

They know it should be they, but still they say them

Stefan Grondelaers

1 Introduction

There is no better way to pay homage to Dany Jaspers than by whetting his appetite in the (linguistic) domains which are near and dear to his heart. The variable I discuss in this paper – the subject use of the object pronoun *hun* “them” as in *Hun hebben* “them have” – is certain to appeal to his taste for the peculiar: “subject-*hun*” (as we will henceforward call it) lay dormant in the grammar of Netherlandic Dutch for a century, before it rose its controversial head to inflame an entire nation by its unstoppable vitality, and this in spite of hysterical disapproval.

In its capacity as most famous standard language violation in The Netherlands, “subject-*hun*” is bound to tickle another sensitivity of Dany’s, his allergy to standard language abuse. I apologize beforehand if I do him injustice, but I believe Dany’s first reaction to subject-*hun* would be to attribute the “mistake” to slipping standards (magnanimously and benevolently, as he is, of course), and to recommend linguistic love and pride to “remedy” the error.

Much as I abhor the rise of subject-*hun* myself, I want to argue in this paper that the new pronoun is not a mistake sloppy Dutchmen make in defiance of their standard language norm, but an accountable innovation in the grammar of Netherlandic Dutch.

In what follows, I will rely on classical Labovian socio-syntax to “tease out” the function of *hun* in a minimal pair-approach, building on a large dataset of tweets with *zij* and *hun* which will also help me determine *hun*’s social meaning and prestige value. For some sort of prestige (change) must be involved if a variant can lead an underground existence for most of its career, before surfacing in the most visible of all positions (sentence-initially).

2 Case

The syntactic alternation investigated in this paper is the rapidly advancing use of the object personal pronoun *hun* “them” as a sentence subject in Netherlandic Dutch (in (2)), in lieu of the standard personal pronoun *zij* “they” in (1):

- (1) Als je zo speelt krijgen zij natuurlijk altijd kansen.
“If you play like that they will always get chances”
- (2) Als je zo speelt krijgen hun natuurlijk altijd kansen.
“If you play like that them will always get chances” (Van Hout, 2006, 277)

The subject use of *hun* was first observed in Vor der Hake (1911), but it received a major impetus in the last decades. The Spoken Dutch Corpus, which was compiled prior to the present acceleration in subject-*hun*'s dissemination, contains 213 tokens, produced by a limited number of speakers (11.3% who only use *hun*, 7.1% who use both subject-*hun* and subject-*zij*). Crucially, subject-*hun* is an exclusively Netherlandic innovation, which is categorically absent in Flemish Dutch.

Although all the available production data converge on the external conditioning of subject-*hun* as young, lowly educated, informal, and unscripted, there is general convergence in the literature (Van Hout, 2006; Van Bergen et al., 2011; Grondelaers & van Hout, 2011; Grondelaers et al., 2016) that *hun* is leaving its original social habitat, and that it is strongly increasing in popularity. The absence of any quantitative evidence to this effect does not invalidate the claim, because the change is so highly noticeable. One tell-tale sign is the fact that whereas our students tried to avoid the non-standard form until recently (or at least corrected it after having inadvertently produced it in a conversation with their teachers), most of them have no reticence (left) to admit that they use it.

We have proposed three reasons to account for the popularity of subject-*hun* in the face of its public disapproval. The first is the fact that there is room for change: if a natural language can ever be inefficient, then Dutch is “deficient” in its pronoun paradigms, which are littered with overlap and double duty: standard *zij*, for instance, is both the singular female and the general plural form of the personal pronoun. *Hun* is hardly an improvement on that score: in addition to a subject pronoun, it is an object pronoun as well as a possessive form.

But necessity and occasion do not suffice for change. The grammar has to benefit from the addition – the new form has to have a meaning/function the available variant does not have (to the same extent) – and there must be a com-

elling (prestige) reason why the Dutch are prepared to challenge norm sensitivities and appear stupid on account of it.

