëtanelle Gil-

- quin, Szilvia Papp and Mara Belén Díez-Bedmar (red.), *Linking up Contrastive and Learner Corpus Research*. Amsterdam/Atlanta: Rodopi, 2008, pp. 177-200.
- Laura van Eerten, 'Over het Corpus Gesproken Nederlands', *Nederlandse Taalkunde* 12, 3 (2007), pp. 194-215.
- Hans-Olav Enger, 'Morphological theory and grammaticalisation: the role of meaning and local generalisations', *Language Sciences* 36 (2013), pp. 18.
- John Fox, 'Effect displays in R for generalised linear models', *Journal of Statistical Software* 8 (2003), pp. 1-27.
- John Fox and Sanford Weisberg, *An R companion to applied regression*, 2nd edn. Los Angeles: Sage, 2011.
- Sylviane Granger, 'From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora', In: Karin Aijmer, Bengt Alsterlund and Mats Johansson (red.), *Languages in Contrast: Text-based Crosslinguistic Studies*. Lund: Lund University Press, 1996, pp. 37-51.
- Sylviane Granger, 'Computer learner corpus research: current status and future prospects', In: Ulla Connor and Thomas A. Upton (red.), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 2004, pp. 123-145.
- Stefan Thomas Gries, *Statistics for linguistics with R. A practical introduction*, 2nd edn. Berlin: De Gruyter, 2013.
- Stefan Thomas Gries and Sandra Deshors, 'Using regressions to explore deviations between corpus data and a standard/target: two suggestions', *Corpora* 9, 1 (2014), pp. 109-136.
- Stefan Grondelaers, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede and Dirk Speelman, 'Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]', *Nederlandse Taalkunde* 5, 4 (2000), pp. 356-363.
- Walter Haeseryn, Kirsten Romijn, Guido Geerts, Jaap de Rooij and Maarten van den Toorn, *Algemene Nederlandse Spraakkunst*, Groningen: Nijhoff, 1997.
- Wayne Harbert, *The Germanic languages*, Cambridge: Cambridge university press, 2007.
- Frank Harrell, 'rms: Regression Modeling Strategies. R package version 4.0-0', 2013, <http://cran.r-project.org/package=rms>.
- Jennifer Hendriks, 'Re-examining the "origins" of the prenominal periphrastic possessive construction Jan z'n boek in Dutch', *Diachronica* 29, 1 (2012), pp. 28-71.
- Jack Hoeksema, 'Adjectivale inflectie op -s: geen geval van transpositie [Adjectival inflection on -s: not a case of transposition]', In: Eric Hoekstra and Caroline Smits (red.), *Morfologiedagen 1996 [Morphology Days 1996]*. Amsterdam: P. J. Meertens-Instituut, 1998a, pp. 46-72.
- Jack Hoeksema, 'Een ondoode categorie: de genitief [An undead category: the genitive]', *Tabu* 28 (1998b), pp. 162-167.
- Joop van der Horst, *Geschiedenis van de Nederlandse syntaxis [History of Dutch syntax]*, Leuven: Universitaire Pers Leuven, 2008.
- Joop van der Horst, *Taal op drift. Lange-termijntwikkelingen in taal en samenleving [Language drift. Long term developments in language and society]*, Amsterdam: Meulenhoff, 2013.
- Joop van der Horst and Kees van der Horst, *Geschiedenis van het Nederlands in de twintigste eeuw [History of Dutch in the twentieth century]*, Den Haag: Sdu, 1999.
- Brian Joseph, 'Diachronic explanation: putting the speaker back into the picture', In: Garry W. Davis and Gregory K. Iverson (red.), *Explanations in historical linguistics*. Amsterdam: John Benjamins, 1992, pp. 123-144.
- Ellen-Petra Kester, 'The nature of adjectival inflection', Utrecht: LEU, 1996.
- Peter Koch and Wulf Österreicher, 'Schriftlichkeit und Kommunikative Distanz [Writtness and communicative distance]', *Zeitschrift für Germanistische Linguistik* 35, 3 (2007), pp. 346-375.

