

Article

Are We Indeed So Illuded? Recency and Frequency Illusions in Dutch Prescriptivism

Marten van der Meulen 

Centre for Language Studies, Radboud Universiteit, 6525 HT Nijmegen, The Netherlands;
marten.vandermeulen@ru.nl

Abstract: In 2005, Arnold Zwicky posited two misapprehensions about language: the Recency Illusion, or the false idea that certain language variation is new, and the Frequency Illusion, the erroneous belief that a particular word or phrase occurs often. Since their conception, these concepts have received widespread attention in popular scientific linguistics, but quantitative research investigating their application is scarce. The purpose of this paper is to provide an empirical investigation of Zwicky's proposed illusions. It does so by collecting statements about recency ('this word is new') and frequency ('this construction occurs often') from a database of Dutch prescriptive publications (1900–2018). I assessed their accuracy by comparing them to linguistic sources, including dictionaries, and usage corpora and other data. Our research showed that recency statements were rare, but that frequency statements, especially using high frequency terms such as *vaak* ('often'), were commonplace. Compared to usage, most prescriptive recency and frequency statements for both lexis and grammar indeed constituted Zwickian illusions. This seems partly due to genuine erroneous or unsupported beliefs by authors, but also partly to prescriptive genre conventions and rhetorical choices. Our explorative research highlights the complex usage–prescriptivism interface, and argues for more research into this aspect of language perceptions.

Keywords: prescriptivism; Dutch; frequency illusion; recency illusion; language norms; language use



Citation: van der Meulen, Marten. 2022. Are We Indeed So Illuded? Recency and Frequency Illusions in Dutch Prescriptivism. *Languages* 7: 42. <https://doi.org/10.3390/languages7010042>

Academic Editors:
Joan Costa-Carreras and Carla Amorós Negre

Received: 24 November 2021

Accepted: 14 February 2022

Published: 22 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2005, American linguist Arnold Zwicky published a post on the well-known blog *LanguageLog* entitled 'Just Between Dr. Language and I' (Zwicky 2005). In what would later be called a "classic post" (Zimmer 2019), he wrote about a language column which claimed that the construction *between you and I* had only started to appear over the last few decades. Zwicky invalidated this claim by showing that its usage dated back at least 150 years, and possibly as much as 400 years. He also speculated more generally about such recency statements, claiming that the claim in this column was not an isolated occurrence, but rather an example of several "systematic dogged misapprehensions" (Zwicky 2006) about language that people entertained. Zwicky described two of these closely related misapprehensions as follows:

- Recency Illusion: If you've noticed something only recently, you believe that it originated recently.
- Frequency Illusion: Once you notice a phenomenon, you believe it happens a whole lot. (Zwicky 2005)

These illusions manifest in metalinguistic statements of various sorts. In the case of the aforementioned *between you and I*, the columnist wrote that "about 20 years ago U.S. English-speakers began switching *me and X* to *X and I* everywhere the phrase occurs" (quoted in Zwicky 2005). Here, the temporal marking is a form of recency statement. Other statements include the word *recent* or variations thereof, such as *this word has recently come into use*, or statements such as *this has lately come into fashion*. As for the Frequency Illusion, Zwicky reports in his original blog post how members of a certain research group expressed

a belief about the common use of quotative *all*, which some even perceived to be used “all the time” (2005). *All the time* is a clear statement about the frequency with which a word occurs; other statements include other high frequency terms such as (*very*) *often* or *repeatedly*, but also frequency terms denoting presumably lower degrees of frequency, such as *sometimes*, *rarely* or *hardly ever*. Even a seemingly neutral description *this word occurs* can implicitly refer to frequency.

The crucial part about these illusions is that the perceived recency and frequency may not actually be based in fact, as Zwicky’s example shows. As such, these illusions are clearly related to the large and well-researched group of cognitive biases, or the “systematic errors in judgment and decision-making common to all human beings which can be due to cognitive limitations” (Wilke and Mata 2012, p. 531). The existence of such ‘errors’ does not mean that people are always wrong in assuming recency and frequency: there are well-known cases in which people make decisions based on demonstrably accurate information (cf. Arnold et al. 2000; Brysbaert et al. 2017). However, for some reason, sometimes this impression is wrong. One of the main reasons why people exhibit cognitive errors is selective attention. More specifically for language, Zwicky claims, people are influenced by the fact that “hardly anyone has a panoptic view of language variation” (2005). The illusions, then, are not only an effect of selective attention to language variation, but also of the (almost inevitable) restricted access that language users have to a language or language variety as a whole. Consciously or subconsciously, language users take their personal experience as representative for a larger variety, lect, or language, resulting in inaccurate estimations of the recency or frequency of certain words or phrases.

