

Most borrowable construction ever!
A **large-scale** approach to contact-induced pragmatic change

Abstract

The construction *Beste boek ooit* ('Best book ever!') comes in different forms in Dutch. Variation is not only attested in the absence or presence of determiners and postmodifiers, but also in code choice: English, Dutch and hybrid (*Beste boek ever!*) variants co-occur.

This article investigates differentiation between instances with *ooit* and instances with *ever*. To ensure sufficient signal, we adopt a **bird's eye perspective**, analyzing over 100,000 observations from a Twitter corpus from the Low Countries (period 2011-2016). Our results reveal that (1) the two constructional variants increase in frequency in the time period under study, (2) this increase is more pronounced for the *ooit*-variant; (3) the *ever*-variant undergoes specialization towards a pragmatically marked form.

Overall, our account complements anglicism research (Andersen 2014) in four ways. First, we foreground constructional borrowing instead of single-word borrowing. Second, in working with Twitter data, we break with the tradition of print media corpora. Third, we explore **methods derived from big data analyses** in a field of research that has mainly relied on manual coding of small-scale datasets. Finally, we illustrate how matter and pattern replication can go hand in hand in contact-induced change.

Keywords contact-induced change; Construction Grammar; sentiment analysis; pattern replication; matter replication

1 Introduction: The socio-pragmatic turn in anglicism research

Apart from some early exceptions (e.g. Prince 1988, Poplack et al. 1988, Meeuwis 1991), anglicism research has long disregarded the social meaning of loanwords and the pragmatic effects at play in lexical borrowing. The main research goals were instead to provide inventories and taxonomies of loanwords, to demarcate lexical borrowing from codeswitching and to describe processes of morpho-phonological adaptation of source language elements to the receptor language structure (see e.g. Franco et al. forthcoming for an overview of research on gender assignment). Although these approaches have provided great insight into the more structural side of the borrowing process, several questions remained unaddressed. To name just a few: why do language users borrow items? When is a loanword preferred over an existing and well-established receptor language alternative? What social meanings are evoked through loanwords? How does contact-induced pragmatic change come about? **Which pragmatic differences exist (or come about) between source and receptor language forms? This final question lies at the core of our contribution.**

Through the rise of English as a global language, the lexical stock of many European languages has over the past decades been enriched with anglicisms, which has resulted in renewed scholarly attention for lexical borrowing in general and for English loanwords in particular. Typically, new studies on the topic focus on the unresolved questions of the past (see Andersen 2014, Onysko & Winter-Froemel 2011, Zenner & Kristiansen 2013), together resulting in what we could call the "socio-pragmatic turn" in anglicism research. This socio-pragmatic turn is also leaving traces in the methods used to study the more traditional questions of loanword adaptation and integration.

Our paper aims to contribute to this new wave of research, focusing particularly on the pragmatics of borrowing: relying on insights from Construction Grammar, we analyze the interplay between matter

and pattern replication (Matras & Sakel 2007) in pragmatic borrowing from English in Dutch. Where the former relates to what is traditionally known as direct borrowing (integrating foreign forms in a given receptor language), the latter concerns indirect borrowing, replicating meanings and constructions found in a source language with receptor language material (see e.g. Doğruöz & Backus 2009). Specifically, we focus on recent evolutions in the Dutch construction [(D) A_{superlative} N *ooit* (Ptcp)] (e.g. *de droogste zomer ooit (gemeten)* ‘the driest summer ever (measured)’), and link these to the English counterpart [(D) A_{superlative} N *ever* (Ptcp)] (*Driest summer ever (measured)*), which also occurs in the discourse of Dutch speakers, as a well-entrenched insertion that accounts as an instance of “constructional borrowing” (Colleman 2016, see below). Additionally, English-Dutch hybrid forms such as *beste boek ever* (‘best book ever’) are also attested, suggesting that there is no clear watershed between the Dutch construction and the English construction used by speakers of Dutch, and that we may merge them in one template [(D) A_{superlative} N *ooit/ever* (Ptcp)]. As will be discussed below, this paper focuses on alternations between *ooit* and *ever* in a Dutch corpus.

In our study of the constructions, we adhere to four convictions, which each complement the current practices of socio-pragmatic anglicism research. Our first conviction ties in with calls made by Onysko (2007), Andersen (2014), Fiedler (2014), Van de Velde & Zenner (2010) and Zenner & Geeraerts (2015) to expand on the restricted attention on single words in anglicism research: contact-induced pragmatic change not only happens at the level of words (see Peterson & Vaattovaara 2014’s exemplary analysis of the English loanword *pliis* ‘please’ in Finnish), but also occurs at the level of constructions and phrases, at the suprasegmental level and at the level of gesture. Our paper provides a case in point, studying pragmatic change and its potential formal markers in the multi-word construction [(D) A_{superlative} N *ooit/ever* (Ptcp)], henceforth referred to as the ‘*ooit/ever* construction’. In part resulting from the strong focus on single lexical units, traditional anglicism research also typically distinguishes quite strictly between direct borrowing (‘matter replication’; loanwords introducing new form and meaning, such as English *computer* in Dutch) and indirect borrowing (‘pattern replication’ such as loan translation and loan rendition, e.g. Dutch *wolkenkrabber* based on English *skyscraper*), whereas both can simultaneously drive contact-induced variation and change. As a second conviction, we insist on studying pattern replication and matter replication simultaneously (Matras 2009, Matras & Sakel 2007). Our third conviction concerns the type of data needed to study such pragmatic change in contact. Where traditionally, studies on English loans rely on print media corpora (Viereck 1980; Yang 1990) or lexicographical reference works, we insist on working with Twitter data, as it allows us to study language contact in a much more volatile environment that is caught between written and spoken, distance and proximity, and is hence specifically relevant for the observation of pragmatic change (see Beers Fägersten & Stapleton 2017). At the same time, as Twitter constitutes a genre in its own right, the analyses presented in this paper cannot be straightforwardly generalized to other usage contexts. Fourth, we are convinced that a pragmatic analysis of one single construction can only be relevant when sufficient signal is ensured to transcend the level of anecdotal evidence. To this end, we here explore the possibilities of a large-scale approach to anglicism research. As will become obvious in the conclusion, such approaches come with both benefits and drawbacks, but at the very least they offer a broader perspective on the construction under scrutiny than would be possible following the traditional approach in anglicism research, which hitherto mainly involves intensive manual coding of small-scale datasets.

The precise way in which these convictions are translated into our research design is explained below. Section 2 starts off with a brief presentation of the construction under scrutiny, also providing the readers with the key notions of Construction Grammar needed for a full understanding of the pragmatic change discussed. Additionally, the main hypothesis of this paper is presented. Next, we turn to a description of our large-scale approach: Section 3 presents our data in more detail, and

discusses the variables that we have worked with. The results of our analyses are presented in Section 4, followed by a discussion and conclusion in Section 5.

2 A Construction Grammar view on the *ooit/ever* construction

To describe the hypothesized changes in the *ooit/ever* construction, we follow the growing field of Construction Grammar. This approach provides us with an insightful framework to study contact-induced variation and change (see Höder 2012, Coleman 2016, Doğruöz & Backus 2009, Boas & Höder forthcoming, Zenner et al. forthcoming). The basic tenets of Construction Grammar are well known (see Goldberg 1995, 2006; Croft & Cruse 2004 and Hoffmann & Trousdale 2013): language is seen as a network of constructions. Constructions are form-meaning pairings of various degrees of complexity (from atomic to multi-word) and various degrees of schematicity (from lexically specific, over partially instantiated, to fully schematic). Goldberg (2006: 5) argues that even patterns that are fully compositional – that is, whose meaning can be computed on the basis of the constituting sub-constructions – are separate constructions themselves, provided they have a frequency that is high enough to make them entrenched in the mind of the language user. So, English has a construction [*Stop acting like a N*], with N standing for a derogatory noun, even if the meaning of the concrete instantiations (*Stop acting like a fool*, *Stop acting like a clown*, *Stop acting like an asshole*, *Stop acting like a douchebag* ...) can be computed on the basis of the subconstructions it contains, namely the imperative construction, the *ing*-complement construction, the NP construction etc. (including the individual words, as lexical entries are also constructions). Semantic compositionality does not entail that the construction is not separately, indeed redundantly, stored in the mental lexicon.

How does this apply to the *ooit/ever* construction? In principle, strings like *de warmste zomer ooit* ('the hottest summer ever') could be analysed as an instance of an ordinary noun phrase. For many speakers, the string is, however, (also) recognisable as an instance of a more specific separate multiword, partially schematic and partially lexically fixed construction [(D) A_{superlative} N *ooit/ever* (Ptcp)]. This construction has one lexically specific element: the adverb *ooit* or *ever*, which is here used as an adverbial postmodifier in the noun phrase.¹ Other elements of the *ooit/ever* construction are the regular constituents of the noun phrase: a determiner, an adjective (in this construction in the superlative), and a head noun.² The determiner and the participle are put in parentheses, as they are optional. This optionality is interesting, and will play a crucial role in the analysis (below).