- Bernd Kortmann and Benedikt Szmrecsanyi, *Linguistic complexity: second language acquisition, indigenization, contact*, Berlin: De Gruyter, 2012.
- Christiaan Wouter Kusters, *Linguistic complexity: the influence of social change on verbal inflection*, Utrecht: LOT, 2003.
- Roger Lass, 'How to do things with junk: exaptation in language evolution', *Journal of linguistics* 26, 1 (1990), pp. 79-102.
- Sterre Leufkens, *Transparency in language: A typological study.*, Utrecht: LOT, 2015.
- Adolphe van Loey, *Middel nederlandse spraakkunst. Deel I. Vormleer*, 9th edn. Groningen: Wolters-Noordhoff, 1980.
- Gary Lupyán and Rick Dale, 'Language structure is partly determined by social structure.', *PLoS one* 5, 1 (2010), pp. e8559.
- Jaap van Marle, 'The unity of Morphology: on the interwovenness of the derivational and inflectional dimension of the word', *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1996, pp. 67-82.
- Yaron Matras, *Language Contact*, Cambridge: Cambridge University Press, 2009.
- Dirk Pijpops and Freek Van de Velde, 'Constructional contamination. A case study.', *Folia Linguistica* (Forthcoming).
- Dirk Pijpops and Freek Van de Velde, 'A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion', *Corpus Linguistics and Linguistic Theory* (2014), Published online, ahead of print.
- R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2013, <http://www.r-project.org/>.
- Tom Ruetten and Freek Van de Velde, 'Moroccorp: tien miljoen woorden uit twee Marokkaans-Nederlandse chatkanalen', *Lexikos* 23 (2013), pp. 475-556.
- Moritz Schönfeld, *Historische grammatica van het Nederlands [Historical grammar of Dutch]*, (Red.) Adolphe van Loey, 8th. edn. Zutphen: W.T. Thieme & Cie, 1970.
- Hendrik Schultink, *De morfologische valentie van het ongelede adjectief in modern Nederlands*, Den Haag: Van Goor, 1962.
- Alan Scott, 'The position of the genitive in present-day Dutch', *Word Structure* 4, 1 (2011), pp. 104-135.
- Alan Scott, *The genitive case in Dutch and German: a study of morphosyntactic change in codified languages*, Leiden: Brill, 2014.
- Ludwig Söll, *Gesprochenes und geschriebenes Französisch [Spoken and written Dutch]*, Berlin: Schmid, 1974.
- Dirk Speelman, 'Logistic regression: A confirmatory technique for comparisons in corpus linguistics', In: Dylan Glynn and Justyna A. Robinson (red.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. (Human Cognitive Processing [HCP]). Amsterdam: John Benjamins, 2014, pp. 487-533.
- John Suler, 'The Online Disinhibition Effect', *CyberPsychology Behavior* 7, 3 (2004), pp. 321-326.
- Taaladvies.net, 'Taaladvies.net', *Nederlandse Taalunie*, <http://taaladvies.net/> (7 April, 2015).
- VRT Taalnet, 'VRT Taalnet', <http://www.vrt.be/taal/> (7 May, 2015).
- Taaltelefoon.be, 'Taaltelefoon.be', *Vlaamse overheid*, <http://www.taaltelefoon.be/> (7 April, 2015).
- Peter Trudgill, *Sociolinguistic typology: social determinants of linguistic complexity*, Oxford: Oxford University Press, 2011.
- Freek Van de Velde, *Iets taalkundig(s): een functioneel georiënteerde analyse van deflexie en de genitiefontwikkeling in het Nederlands [Something linguistic: a functionally oriented analysis of deflexion and the development of the genitive in Dutch]*, Leuven: University of Leuven MA thesis, 2001.

- Freek Van de Velde, *De nominale constituent. Structuur en geschiedenis* [*The noun phrase. Structure and history*], Leuven: Leuven University Press, 2009.
- Freek Van de Velde and Joop van der Horst, 'Homoplasmy in diachronic grammar', *Language Sciences* 36, 1 (2013), pp. 66-77.
- Freek Van de Velde and Muriel Norde, 'Exaptation: taking stock of a controversial notion in linguistics', In: Muriel Norde and Freek Van de Velde (red.), *Exaptation in language change*. Amsterdam: John Benjamins, 2016, pp. 1-35.
- Freek Van de Velde and Fred Weerman, 'The resilient nature of adjectival inflection in Dutch', In: Petra Sleeman, Freek Van de Velde and Harry Perridon (red.), *Adjectives in Germanic and Romance*. (Linguistik Aktuell/Linguistics Today). Amsterdam: John Benjamins, 2014, pp. 113-145.
- William Venables and Brian Ripley, *Modern applied statistics with S*, 4th edn. New York: Springer, 2002.
- Freek Weerman and Petra De Wit, 'The decline of the genitive in Dutch', *Linguistics* 37, 6 (1999).

About the authors

Dirk Pijpops, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, FWO Flanders, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: dirk.pijpops@kuleuven.be

Freek Van de Velde, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: freek.vandevelde@kuleuven.be