The above explains *why* these illusions occur; Zwicky also hypothesizes about *who* are likely to suffer from them, and *what* language is prone to being the subject of these illusions. According to him, education plays a key role in determining which language users are susceptible to false claims about recency and frequency (similarly, Cameron notes that knowledge of prescriptive discourse is ‘common’ for educated language-users (Cameron 1995, p. viii). Especially those with higher levels of education are vulnerable, as they have more “faith in the engines of correctness” (Zwicky 2008). Language professionals, linguists, and really anyone who is “reflective about language” are also susceptible (Zwicky 2005). As for which language is targeted, Zwicky claims that language users are particularly disposed towards certain types of language, which he states are “non-standard, informal, and spoken variants” (Zwicky 2008). One thing that variants with these parameters have in common is that they are often the subject of normative attention. For example, the whole point of the aforementioned column about *between you and I* was to eradicate mistakes in the use of this phrase. It seems, then, that recency and frequency statements are probably made in relation to *usage items*, or particular cases of *disputed variation*.

This combination of who (reflective language users) and what (disputed language variation) means that prescriptive publications, such as usage guides and style guides, are particularly likely to contain Recency and Frequency Illusions. After all, when we talk about these so-called ‘engines of correctness’, then surely prescriptive publications, which contain nothing but usage items, and aim at fixing language mistakes and eradicating unwanted variation, are the vehicles that run on these engines. Earlier research has shown that such works do indeed contain statements about usage, particularly relating to frequency (Sundby et al. 1991, p. 38; Kostadinova 2018, pp. 154–66; Ayres-Bennet 2020). However, this research is limited in scope, and recency statements do not seem to have been studied at all. Whether there is any systematicity to the use of frequency and recency statements, for example depending on the variable, author, or time period, is as of yet unclear.

Equally unclear is to what extent prescriptive statements about usage are an accurate reflection of actual language use. Recent work has shown that the position of prescriptivism being “often in defiance of normal usage” (Trask 1999, p. 246) may not be as categorically true as was long assumed to be the case. For example, van der Meulen and Rutten (2022) show that in a work by the Dutch linguist Matthijs Siegenbeek (1774–1854) there is a relation between the degree of frequency term used to describe a disputed variant and

the actual frequency with which this variant occurs in usage (i.e., higher terms, such as *vaak* ('often') show higher occurrences of usage). Similarly, in their study of the *one of the (few) X who Y-construction*, Hogeweg et al. claim that a frequency statement in a particular style guide is correct in asserting that this construction occurs *vaak* ('often') with singular agreement (Hogeweg et al. 2018, p. 339).¹ By contrast, in the one example in which an alleged Recency Illusion is explicitly tested against usage data, Liberman (2010) comes to an uncertain conclusion about whether the claim of recency for the sentence-initial *so* is correct. However, he admits that his analysis is "crude", as he only looks at frequencies per million words in the Corpus of Historical American, without checking whether instances are relevant or not, without an operational definition of recency, and without taking potential genre differences into account.

In summary, although we have some evidence for recency and frequency statements being used in prescriptive publications, we know little about the extent and manner of their use; whether or not these statements are accurate is similarly uncertain. The present paper addresses these two issues, by first mapping recency and frequency statements in Dutch prescriptive publications from 1900 onwards. Second, I assess whether these statements are illusions in the Zwickian sense, or whether they conform to actual usage patterns, by comparing the contents of these statements against actual language usage. In order to investigate whether certain types of linguistic variable are more likely to be correctly spotted than others, I look at both lexical and grammatical usage items. In the absence of any quantitative work exploring these matters, the purpose of the present paper is explicitly explorative: I want to see whether this type of methodology works, and what issues we encounter. The methodology and the materials are discussed in Section 2. Next, I look at the results in the prescriptive publications for recency (Section 3.1) and frequency (Section 3.2), after which I turn to an assessment of the accuracy of recency statements (Section 3.3) and frequency statement (Section 3.4). I end the paper with a discussion of all these findings (Section 4).

2. Materials and Methods

In this section, I first describe our prescriptive material, how I extracted recency and frequency statements from this material, and the tags I gave to these statements (Section 2.1). Following that, I explain how I tested the veracity of these statements. For recency (Section 2.2) and frequency (Section 2.3), I detail how each concept is operationalized, which statements I used for our evaluation of accuracy, and what usage data I base my evaluation on.

2.1. Prescriptive Data

To study prescriptivist statements about usage, I searched the Normative Database of Dutch. This database contains 5678 entries from prescriptivist publications aimed at mother tongue speakers of Dutch in the Netherlands for the period 1910–2018 (see van der Meulen 2020 for more details). The database focuses primarily on morphosyntactic usage items, but includes a substantial number of entries on lexis, especially for the earlier decades, as well as some entries for other linguistic levels, such as spelling or pronunciation. Entries in the database are quite heterogeneous in terms of length and contents, because their structure follows the varying build-up of the publication they originate from. Consequently, while many entries pertain to a single usage item, some contain more than one. Any usage item can attract more than one recency or frequency statement.² I approached the data using #Lancsbox (Brezina et al. 2020).

To investigate the Recency Illusion, I used a set of queries based on words denoting recency, manually cleaning up the results.³ Over the course of the investigation, it became apparent that there are in fact two types of recency statements. Next to the 'narrow' Zwickian type, which is about recent *origins* (Example 1), there are also statements that comment on a recent *development in frequency*, often by using the word *tegenwoordig* ('nowadays',

Example 2). Although this latter type of pronouncements is not strictly Zwickian, it does constitute a recency claim, and thus I included them in my initial survey of statements.