Given its frequency and its behavior, to be discussed below, the idea that the *ooit/ever* construction is stored as a construction in its own right is plausible enough. It is, however, not clear whether it is just one construction, or rather many related constructions. Are the Dutch variant (*Warmste zomer ooit!*)

¹ Some analyses would consider *ooit/ever* as a discontinuous postmodifier of the superlative adjective, but alternative analyses are possible as well (see Van de Velde 2009:115-118, where other, more marginally occurring variants of the construction are also discussed). One could argue that the adverb submodifies a participle which itself submodifies the adjective. In a dependency grammar, this could be rendered as N -> A -> Ptcp -> Adv (c.q. *zomer* -> *droogste* -> *gemeten* -> *ooit*), and in the construction without the participle, there is an ellipsis. Construction Grammar eschews zero elements and ellipsis, however, because it works on output-oriented configurations (Booij & Audring, forthcoming; Van de Velde, forthcoming). We will not pursue the matter here.

² To avoid overburdening the formalisation of the construction, we leave out other possible slots in the NP, such as predeterminers and peripheral modifiers (see Payne & Huddleston 2002), and other postmodifiers of the noun *de meest onverantwoorde moeder in spe ooit* 'the most irresponsible mother to be ever'. We also ignore the paradigmatic variation on the participle (Ptcp), which can be substituted by *te + infinitive* (e.g. *het hoogste niveau ooit te bereiken* 'the highest level ever to be reached'), as well as further dependents of the participle (e.g. *de eerste film ooit vertoond in een Franse filmzaal* 'the first movie ever shown in a French movie theater').

and the English variant (*Hottest summer ever!*) **two variants** of the same construction, or are they two different constructions? This is not an easy question, as hybrids are also attested: the English adverb *ever* can easily be combined with a Dutch adjective and noun. This suggests we are dealing with constructional variants, rather than with separate constructions.³ In our focus on the co-occurrence of matter and pattern replication, this paper studies those constructional variants that occur in Dutch contexts, focusing on the alternation between *ever* and *ooit*.

Another thorny issue concerns the question whether the different formal manifestations count as separate constructions, irrespective of the language choice. Take the Dutch variant (*Warmste zomer ooit!*): as indicated by the round brackets in the formalization above, it has two optional parts: the determiner and the participle. We will call the construction without determiner and without the participle (*Warmste zomer ooit!*) the bare variant, and the construction where one or both elements are realized (e.g. *De warmste zomer ooit gemeten* ‘the hottest summer ever measured’) the non-bare variant.⁴ Is this an instance of free variation? One argument to treat them as distinct, non-free variants is that the bare construction seems to come with exclamative mood by default, whereas the non-bare version can, but need not have exclamative mood. This can be seen in (1), which is a factual statement without exclamative mood. The bare construction cannot normally be used in an objective, non-exclamative context. The same adjective-noun combination can be used in the bare construction, but this coerces an exclamative meaning on the noun phrase, which may sound odd when the adjective-noun combination suggests a more objective reading, as in (2), unless it is used in highly specific contexts, e.g. as an advertisement in the car industry.

- (1) *de p1800 is misschien wel het meest internationaal bekende volvo-model ooit*
‘the p1800 may well be the most internationally renowned Volvo model ever’
- (2) *meest internationaal bekende volvo-model ooit!*
‘most internationally known volvo-model ever’

A second argument to see them as separate (i.e. non-free) constructional variants is that they behave differently in their external syntax: it is harder for the bare variant to be integrated as a noun phrase in a clause. Rather, it stands on its own as an exclamation.⁵ This is true for the English and the Dutch variant, as illustrated in (3)-(7). Exceptions can be found though, as in (9), where the bare *ooit* construction *is* integrated in a clause. In fact, such examples are quite common in our dataset. The hybrid/English variant with *ever* seems to be more explicitly tied to the bare use, as is seen in (8). Of course, follow-up perception studies are needed to inventory grammatically judgements for these constructions.

- (3) *this is the worst movie ever released*
- (4) *(*this is) worst movie ever!*

³ This raises the question what exactly the difference is between a separate construction and a constructional variant. The most straightforward approach is to use the meaning as the decisive criterion: if two distinct forms have different meanings, they count as separate constructions. If they have the same meaning, they count as variants of one construction. The latter situation is unproblematic in cases of free variation, but the question is of course, whether variation can ever be free in language, especially if we take into account difference in pragmatic contexts, see Baayen et al. (2013:254-256) for an attempt at clarifying the continuum. We will treat the two formal realisation of the *ooit/ever* construction as constructional variants, rather than as separate constructions, without implying that there are no differences in the spheres of use.

⁴ The determinerless cases cannot only be linked to Englishization, but are also – specifically in less subjectified contexts – connected to the construction’s use in newspaper headlines.

⁵ It is not uncommon for stand-alone noun phrases to express non-declarative mood (see Sadock & Zwicky 1985: 187-188). Exclamatives in particular are wont to take forms other than regular full finite clauses (see Haeseryn et al. 1997:1433-1436 for Dutch).

- (5) *Dit is de slechtste film ooit vertoond*
- (6) *(*Dit is) slechtste film ooit!*
- (7) *(*Dit is) slechtste film ever!*
- (8) *?Dit is de slechtste film ever vertoond*
- (9) *wazigste droom ooit gehad en precies om 8:40 weer wakker (Twitter corpus)*
‘haziest dream ever had and precisely at 8:40 awake again’

We can assume that the Dutch bare construction is a result of English influence, counting as an instance of pattern replication. In English the bare construction has existed since the beginning of the 20th century, at least in American English (OED, s.v. *ever*, adv.). In Dutch, by contrast, the bare construction starts to occur in the latter quarter of the 20th century only (Van der Horst & Van der Horst 1999: 231-232), at a time when (particularly American) English influence made itself felt in Dutch. This potential influence of English has not gone unnoticed in normative accounts: the construction is discussed on official language planning websites, such as taaladvies.net (a sub-branch of taalunieversum.org)⁶ and onzetaal.nl⁷. The construction also occurs in the list of examples provided for anglicisms on Dutch Wikipedia⁸. The two official norming websites accept the construction as “correct” Dutch, but not without explicitly referencing to English influence in its emergence and providing an alternative expression.

In the remainder of this paper, we will investigate pragmatic change in the Dutch use of the construction, studying possible indications of the influence by the English construction in (6) – both in the hybrid/English variant with *ever* and in the Dutch version with *ooit*. As we discussed above, the latter can be classified as an instance of pattern replication. The form with *ever* instead counts as an instance of (partial) matter replication (*ever* instead of *ooit*, possibly accompanied by other English slots in an otherwise Dutch context) combined with pattern replication. The use of the construction in the source language is not in itself central to our analysis: the fully English construction only factors in when it is included in a Dutch context. Elaborate alternational switches and full shifts to English (Muysken 2000) are explicitly excluded, as we focus on the use of the borrowed construction in an otherwise Dutch context.

3 Data and methods

Precisely how we can determine whether a specific constructional change in Language 1 is influenced by a related construction in Language 2, is an unresolved matter in linguistics (see Poplack et al. 2012, Coleman 2016). For this paper, we combine the principles of variationist linguistics with a pragmatic analysis of the construction under scrutiny (compare Levey et al. 2013). Additionally, as both matter and pattern replication are at play, we hope to uncover the potential influence of language contact in the recent development of the construction somewhat more confidently. Adopting a large-scale approach, we additionally attempt to base our description of change on sufficient signal. However, as will become clear below, relying on a sizeable corpus comes with both benefits and drawbacks.

3.1 Twitter corpus

As we are dealing with a construction that has subjective semantic and pragmatic meaning, and as we are studying what looks like a change in progress, we need an informal, recent, diachronically sliceable corpus of sufficient size. For this reason, we focus on Twitter data, although as mentioned above, we should be careful not to generalize our results to other genres: with its character limitations and

⁶ <http://taaladvies.net/taal/advies/vraag/624/>, consulted February 2018

⁷ <https://onzetaal.nl/taaladvies/anglicismen/>, consulted February 2018

⁸ https://nl.wikipedia.org/wiki/Anglicisme#cite_note-2, consulted February 2018

specific conventions (such as hashtags, retweets), Twitter can be considered as a separate genre, a sub-community in which language is used in a specific way.