In previous work, we have found evidence for both factors (Grondelaers et al., Forthcoming). In a series of corpus analyses, we discovered that *hun* is an unreducible form which often bears stress, and which is significantly preferred over *zij* in contexts of “engaged negative contrast”:

- (3) Wij zijn Ajax, hun moeten oprotten (from a soccer fansite)
“We are Ajax, them have to fuck off”.
- (4) suriname mensen zijn best wel dom en hun denken dat wij in nederland dom zijn.
“suriname people are pretty stupid and them think that we in the Netherlands are stupid.”
- (5) wat hun denken is fictie. wat wij hebben is een #feit
“what them think is fiction. Wat we have is a #fact”

Whereas the available data have supported this “Contrast hypothesis” time and again, they failed to substantiate the widely accepted “Animacy hypothesis” (Van Bergen et al., 2011) that *hun*’s popularity is a consequence of the fact that it exclusively refers to animate and especially human referents, whereas *zij* can also denote inanimate referents.

In a sequence of speaker evaluation experiments, we found that *hun* indexes new, dynamic prestige. By using *hun*, speakers portray themselves as mildly provocative, assertive, and cool. The fact that subject-*hun* indexes this new social meaning reveals that important ideological changes are taking place in Netherlandic Dutch. The emergence of *hun*, more particularly, embodies the demise of the conservative standard language ideology which hierarchizes one variety of Dutch – uniform Standard Dutch – as the only correct, pure, and beautiful variety. The fact that *hun* can acquire prestige – even if it is only “new” prestige – suggests the emergence of a parallel ideology which “legitimizes” forms excluded by the conservative ideology. If prestige is a socio-psychological motivation for language users to copy linguistic behaviour, then *hun*’s modern prestige may explain why it is so popular.

A concern which complicates our research into *hun* is the fact that the validation of *hun*’s linguistic and social meaning requires different empirical techniques, corpus analysis for the former, and psycho-social experimentation for the latter. In what follows I argue that the micro-blogging platform Twitter represents the best possible data source to investigate the linguistic and the social

meaning of *hun* simultaneously, in one integrated analysis.

3 Twitter as a corpus

We believe that tweets can remedy the token shortage problem in syntactic variation research, the fact that constructional alternations are much less frequent in texts and conversations than words and sounds (Milroy & Gordon, 2003). This frequency problem becomes all the more pressing if the research topic is a stigmatised non-standard variant which users are taught to avoid. No matter how extensive they are, classical corpora offer no solution for the token shortage issue: the Leuven News Corpus totals half a billion tokens, but contains no instances of *hun* because the editors of the newspapers of which it consists typically ban prescriptively deviant usage.

In order to study subject-*hun* on the basis of non-elicited naturalistic data, we need language materials in which prescription and norm sensitivity play a lesser role. Internet chat is a case in point, but the gigantic Twitter corpus compiled by the Meertens Instituut (Tjong Kim Sang & Van den Bosch 2013) is an even better data source. Since tweets typically escape editorial control, they are known to feature characteristics of orality and non-standard usage, as a result of which they are eminently suited as “supplementary data for investigating non-frequent, non-canonical phenomena in spoken language” (Rehbein, 2014, 20). An empirical issue which remains to be resolved is whether the short format of tweets (max. 140 characters) renders them suitable to study constructional alternation.

We have extracted tweets with subject-*hun* or subject-*zij* from the Dutch eScience Centre corpus (Tjong Kim Sang & van den Bosch, 2013). In order to guarantee that our automatic queries extracted the intended use of *hun* as a subject (not as an object or possessive), we limited them as much as possible to non-ambiguous patterns in which *hun* or *zij* are sentence-initial, and precede the verb forms which collocate most often with them (as revealed by previous corpus analyses). After manual control, we ended up with 13.947 tweets, in which *hun* (n = 7260) is somewhat more frequent than *zij* (n = 6714). This is the sort of quantity which allows us to investigate the factors which determine the competition between two pronoun variants.