Computer mediated communication (CMC) has drawn quite some attention from (contact) linguists over the past decade (Mahootian 2005, and see e.g. De Decker 2015 for Dutch). As CMC embeds language users in the global world wide web while it simultaneously also connects them to their individual network, online environments provide the ultimate location for contact between English as the global *lingua franca* and the mother tongues of broad groups of speakers (compare Androutsopoulos 2013 for the broader notion of networked multilingualism). Combining elements from writing and speaking, playing on distance and proximity, social media platforms add to this the volatility needed to study change in progress (Beers Fägersten & Stapleton 2017). The use of CMC for research on anglicisms is so far however quite rare (though see e.g. Leppänen et al. 2009, Sharma 2012), with a particular lack of quantitative large-scale approaches (Zenner et al. 2017). This is surprising, given the possibilities that have opened up through ‘the fourth paradigm’ (data-intensive scientific discovery; Hey et al. 2009) (see e.g. Alex 2008 for one of the rare NLP approaches to anglicisms at this point).

In our focus on a short, expressive and subjective construction, Twitter appears to be the most appropriate platform. As it imposes a limit of 140 characters to messages, users typically (i) resort to shorter expressions and (ii) tweet about content they feel opinionated on. From a more practical perspective, it is quite easy to collect a large corpus of location-specific tweets in a short amount of time, thanks to Twitter’s developer’s API (application programming interface) (see also Eisenstein 2014, Yuan et al. 2015).

Our corpus was collected through *twiqs.nl*, a website that allows researchers and students to search Dutch tweets (Tjong et al. 2013), in combination with Twitter’s own API. This means that *twiqs* automatically imposes a language filter: only tweets recognized as Dutch are included in the database.⁹ We here focus on tweets collected in a window from 1 January 2011 to 31 December 2015.¹⁰ A direct consequence of the language filter is that tweets including the fully English version of our construction do not make the pass, except when they occur in an explicitly Dutch context, as in (10). Of course, Dutch speakers also use the English constructional variant without embedding it in a Dutch tweet. We explicitly exclude such elaborate switches and full shifts to English (Muysken 2000) and focus on the use of the borrowed construction in an otherwise Dutch context. Additionally, based on manual coding of a random sample of 1000 tweets with *ever*, it is safe to say that the majority of tweets with *ever* (about 70%) that pass *twiqs*’ language filter are in fact instances of the hybrid construction (7).

(10) *je bent zo druk en heb geen tyd, dat je het op twitter zet #biggest lie ever*
‘you are so busy and have no time, so that you put it on twitter #biggest lie ever’

All tweets that contained *ever* or *ooit* were selected, which of course led to noise such as (11-12). A further selection was made based on linguistic criteria. Firstly, we relied on morphological markers like superlative suffix *-st/-ste* as in *grootste huis* ‘biggest house’ or the analytic superlative, as in ‘meest indrukwekkende prestatie’ (‘most impressive accomplishment’), to increase precision. Secondly, from

⁹ The *twiqs* crawler changed its language identification procedure in May 2014, but a manual comparison between both procedures by the *twiqs* administrator showed an overlap in correct Dutch language assignment for 914 out of sample of 1000 tweets (<https://ifarm.nl/erikt/twin/>, posted by admin on 09/05/2014).

¹⁰ Searching the *Twiqs* database for the literal strings “ooit” and “ever” in the period 01/01/2011 00h00m00s to 31/12/2015 23h59m59s resulted in 8,725,619 tweet IDs. These were then queried from the Twitter API using the *tweepy* python library, resulting in 8703619 tweets. Exactly 22,000 tweets could not be retrieved. Manual inspection showed this was mainly because they were sent from discontinued Twitter accounts.

the remaining tweets, we excluded cases where *ooit/ever* occurred in relative clauses following the head noun, as in (13).

- (11) *ik heb'm ooit wel eens ontmoet*
'I met him once' (literally: 'I have him ever once met')
- (12) *dit is de beste film die ik ooit heb gezien*
'this is the best movie that I have ever seen'
- (13) *Dat is de beste film die ik ooit gezien heb*
lit. 'that is the best movie that I ever seen have'

Based on these criteria, 465,000 observations remained (out of 8,725,619). We do not have any information on recall, but given the size of our database this seems less relevant. Precision rates are more crucial, and are satisfactory: comparing manual coding of a random sample of 5,000 occurrences with the automatic selection criteria reveals a precision rate of over 95%. To keep the computational efforts feasible, a random sample of 105,283 tweets was selected from this database for further analysis, which exploits the maximal capacity of the spreadsheet software used for exploratory analyses (Microsoft Excel). The main purpose is now to scrutinize possible frequency effects in the constructions with *ooit* and *ever*, and to verify which tweets show characteristics of the pragmatic change described above, visual in both formal (e.g. loss of determiner) and semantic (e.g. increased subjectivity) features. Below, we describe the selection and operationalization of the features relied on in more detail.

3.2 Variables

In our focus on patterns of change, a first and rather obvious variable is the period in which the tweet was published. Our database contains the exact publication date of the tweets, to the minute. As this is not the most insightful resolution for measuring the kind of linguistic change we are interested in, we transformed the timestamp to a categorical variable, with one level per 6 months for each of the 5 years in our database. This leads to 10 periods, with an average number of about 300 million tweets identified as Dutch by *twiqs.nl* per period in total, and an average number of 10,528 tweets per period containing our construction (standard deviation 2,691). Our results indicate that the Dutch *ooit* construction experiences a frequency boost from 2013 onward (see Section 4 for a more detailed discussion). Part of the analysis will be focused on the period 2013-2015, with 6, rather than 10 time intervals.

Second, we need to define a number of diagnostics to detect a change. On the side of linguistic codes, we make a distinction between instances of the *ever*-variant (*grappigste boek ever*) or the *ooit* variant (*grappigste boek ooit*). A total of 20,659 constructions in our database contain *ever*, the other 84,624 use *ooit*.

Next, we introduce some distinctions to witness possible pragmatic specialization of the construction. Specifically, we aim to reveal whether the in itself expressive construction [(D) A_{superlative} N *ooit/ever* (Ptcp)] is gaining in subjectivity and is used in a more exclamative fashion. To this end, we code for formal and semantic features. Of course, manually coding over 100,000 examples is not a realistic endeavor, which means that – following what is typical in big data approaches – we aim overall for 'good enough'-precision in coding rather than solid 100% correct coding.

3.2.1 Formal features

On the formal side, we distinguish between the bare and the non-bare variant, illustrated in (15) and (14), respectively (see Section 2, above).

- (14) *ik vind het toch de beste film ooit*
 'I consider it to be the best movie ever'
- (15) *beste film ooit!*
 'best movie ever!'

To isolate bare constructions, we looked at the presence of an overt determiner. Concretely, the absence or presence of the determiner was coded in a semi-automatic fashion. After isolating the slot preceding the adjective in the noun phrase, all slots with a token frequency over 20 were manually coded (e.g. *de* 'the' and *mijn* 'my' received the code "determiner", whereas *is* 'is' received the code "noDeterminer"). Thanks to the Zipfian distribution of linguistic elements, including determiners **typically ranking among the top frequency elements**, this procedure helped us classify 84% of the data already (60,891 with determiner, 26,952 without). After manual selection of the remaining 16%, we decided that the best way to proceed was coding these as instances of "noDeterminer". Taking a random sample of 500 tweets among these approximately 27,000 instances, and manually checking the automatic "noDeterminer"-code, the precision of the determinerless category is 91.8%. The remaining 8.2% include false positives of the construction as such, as well as false hits of determinerless instances.

To increase the reliability of our formal approximations of the anticipated specialization in the construction, we took advantage of the observation that the prototypical bare construction in our corpus is not syntactically embedded in a clause, see (15). Coding for syntactic embeddedness is unfortunately not so easy. In principle, one could rely on automatic parsers developed in corpus linguistics. **This is not a feasible solution, however, as the traditional parsers for Dutch (such as Alpino, see van Noord 2006) are known to severely underperform on Twitter data (van Noord et al. 2013).** Instead, based on manual scrutiny of the data we derived ten rules to estimate (un)embeddedness which we then implemented in a Python script that ran through all tweets in our database. As an example, our simplest rule for estimating "unembeddedness" states that the construction can be classified as syntactically unembedded when it is preceded and followed by punctuation (see (16)). Our simple rule for estimating "embeddedness" states that constructions can be classified as embedded when they are preceded and followed by three consecutive words uninterrupted by punctuation (see (17)).

We calculated precision and recall rates for each rule on a random sample of 1,000 tweets. For the manual coding of the test sample, we adopted a strict interpretation of non-embeddedness: instances with post-modifiers were classified as embedded (see (18)). Next, we combined the rules that produced satisfactory results. This way, we arrived at recall of 95.2% for our test set: 952 tweets in the set receive an embeddedness code. Precision of the coding procedure is 91.1% overall: 747 of the 822 tweets that are automatically classified as embedded are actually embedded (90.8%), and 121 of the 130 tweets that are automatically classified as unembedded are in fact unembedded (93.1%).

- (16) *@brouwertje ben het nu aan het terugkijken_ beste ajax ooit_ genieten. gaan we niet meer zien, zo'n team of dit voetbal in nederland.*
 '@brouwertje am now watching it again_ best ajax ever_ love it. we will not witness this again, this type of team or this type of football in the Netherlands.'
- (17) *@vi_nl en waar is de duurste speler ooit die het bedrag niet waard is?*
 '@vi_nl and where is the most expensive player ever who is not worth the pay?'
- (18) *#nieuwstwitter grootste winst ooit voor chelsea onder abramovitsj | sport*
<http://t.co/blh8h00wqp> #actueel
 '#nieuwstwitter biggest profit ever for chelseae under abramovitsj | sport
<http://t.co/blh8h00wqp> #actueel'

For the database as a whole, we obviously cannot provide precision rates. Recall is 95.9%, with 4,733 examples that are left uncoded. Of the 100,550 coded examples, 14,729 are coded as unembedded (14.6%).

What we call the ‘bare’ construction in the analyses in Section 4, is a combination of: (a) absence of a determiner, (b) absence of a postmodifying participle and (c) non-integration in a clause (“unembedded”).

3.2.2 Semantic features

On the semantic side, providing good diagnostics for a potential change in the *ooit/ever* construction is not obvious. Assessing the degree of subjectivity or exclamative mood by manual human coding based on introspection is problematic. The manual coding runs the risk of not being objective enough, unless it is done by multiple raters whose judgments are subjected to a measure for inter-rater agreement. This was not an option because of the large number of datapoints, exceeding what is even feasible for a single rater to go through in a reasonable amount of time. We decided to turn to automatic Sentiment Analysis. Although such ‘big data’ analysis tools are still in full development, especially for a smaller language like Dutch, we do want to make the methodological case for integrating them into large scale linguistic analyses of pragmatic usage of constructions.

Sentiment Analysis is one of the current hot topics in natural language processing, as opinionated texts make up an increasingly large part of CMC. One direct benefit is that open-source programs are available to assess the degree of subjectivity and polarity of specific adjectives and of sentences at large. Our coding relied specifically on the Python module Pattern¹¹, developed by the research unit CLiPS at Antwerp University (see De Smedt & Daelemans 2012b for an outline of the method) and used for insightful sentiment analyses of Dutch Twitter data in previous studies (Boonzajer-Flaes, Rudinac & Warring 2016; Walasek 2017)¹². Sentiment scores were collected at the level of the entire tweet (as opposed to the construction or adjective only)¹³. Scores were available for 73,654 tweets in our corpus. Polarity averages on 0.29, with scores located between -1 (maximally negative sentiment) and 1 (maximally positive sentiment), with 0 serving as neutral point. Subjectivity, which measures the degree of personal involvement in a statement (the term “covers the fact that a particular element or construction requires reference to the speaker in its interpretation”; De Smet & Verstraete 2006), averages on the remarkably high score of 0.74, with scores ranging from 0 (no subjective meaning) to 1 (maximally subjective meaning). Table 1 contains tweets from our dataset with different combinations of polarity and subjectivity scores. Tweet 22582 illustrates how a tweet can have negative polarity and be fairly objective at the same time: it refers to a fact (the airing of the last episode of Dexter) and an (observable) expression of negative emotion.

¹¹ <https://www.clips.uantwerpen.be/pages/pattern-nl> (De Smedt & Daelemans 2012a)

¹² Both studies included a small scale benchmarking against other systems. Boonzajer-Flaes et al. 2016 showed Pattern to outperform NLTK’s Naive Bayes sentiment classifier for polarity. Walasek 2017 (p.22) benchmarked Pattern against her own classifiers and showed similar F1 scores for dominant class polarity scores (0.65). Although we did perform cursory inspection and “sanity checks” (see next footnote), validation of Pattern’s sentiment scores against human judgements was outside the scope of the current study. We therefore interpret our results with caution, stressing the methodological issue of using “big data methods”, including automatic sentiment analysis, on very large linguistic datasets for identifying pragmatic usage patterns.

¹³ As a sanity check on Pattern’s coverage (and its underlying lexicon), we also looked at the individual adjectives occurring the *ever/ooit* construction as the dominant sentiment bearing items in our tweets. Adjective-specific scores were available for 1,216 of the 4,360 adjective types in the constructions our database, together accounting for 85,654 adjective tokens in the constructions.

ID –country date ¹⁴	Tweet	Translation	Polar.	Subj.
3255 - NL 10/04/2015	[...] voortreffelijk gedineerd! heerlijkste #amuse ever! [...]	[...] <i>superbly dined! Most delicious #amuse ever!</i> [...]	1.0	1.0
42250 - NL 08/04/2014	grootste feest ooit! miljoenen mensen komen! en ik ben uitgenodigd door #jesus wow!	<i>Biggest party ever! Millions of people coming! And I am invited by #jesus wow!</i>	0.98	0.2
69459 - BE 19/08/2013	op 17 augustus 1982, produceerde polygram - een onderdeel van philips - de allereerste compact disc ooit. #audiorevolutie #tijdvoor80	<i>On August 17th 1982, polygram – a division of Philips – produced the very first compact disc ever. #audiorevolution #timefor80</i>	0.0	0.1
22582 - NL 30/09/2013	de laatste #dexter ever, daar gaan we! #huilen [...]	<i>the last #dexter [episode] ever, there we go! #cry</i>	-0.5	0.2
46830 - NL 21/09/2013	als je wifey jaloers is omg.. ergste nachtmerrie ever.	<i>If your girlfriend is jealous OMG... worst nightmare ever.</i>	-0.6	1.0
8202 - BE 24/07/2013	jaws 3, meest idiote film ooit.	<i>Jaws 3, most idiotic film ever.</i>	-0.8	1.0

Table 1 - Example tweets sorted from positive to negative polarity and with different subjectivity scores

With these measures, we are in a position to gauge what diachronic trends are borne out in the data. Possible trends we can expect to find:

- (i) A growth in the frequency of the *ever* variant, due to matter replication
- (ii) A growth in the frequency of the *ooit* variant, due to pattern replication
- (iii) An increase of the bare construction and subjectivity in the *ooit* variant, due to pattern replication
- (iv) An increase in the bare construction and subjectivity in the *ever* variant, due to pattern replication

We clearly expect to find the same direction in the evolution of the Dutch and hybrid/English variants of the construction. What we are particularly interested in, however, is comparing the speed of evolution:

- (v) Does the *ooit* construction grow faster than the *ever* construction?
- (vi) Is the pragmatic specialization more outspoken for the *ever* variant?

4 Results and discussion

With over 100,000 tweets, even the smallest trends will easily reach significance when working with traditional hypothesis-testing inferential statistic tools. Indeed, in fourth-paradigm research (Hey et al. 2009), operating on **datasets of this size**, p-values tend to become meaningless. There are solutions, by including effect size, and by validation on other corpora, but here we have opted to emphasize data visualization to uncover the trends in our data.

4.1 Overall frequency and formal features

Figure 1 shows the diachronic evolution of the *ooit/ever* construction (and see Table 2 for absolute and relative numbers). The *ooit* variant outnumbers the *ever* variant throughout the whole time period, though we should not forget that elaborate alternating switches to English and fully English tweets

¹⁴ Tweets have been anonymized by removing handles, user names and exact timestamps/locations. IDs are random indices from our dataset, which can be obtained on request to the second author (NAME).

sent out by Dutch users including the *ever* variant are not sampled in our database. What is clear, however, is that the difference between *ooit/ever* becomes markedly bigger after the first half of 2013, when there is a sudden jump in the use of the *ooit* variant. The *ooit* variant more than quadruples in frequency, from 17.7 instances per million tweets to 87.5 per million tweets.¹⁵ This jump visually obscures the fact that the *ever* variant undergoes a steady increase as well, though: it doubles its significance from 5.4 instances per million tweets in 2011 to 11.1 instances per million tweets in 2015.