Unfortunately, there is almost no prior research into the linguistic features of tweets. A number of things stand out immediately, though. There is an overwhelming amount of (very) short messages which don't even reach the 140 char-

hypothesis (values “human” vs. “organization” vs. “non-human animate” vs. “non-animate”) and the Contrast hypothesis. In the short format of a tweet, it is not evident to find convenient implementations of involved negative contrast, but we can predict that if *hun* is a negative contraster, it will occur more frequently in syntactic copula+predicate environments (*we are good, x are bad*), and in tweets containing intensifiers (as in (11)-(12)), interjections (as in (13)), and non-linguistic symbols expressing agitation and ill-will:

- (11) hun zijn **tering** veatt!!
 “Them are fucking fat (meaning “cool”)”
- (12) <@> hahaha schatje niks tegen homo’s hoor maar hun doen **kanker** sneu #loveyou
 “Hahaha, deary, nothing against gay people, but them behave fucking weird”
- (13) kijk wat hun doen **wtf** #rtl5
 “Look what they are doing wtf #rtl5”

Crucially, all the predictions are supported by the data in tables 1 and 2. In table 1, the copulative schema triggers an outspoken preference for *hun*, and so do intensifiers and interjections in the containing tweet (see table 2). For non-linguistic additions, the effect is significant but somewhat less outspoken. Regression analysis reveals that all predictors added in function of *hun*’s alleged contrastive function strongly improve model fit; model prediction precision goes up from 52% (without predictors) to 73.9% (with predictors).

	<i>hun</i>	<i>zij</i>
copula+adjective <i>X are good</i>	4327	1428
copula+noun <i>X are assholes</i>	306	487
other	2627	4799

Table 1: Distribution of *hun* and *zij* in the copulative schema

Interestingly, this is the first analysis which returns a significant effect of Animacy (Animacy, in fact, has the highest effect size of all predictors). As predicted by Van Bergen et al. (2011), *hun* overwhelmingly refers to explicitly human referents

	<i>hun</i>	<i>zij</i>
-intensifiers	3520	3104
+intensifiers	5307	972
-interjections	5106	5778
+interjections	1518	501
-non-linguistic additions	5825	5862
+non-linguistic additions	799	417

Table 2: Distribution of *hun* and *zij* in combination with intensifiers, interjections, and non-linguistic additions

(*hun* 95.18%, vs. *zij* 75.25%), but contrary to prediction the Animacy effect is not due to a significantly larger frequency of *zij* in reference to inanimate objects: neither *zij* nor *hun* refer to non-animate referents often ($n = 28$, which is only 0.2% of all cases), and the difference is not categorical: there is one tweet in which *hun* refers to an inanimate entity:

- (14) okeee , woow hun zijn goed die rollschaatsennn . #hgt
 “okay, wow them are good these roller skatesss.”

If anything, the huge Animacy effect is the result of a skewed distribution (*zij* $n = 1657$ vs. *hun* $n = 349$) on value 2, when the pronoun refers to the people who constitute collectives and organizations, as in ...*I asked the housing department but they told me that...* Although Animacy is the best *hun*-predictor in the regression, it is impossible to explain the contrast effects in terms of animacy, whereas animacy can easily be accounted for as a by-product of negative contrasting: the most heated oppositions on Twitter are between (groups) of humans (as nearly all the examples testify to).

In sum, there is sufficient evidence that the use of *hun* as a subject is motivated by its comparatively richer functionality than the standard pronoun *zij*. However, this language-internal motivation does not explain *hun*'s soaring vitality: in order for *hun* to disseminate this rapidly, there must be a reversal or at least a radical weakening in the value system (language ideology) which frames and protects Netherlandic Standard Dutch as the only correct variety. Following Kristiansen (2009), who reported experimental evidence in support of a conservative and a progressive standard language ideology for Danish, and in support of a Danish standard “for the school” and a standard “for the media”, we found

evidence in previous experimental work for a progressive value system which renders variants like subject-*hun* (but also strong female Randstad accents, and even some Tussentaal features in Belgian Dutch) dynamically prestigious. A question which remains to be solved is whether we can document this dynamic prestige in Twitter data. While it is theoretically possible to use qualitative analysis to infer prestige values from specific contexts, we restrict ourselves in this paper to delimiting tweet contexts which boost the use of *hun*'s surmised prestige value. To this we turn in the final section.