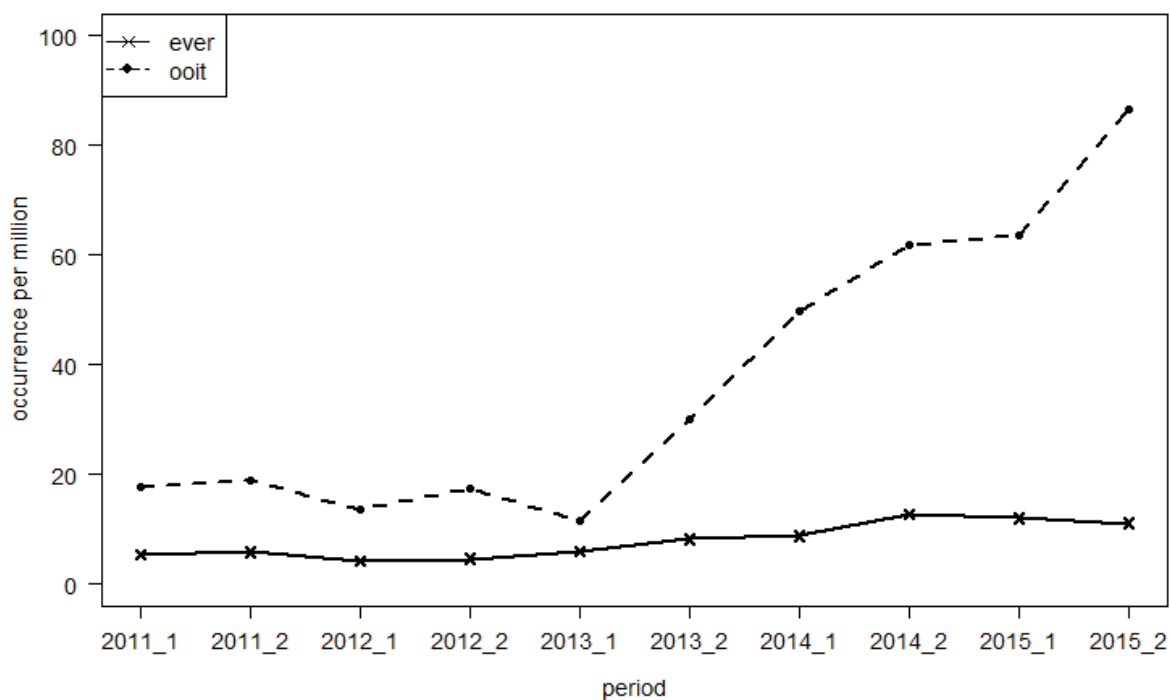


Figure 1 – diachronic evolution in occurrences per million (*ooit* and *ever*)

period	total tweets	EVER		OOIT	
	n	n	n per million	n	n per million
2011_1	339,423,570	1,822	5.368	6,004	17.689
2011_2	384,689,151	2,255	5.862	7,230	18.794
2012_1	378,917,209	1,626	4.291	5,187	13.689
2012_2	396,009,050	1,851	4.674	6,940	17.525
2013_1	433,837,633	2,577	5.940	5,080	11.709
2013_2	361,364,684	2,987	8.266	10,857	30.044
2014_1	226,456,201	1,996	8.814	11,262	49.731
2014_2	165,583,089	2,119	12.797	10,220	61.721
2015_1	157,386,779	1,903	12.091	9,991	63.481
2015_2	136,995,838	1,523	11.117	11,853	86.521

Table 2 – diachronic evolution in occurrences per million (*ooit* and *ever*)

¹⁵ Note that the actual occurrence of the constructions will be about four times as high, as our analysis zooms in on a randomly selected subset of the data

The trends in Figure 1 take into account all observations, irrespective of whether they are instances of the bare construction or not. One might expect that the increase of the *ooit/ever* construction is due to the use of the bare construction, which has carved out a separate niche by more explicitly expressing exclamative mood than the original construction. This does not seem to be the case, however. Figure 2 shows the proportion of the bare construction over the total *ooit/ever* observations. As can be appreciated from scrutinizing Table 3, the proportion remains fairly stable over time: the bare construction accounts for roughly between 7% and 12% of the total number of observations, for *ooit* and for roughly between 12% and 17% for the *ever* variant. This shows that the exclamative use is tied more explicitly to the English construction, which is visually supported by Figure 2: the line for the *ever* variant is consistently higher on the Y-axis than the *ooit* construction, which reveals that the exclamative use is tied more explicitly to the English construction.

What we seem to have here is an instance of what one may refer to with an aphorism as ‘a rising tide lifts all boats’. Overall, the *ooit/ever* construction is used more and more through time, and though one might assume that this rise would be felt first and foremost in the more specialist bare construction, this specialised variant cannot monopolize the market.

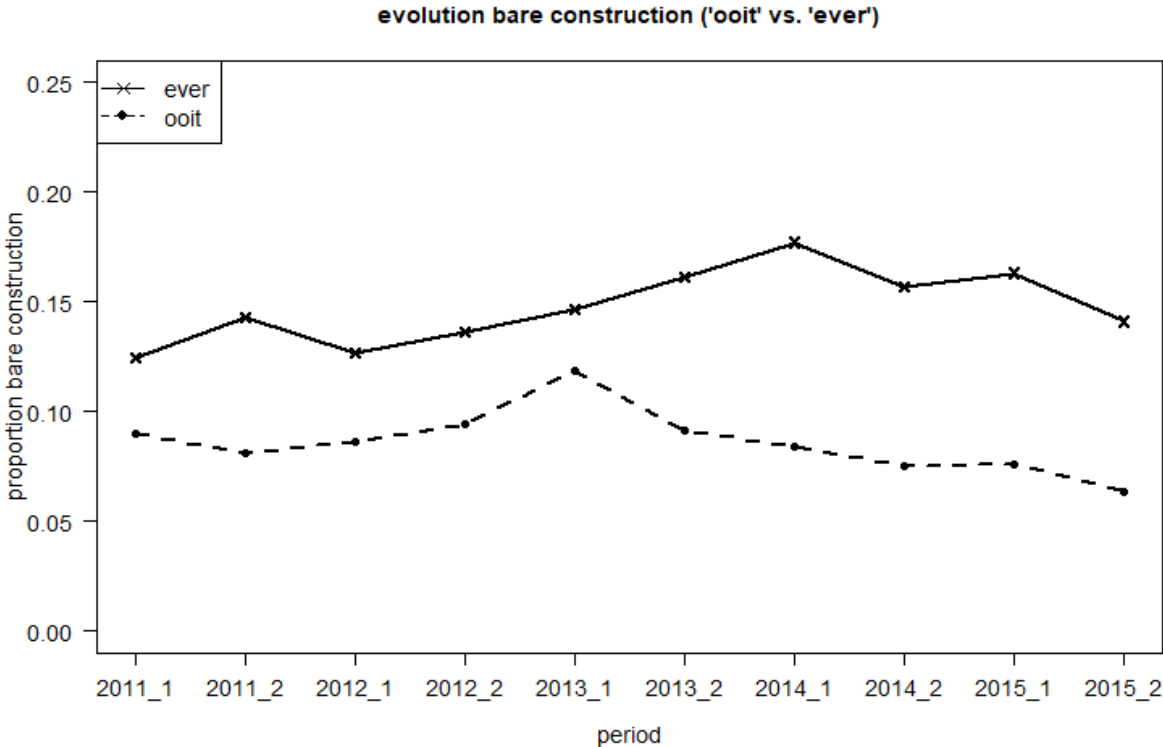


Figure 2 – diachronic evolution in the relative occurrence of the bare construction (*ooit* and *ever*)

period	EVER			OOIT		
	n (bare)	n (total)	% (bare)	n (bare)	n (total)	% (bare)
2011_1	227	1,822	12.46	539	6,004	8.98
2011_2	322	2,255	14.28	584	7,230	8.08
2012_1	206	1,626	12.67	447	5,187	8.62

2012_2	252	1,851	13.61	654	6,940	9.42
2013_1	378	2,577	14.67	602	5,080	11.85
2013_2	482	2,987	16.14	994	10,857	9.16
2014_1	353	1,996	17.69	949	11,262	8.43
2014_2	332	2,119	15.67	770	10,220	7.53
2015_1	310	1,903	16.29	762	9,991	7.63
2015_2	215	1,523	14.12	754	11,853	6.36

Table 3 – numbers for the diachronic evolution in the bare construction (*ooit* and *ever*)

The same information is presented in a different fashion in Figure 3, which gives the proportion of the use of the *ever* variant over the total uses of the *ooit/ever* construction from 2013 onward, distinguishing the bare from the non-bare construction (and see Table 4 for numbers). The overall trend is that *ooit* is gaining increasingly more ground, which is consonant with the idea that the English matter replication of the *ever* variant (also in hybrid forms) blazed the trail for pattern replication. It is also clear that the takeover by *ooit* happens more rapidly in non-bare construction. This may be taken as an indication that *ever* in Dutch is undergoing a process of specialization, retreating more and more to the expression of exclamatives, at least in Dutch contexts. It remains hard to assess what *ever* does in fully English tweets sent out by Dutch users. However, this is not the focus of our analysis (see above).

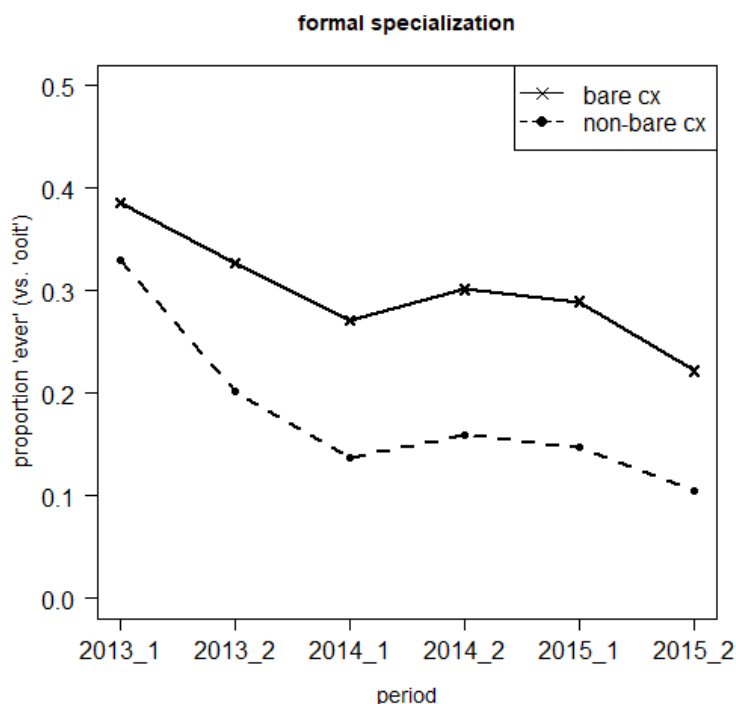


Figure 3 – relative preference for *ever* over *ooit* in bare constructions and non-bare constructions

period	BARE			NON-BARE		
	ever	ooit	% (ever)	ever	ooit	% (ever)
2013_1	378	602	38.57	2,199	4,478	32.93

2013_2	482	994	32.66	2,505	9,863	20.25
2014_1	353	949	27.11	1,643	10,313	13.74
2014_2	332	770	30.13	1,787	9,450	15.90
2015_1	310	762	28.92	1,593	9,229	14.72
2015_2	215	754	22.19	1,308	11,099	10.54

Table 4 – number for the preference for *ever* over *ooit* in bare and non-bare constructions

4.2 Semantic features

Now let us turn to the semantic side of the construction. As explained in Section 2, the tweets were subjected to a Sentiment Analysis gauging their subjectivity and polarity. Focusing on the former, we can see that the diachronic trend in the subjectification seems to follow the trend in the proportion of the bare construction (see Figure 4, which includes both bare and non-bare constructions). From 2011 to 2013, there is a rise in subjectivity, followed by a decrease from 2013 to 2015. Since there is no serious correlation between subjectivity and the bare construction (kendall correlation = 0.06), this mirroring of the diachronic trend is not merely due to the constructional link between form and function, but rather follows the same underlying trend of the ‘rising tide’: up until 2013, the *ooit* variant is used increasingly subjectively. Then, the construction is entrenched, and occupies a strong enough position from where it may spread to less prototypical contexts. For the *ever* construction, by contrast, we do see a clear rise of subjectivity of the tweets over the periods, but with a strong ceiling effect¹⁶. As we will discuss at the end of this section, this ceiling effect can be attributed to the increasing dominance of subjective-evaluative *best* in the adjective slot. This further underlines our idea of *ever* retreating to a pragmatically specialized niche.



Figure 4 – diachronic evolution in subjectification rates for the *ooit* variant

¹⁶ Shown in the boxplots by increasing median values within interquartile ranges lodged at the scale maximum.

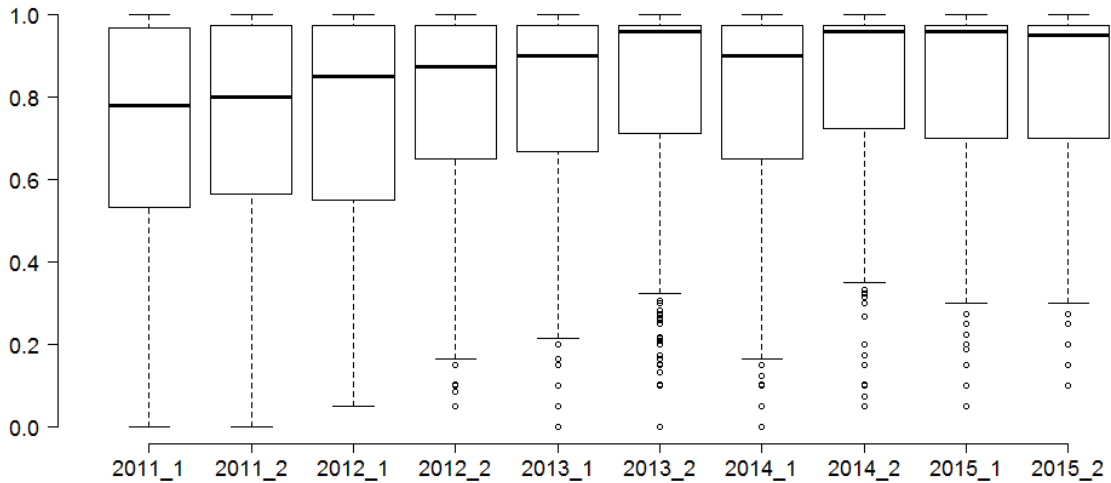


Figure 5 – diachronic evolution in subjectification rates for the *ever* variant

Turning to polarity then, we get the same image:¹⁷ the polarity of *ooit* tweets follows the camel hump in 2013 (although less outspoken than for subjectivity), but in attenuated form compared to the trend in Figure 4. For *ever*, the rise of the subjectivity value is mirrored in the polarity measurement.

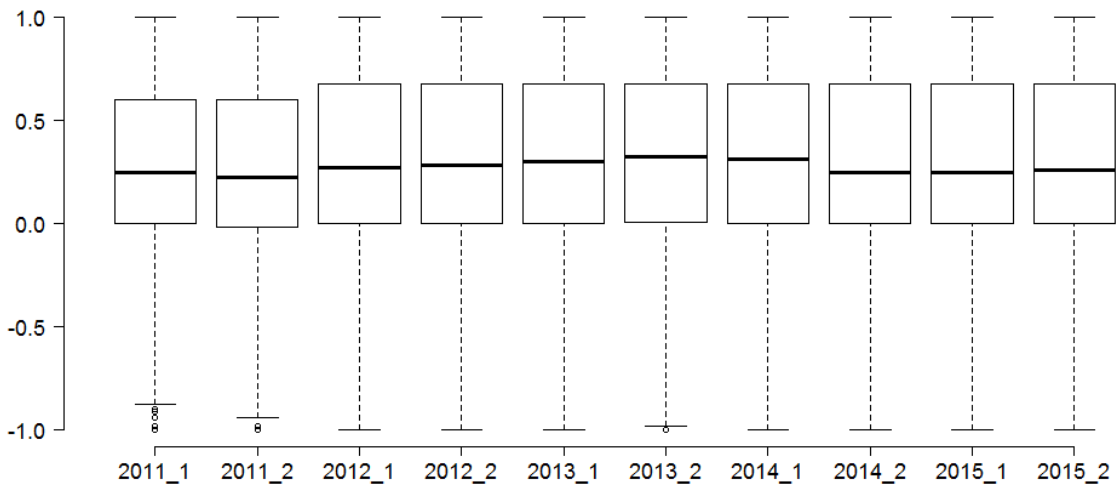


Figure 6 – diachronic evolution in polarization rates for the *ooit* variant

¹⁷ Subjectivity and polarity scores show a spearman rank correlation of 0.65. Although highly evaluative expressions are often highly subjective (“best party ever”), both scores do not completely overlap (e.g. “highest approval ratings ever” has positive polarity while being relatively objective measurements)

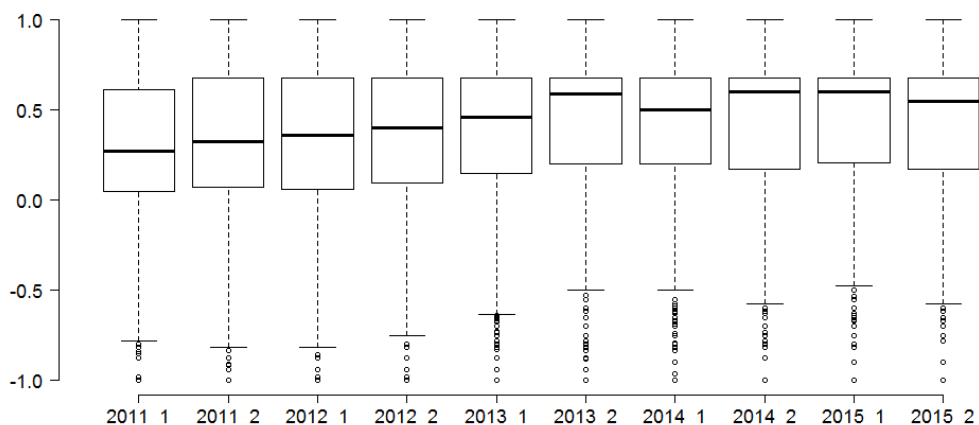


Figure 7 – diachronic evolution in polarization rates for the *ever* variant

The analysis of the formal features of the *ooit/ever* construction above had revealed that *ever* tweets seem to undergo a process of specialization: they hold out better against the post-2013 *ooit* gulf in the bare construction. A similar process can be observed on the semantic side. After 2013, the *ever* construction in Dutch retreats to the highly subjective and highly polarized contexts. This retreat can also be seen at the lexical level. The topmost adjective in all periods is *best(e)*. Focusing on the bare construction only, Figure 8 shows that for the *ever* construction, the adjective *best(e)* accounts for an increasing proportion of the attested adjectives through the years. For the *ooit* construction, by contrast, there is no clear trend.