5 Corpus analysis: *hun*'s social meaning

In section 3, we argued that intentional misspelling on Twitter in many cases is a self-styling tool to portray the author as a dynamic, non-conformist, witty persona through conscious challenge of language norms (Rampton, 2001). If *hun* has the dynamic social meaning we have attested, it is an evident styling option in this respect. We classified all tweets in "correct", "intentionally incorrect", and "error". The qualification "intentionally incorrect" was attributed to cases in which the tweeter voluntarily deviated from standard orthography. Dubious cases were coded as "error". Table 3 contains the distribution of *zij* and *hun* over the three classes:

	<i>hun</i>	<i>zij</i>
correct	5578	6058
intentionally incorrect	1229	360
error	453	296

Table 3: Distribution of *hun* and *zij* over the classes "correct", "intentionally incorrect", and "error"

As predicted, *hun*-use is a prime feature of dynamic self-stylisation (it is about four times more frequent than *zij* in intentionally erroneous tweets). But that is not all: separate regression analyses on the correctly and the intentionally incorrectly spelt tweets reveal that subject-*hun* is not only used more frequently in the latter category: the fact that the regression model for the intentionally misspelt tweets shows fewer and less significant predictors, as well as smaller effect sizes, demonstrates that *hun*-use is (much) less internally conditioned than

in the correctly spelt tweets, which in turn suggests that it is used much more consciously in the intentionally incorrect than in the correct tweets.

Subject-*hun*, as a consequence, functions on two levels: as an unconsciously used pronoun variant, and as a consciously deployed stylisation strategy. In both cases, however, a specific social meaning is at the heart of its usage. We could even go further and claim that *hun*'s social meaning (the *hun*-user as dynamic agent provocateur) and its linguistic meaning (negative contrast) partially overlap: provocation requires a third party who is the victim of this antagonistic behaviour.

6 Conclusions

Let us come to a number of conclusions. Empirically, it will be obvious by now that subject-*hun* is not going anywhere: it is not an error or perversion committed by stupid or lazy people in defiance of the official norm, but a grammatical option whose behaviour can be modelled and predicted.

Methodologically, we hope to have shown that in spite of the small size of individual tweets, Twitter is an extremely valuable data source for (socio)linguistic research; it can be used to investigate both the linguistic and social meaning determinants of syntactic change.

References

- Grondelaers, S., P. van Gent & R. van Hout. Forthcoming. On the inevitability of social meaning and ideology in accounts of syntactic change. Evidence from pronoun competition in Netherlandic Dutch. In T.K. Christensen & Jensen T.J. (eds.), *Explanations in sociosyntax. Dialogue across paradigms*. Amsterdam/Philadelphia: John Benjamins.
- Grondelaers, S. & R. van Hout. 2011. The standard language situation in the Low Countries: Top-down and bottom-up variations on a diaglossic theme. *Journal of Germanic Linguistics* 23. 199–243.
- Grondelaers, S., R. van Hout & P. van Gent. 2016. Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the Low Countries. *Taal & Tongval* 68(2). 119–149.
- Kristiansen, T. 2009. The macro level social meaning of late modern Danish accents. *Acta Linguistica Hafniensia* 40. 167–192.

- Milroy, L. & M. Gordon. 2003. *Sociolinguistics: method and interpretation*. Oxford: Blackwell.
- Rampton, B. 2001. Critique in interaction. *Critique of Antropology* 21. 83–107.
- Rehbein, I. 2014. Using Twitter for linguistic purposes: Three case studies. Paper presented in the workshop Web Data as a Challenge for Theoretical Linguistics and Corpus Design of the 36th Annual Conference of the German Linguistic Society, March 2014.
- Tjong Kim Sang, E. & A. van den Bosch. 2013. Dealing with big data: The case of Twitter. *Computational Linguistics in the Netherlands* 3. 121–134.
- Van Bergen, G., W. Stoop, J. Vogels & H. de Hoop. 2011. Leve Hun! Waarom hun nog steeds hun zeggen. *Nederlandse Taalkunde* 16. 2–29.
- Van Hout, R. 2006. Onstuitbaar en onuitstaanbaar. Ontstaan en verspreiding van het onderwerp “hun”. In N. van der Sijs, J. Stroop & F. Weerman (eds.), *Wat iedereen van het Nederlands moet weten en waarom*. 277– 286. Amsterdam: Bert Bakker.
- Vor der Hake, J. 1911. Is de beleefdheidsvorm U ’n verbastering van UEd.? *De Nieuwe Taalgids* 5. 16–34.