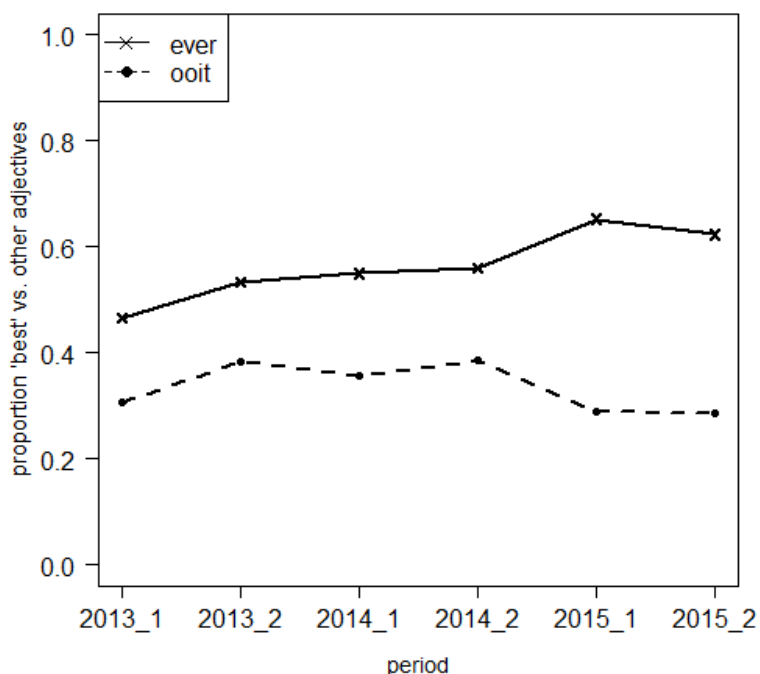


Figure 8 – proportion of bare constructions with *best* in the adjective slot (*ooit* and *ever*)

Two comments are in order. First, despite our large-scale approach which started off with over 100,000 examples, we see that funneling the data has led to a (relatively speaking) scarce amount of data: Table 5 reveals an average of only 192 constructions combining *best* and *ever* per period. Future research

will have to blow up the scale of this study even more to arrive at convincing results at the level of individual types of adjectives. Second, a noteworthy pattern is nevertheless revealed in Figure 8 and Table 5: by the second half of 2015, over 60% of all bare *ever* constructions contain the adjective *best*. This is perhaps an indication that the adjectival slot of the construction is becoming fixed, leading to the exclamative construction [*best N ever*]. Corroborating this pattern depends on follow-up analysis are needed to track the future development of the construction in Dutch.

period	EVER (bare)			OOIT (bare)		
	n (<i>best</i>)	n (other A)	% (<i>best</i>)	n (<i>best</i>)	n (other A)	% (<i>best</i>)
2013_1	176	202	46.56	185	417	0.31
2013_2	257	225	53.32	382	612	0.38
2014_1	194	159	54.96	338	611	0.36
2014_2	186	146	56.02	298	472	0.39
2015_1	202	108	65.16	221	541	0.29
2015_2	134	81	62.33	217	537	0.29

Table 5 – numbers for bare constructions with and without *best* in the adjective slot (*ooit* and *ever*)

5 Conclusion

Our diachronic analysis of over 100,000 tweets with the *ooit/ever* construction in the period 2011 to 2015 has shown that the construction undergoes a precipitous increase in frequency: instances of the *ever* variant double in a time span of five years, and instances of the *ooit* variant quadruple in the same time period. We had expected that the increase in frequency would have been mirrored by a concomitant change in the morphosyntax, with an increase in the so-called bare construction, i.e. the variant without a determiner, unembedded in a higher clause, and without a participle accompanying *ooit/ever*. This construction arose in Dutch through pattern replication of an English construction. However, we found no indication of a propagation of this variant: the proportion of the bare construction remains relatively stable throughout time. It seems then, that the frequency increase benefitted both the more innovative bare construction and the non-bare construction. We may speculate that this could be due to some sort of priming effect: in the slipstream of the innovative construction, the old variant is used with increasing frequency as well. We had also expected that the increase in frequency would have had a concomitant change in the semantics of the construction. Using methods from Sentiment Analysis, we measured the subjectivity and the polarity of the tweets. Only the *ever* construction showed an increase on both dimensions, with tweets becoming semantically more subjective and more positive over time.

The overall picture that emerges then is one where the *ooit* variant gets a boost in frequency in all its manifestations, bare as well non-bare, more as well as less subjective and positive as well as negative contexts. This follows naturally from the fact that the pragmatic specialization is one of degree here. Rather than going from a ‘neutral’, unmarked construction to a subjectified one, we see an increase in subjectivity of an in essence subjectified construction: about 80% of all *ooit* constructions in our database are subjective to some degree.

Turning to the *ever* variant we also see a frequency boost, which is however less outspoken than for *ooit*. We specifically noted that the *ever* variant seems to be retracting to a specialized niche. Increasingly, *ever* constructions are found in subjective, positive contexts, holding out better against

their Dutch competitor with *ooit* in the special bare construction, and with a more limited type promiscuity. Perhaps we could go as far as to say that the adjective slot is becoming lexically fixed for the bare *ever* construction in Dutch: where the bare *ooit* construction shows no immediate lexical restraints, we see an evolution towards [*best N ever*] (instead of [$A_{\text{superlative}} N \text{ ever}$]) for the *ever* variant.

Despite these insights, our study inevitably has some serious limitations. For one thing, the ‘good enough’-approach typical for **large-scale** data analysis means that we do not have full control over what is going on in our database. At the same time, sufficient control is ensured through manual coding of random samples, and the sheer size of our database should prevent possible remaining noise from greatly interfering with our results. Second, our analysis does so far not take into account diatopic ‘lectal’ variables, like the geographical provenance of the tweets, or the social status and social aspirations (in terms of tweets and retweets), age, gender and socio-economic status of the tweeters. **Related to this is our restricted focus on Twitter data: as Twitter can be considered as a sort of ‘virtual subculture’, we cannot simply generalize our results. Further analyses should focus on different genres and registers.** Furthermore, our time frame was limited to five years, whereas the *ooit/ever* construction has been in use for a much longer period. To see more spectacular diachronic trends, it may be advisable to zoom out and use a larger timespan with lower resolution. **At the same time, adopting a worm’s eye perspective to the data will necessarily have to complement our bird’s eye view, which at the very least show that a more detailed pragmatic analysis of the tweets will most likely reveal interesting patterns of differentiation in the use of the *ever* and the *ooit* construction.** Finally, if we want to truly understand the sudden rise in the *ooit* construction in 2013, we will have to study the *ever* construction in the source language alongside our language filtered database of Dutch tweets.

Still, we hope to have shown that **a large-scale approach and a bird’s eye perspective** can be used to provide empirical answers to such elusive concepts as pattern replication and pragmatic contact-induced change. Traditionally, (diachronic) research into lexical borrowing has focused on the transfer of concrete lexemes in small datasets. While these studies cannot by any means be denied to have yielded valuable insights, there is, in our opinion, room for working on larger datasets and semi-automatic annotation to track trends in loan behavior, that can then be fed back to more qualitative, small-scale analysis of the construction under scrutiny. For one thing, we see that a peaceful co-existence of pattern replication and matter replication of the same construction seems untenable. More specifically for our database, a sudden rise of the *ooit* construction (pattern borrowing) was answered by a retreat of the *ever* construction (matter borrowing). Our results indicate pragmatic meaning differentiation that provided both constructions with their *raison d’être* in our corpus.

Relying on massive data to come to this conclusion of course also holds a danger: all trends easily reach significance, even with small alpha levels of 0.001 or less. Significance testing thus loses its value, and we are better off visually inspecting data. Eyeballing trends is of course liable to a certain level of subjectivity. We hope to have been cautious enough here to not over-interpret our data, and to have derived conclusions for which we have sufficient support in the graphs. At the very least, we hope to have inspired readers to undertake similar **large-scale** approaches to the study of contact-induced variation and change, or to use our results as point of departure for more in-depth analysis of the *ever/ooit*-construction and its evolution.

Acknowledgements

Eternal gratitude to Tom Ruetten for jumping in and saving the day for the automatic annotation of embeddedness. **We additionally want to thank two anonymous reviewers and the special issue editors for their useful comments to earlier drafts of this paper. Needless to say, all remaining errors are our own.**

References

- Alex, B. *Automatic Detection of English Inclusions in Mixed-lingual Data with an Application to Parsing*. 2008. PhD Thesis. University of Edinburgh, Edinburgh, UK.
- Andersen, G. 2014. Pragmatic borrowing. *Journal of Pragmatics* 67: 17-33
- Androutsopoulos, J. 2013. Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism* 1-21.
- Baayen, H., A. Endresen, L.A. Janda, A. Makarova & T. Nessel. 2013. Making choices in Russian: pros and cons of statistical methods of rival forms. *Russian Linguistics* 37(3): 253-291.
- Beers Fägersten, K. & K. Stapleton. 2017. Introduction. Swearing research as variations on a theme. In K. Beers Fägersten & K. Stapleton (eds), *Advances in swearing research. New languages and new contexts*, pp. 1-15. Amsterdam: John Benjamins.
- Boas, H. C. & S. Höder (eds.) (under review). *Constructions in Contact*. Amsterdam/Philadelphia: John Benjamins.
- Booij, G. & J. Audring. Forthcoming. 'Multiple motivation, partial motivation: the role of output schemas in morphology'. In: G. Booij (ed.), *The construction of words. Advances in Construction Morphology (CxM)*. Dordrecht: Springer.
- Boonzajer Flaes J., Rudinac S., Worrying M. (2016) What Multimedia Sentiment Analysis Says About City Liveability. In: Ferro N. et al. (eds) *Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science, vol 9626*. Springer, Cham.
- Colleman, T. 2016. A reflection on constructionalization and constructional borrowing, inspired by an emerging Dutch replica of the 'time'-away construction. *Belgian Journal of Linguistics* 30(1): 91-113.
- Croft, W. & D. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- De Decker, B. 2015. Prototypische chatspeakkenmerken in Vlaamse tienerchattaal. De invloed van gender, leeftijd en medium. *Taal en tongval* 67(1): 1-41.
- De Smedt, T. & W. Daelemans. 2012a. Pattern for Python. *Journal of Machine Learning Research*, 13: 2031-2035.
- De Smedt, Tom & W. Daelemans. 2012b. "Vreselijk mooi!" (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* 3568-3572.
- De Smet, H. & J.-C. Verstraete. 2006. Coming to terms with subjectivity. *Cognitive Linguistics* 17(3): 365-392.
- Doğruöz, A. Seza & Ad Backus. 2009. Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change. *Bilingualism: Language and Cognition* 12(1): 41-63.
- Eisenstein, Jacob. 2014. "Identifying Regional Dialects in Online Social Media." Georgia Institute of Technology. <https://www.cc.gatech.edu/~jeisenst/papers/dialectology-chapter.pdf>
- Fiedler, S. 2017. Phraseological borrowing from English into German: Cultural and pragmatic implications, *Journal of Pragmatics* 113: 89-102.
- Franco K., Zenner E., D. Speelman. Forthcoming. Let's agree to disagree. (Variation in) the assignment of gender to nominal anglicisms in Dutch. *Journal of Germanic Linguistics*.

- Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. 2006. *Constructions at Work: the Nature of Generalization in Language*. Oxford: Oxford University Press.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. 2nd edn. Groningen: Nijhoff.
- Hey, T., S. Tansley & K. M. Tolle. 2009. *The fourth paradigm: data-intensive scientific discovery*. Vol. 1. Redmond, WA: Microsoft research.
- Höder, Steffen. 2012. Multilingual Constructions: a Diasystematic Approach to Common Structures. In Kurt Braunmüller & Christoph Gabriel (eds.), *Multilingual Individuals and Multilingual Societies*, 241-257. Amsterdam: John Benjamins.
- Hoffmann, T. & G. Trousdale (eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
- Leppänen, S., A. Pitkänen-Huhta, A. Piirainen-Marsh, T. Nikula, & S. Peuronen. Young People's Translocal New Media Uses: A Multiperspective Analysis Of Language Choice And Heteroglossia. *Journal of Computer-Mediated Communication* 14: 1080-1107.
- Levey, Groulx & Roy 2013 in LVC: A variationist perspective on discourse-pragmatic change in a contact setting
- Mahootian, Shahrzad. 2005. Linguistic change and social meaning: Codeswitching in the media. *International Journal of Bilingualism* 9(3/4): 361-375.
- Matras, Y. 2009. *Language Contact*. Cambridge: CUP.
- Matras, Y. & J. Sakel 2007. Investigating the mechanisms of pattern-replication. *Studies in Language* 31: 829-865.
- Meeuwis, M. 1991. A pragmatic perspective on contact-induced language change. Dynamics in interlinguistics. *Pragmatics* 1 (4): 481-516.
- Muysken, P. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: CUP.
- Onysko, A. 2007. *Anglicisms in German. Borrowing, lexical productivity and written codeswitching*. Berlin/New York: Walter de Gruyter.
- Onysko, A. & E. Winter-Froemel. 2011. Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43(6): 1550-1567.
- Payne, J. & R. Huddleston. 2002. Nouns and noun phrases. In: R. Huddleston & G. Pullum (eds.), *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press. 423-523.
- Peterson, E. & Vaattovaara 2014. *Kiitos* and *pliis*: The relationship of native and borrowed politeness markers in Finnish. *Journal of Politeness Research* 10(2): 247-269.
- Poplack, S., D. Sankoff, & C. Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation, *Linguistics* 26: 47-104.
- Poplack, S., L. Zentz & N. Dion. 2012. Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition* 15: 203-225.
- Prince, E.F., 1988. On pragmatic change: the borrowing of discourse functions. *Journal of Pragmatics* 12: 505-518.

- Sadock, J.M. & A.M. Zwicky. 1985. Speech act distinctions in syntax. In: T. Shopen (ed.), *Language typology and syntactic description*. Vol. 1, 155-196. Cambridge: Cambridge University Press.
- Sharma, B. K. 2012. Beyond social networking: Performing global Englishes in Facebook by college youth in Nepal. *Journal of Sociolinguistics* 16(4): 483-509.
- Tjong, E., K. Sang & A. van den Bosch. 2013. Dealing with Big Data: the Case of Twitter. *Computational Linguistics in the Netherlands Journal* 3.
- Van der Horst, J. & K. van der Horst. 1999. *Geschiedenis van het Nederlands in de twintigste eeuw.* 's-Gravenhage: Sdu.
- Van de Velde, F. 2009. *De nominale constituent. Structuur en geschiedenis*. Leuven: Leuven University Press.
- Van de Velde, F. Forthcoming. Iterated exaptation. In: Booij G. (Eds.), *The construction of words. Advances in Construction Morphology (CxM)*. Dordrecht: Springer.
- Van de Velde F. & Zenner E. 2010. Pimp my Lexis: het nut van corpusonderzoek in normatief taaladvies. Hendrickx E., Hendrickx K., Martin W., Smessaert H., Van Belle W., Van der Horst J. (eds.), *Liever meer of juist minder? Over normen en variatie in taal*, 51-68. Gent: Academia press.
- van Noord, G. 2006. At Last Parsing Is Now Operational. *TALN 2006*: 20-42.
- van Noord, G. et al. 2013. Large Scale Syntactic Annotation of Written Dutch: Lassy. In: P. Spyns and J. Odijk (eds.), *Essential Speech and Language Technology for Dutch: Results by the STEVIN-programme*, 147-164. Berlin, Heidelberg: Springer.
- Viereck, K. 1980. *Englisches Wortgut, seine Häufigkeit und Integration in der österreichischen und bundesdeutschen Pressesprache*. Frankfurt am Main: Peter Lang.
- Walasek, N. 2017. Semi-supervised opinion mining: learning sentiment towards vaccination on Dutch tweets. M.Sc. Thesis, Univ. of Tilburg. (http://www.ru.nl/publish/pages/769526/n_walasek.pdf)
- Yang, W. 1990. *Anglizismen im Deutschen*. Tübingen: Max Niemeyer Verlag.
- Yuan H., Guo, D., Kasakoff, A., & Grieve, J. (2015). Understanding US regional linguistic variation with Twitter data analysis. *Computers, Environment and Urban Systems*. (available online).
- Zenner, E., A. Backus & E. Winter-Froemel (eds.). Forthcoming. *Cognitive Contact Linguistics*. Berlin/Boston: Mouton de Gruyter.
- Zenner, E. & D. Geeraerts. 2015. I'm queen of the world! (Semi-)fixed English expressions and constructions in Dutch. *Taal en Tongval* 67(2): 247-274.
- Zenner, E. & G. Kristiansen. 2013. *New perspectives on lexical borrowing. Onomasiological, methodological and phraseological innovations*. Berlin/Boston: De Gruyter.
- Zenner, E., T. Ruetten & E. Devriendt. 2017. The borrowability of English swearwords: An exploration of Belgian Dutch and Netherlandic Dutch tweets. In K. Beers Fägersten & K. Stapleton (eds), *Advances in swearing research: New languages and new contexts*, 107-138. Amsterdam: John Benjamins.