1. In dit geval gebruikt de jongere taal niet zelden het woord *ontbranden* ('In this case, recent language not uncommonly uses the word *ontbranden*') (Moortgat 1925, p. 50)
2. Per. Tegenwoordig een veel gebruikt voorzetsel: *per 1 januari* ('Per. Nowadays an often-used preposition: *per January 1st*') (Meijers 1959, p. 103)

For frequency statements, I extracted utterances from the database that referenced any kind of frequency of use.⁴ Again, I cleaned up the results, including only explicit statements about actual usage (Example 3), rather than opinions about usage, as this latter group of statements does not have to be grounded in the reality of usage (cf. Theissen 1978, p. 7). This meant that I excluded statements about how usage ought to be (Example 4), as well as statements about attitudes (Example 5).

3. *Driedubbel*: vaak gebruikt waar *drievoudig* wordt bedoeld. *Driedubbel* is zes. ('*Driedubbel*: often used where *threefold* is meant. *Driedubbel* is six') (de Raat 2012, p. 29)
4. *Waar kan nooit redengevend zijn*. ('*Waar* should never indicate a reason') (Algemeen Nederlandsch Verbond 1925, p. 4)
5. Ondertussen raakt dit gebruik van *betreffend* echter steeds meer geaccepteerd. ('Meanwhile this usage of *betreffend* is becoming more and more accepted') (Tiggeler 2001)

As I was interested in delving deeper into the patterns of recency and frequency statements, I tagged all of them for linguistic level (i.e., grammar, lexis, and spelling, etc.) and usage item (for example, comparative conjunction *als/dan* 'as/than'). In certain cases, I grouped certain usage items together. This applied, for example, to several closely related syntactic usage items about the correct combination of dependent and independent clauses. Next, as I was also interested in how prescriptivists applied recency statements according to whether they targeted recent *origins* or recent *increased frequency*, I tagged this dimension as well. I also labelled the frequency statements as *absolute* when they denoted a fixed frequency (e.g., *vaak* 'often', *soms* 'sometimes'), or *diachronic* when they referred to a development over time (e.g., *meer en meer* 'more and more', *steeds vaker* 'increasingly often'). Finally, I classified all frequency statements according to their grade. Terms could be either low, such as *soms* ('sometimes'); middle, such as *gebruikt* ('in use'); or high, such as *vaak* ('often'). For both absolute and diachronic, as well as grade, I started from the classification scheme in van der Meulen and Rutten (2022), but as the present research included many more terms, I expanded this framework where necessary.

2.2. Establishing Recency

To evaluate recency statements, I had to date the origins of the word or construction under scrutiny. Although very common, especially in lexicography, this practice nevertheless comes with a host of issues (for an overview see van der Sijs 2001, pp. 36–43), some of which complicated the current research. For example, when Moortgat (1925, p. 33fn2) questioned the recent use of *halfmens* ('halfman'), he objected not to the wordform as such, which had existed since at least the 16th century (*Woordenboek der Nederlandsche Taal*, WNT, s.v. *half*^f), but to a new meaning. Such a semantic shift was hard to distinguish from the already existing meaning of a word. A second problem was that a word could be established in specialized domains or genres, such as scientific publications or medical jargon, before it percolated to general usage. However, an early occurrence of a word in a specialized context does not prove general usage, nor does it seem feasible to expect prescriptivists to come into contact with, base themselves on, or police such specialized occurrences (unless specifically mentioned).

Because of these issues, I limited my investigation to recency statements denoting new wordforms, rather than new meanings. I dated these wordforms based on general usage only, which we can reasonably expect prescriptivists to be aware of, or have come into contact with. It is very likely that certain authors also came into contact with more specialized genres, but to take these into account would require both a more fine-grained

approach than this paper aims at, and specialized data that we do not have. I only looked at recency statements concerning recent origins (Example 1 above), rather than statements denoting an increase in frequency (Example 2 above). Although the latter type of statement does concern recency, and is thus included in the first half of this paper, the survey of statements does not fall within the strict Zwickyian interpretation of this concept. Moreover, as I also note below, Dutch lacks the proper data to investigate recent changes in frequency for usage items. From the statements that were left after these selection criteria, I randomly selected a maximum of two usage items per decade, for both lexical items and grammatical ones. In taking this coarse-grained approach, I avoided basing our conclusions on the perceptive ability of particular authors. Moreover, this approach enabled me to see whether the accuracy of recency statements showed a development over time with regard to prescriptivists' perceptions.

The data I used to pinpoint the origins of our usage items came both from primary and secondary sources. Firstly, I checked whether the earliest known example of a usage item was mentioned in one of four secondary sources: the historical *Woordenboek Nederlandsche Taal* (WNT),⁵ the *Chronologisch Woordenboek* ('Chronological Dictionary', van der Sijs 2001), the *Etymologiebank* ('Etymology bank', van der Sijs 2010), or the *Geschiedenis van het Nederlands in de twintigste eeuw* ('History of Dutch in the twentieth century', van der Horst and van der Horst 1999). Secondly, I used primary usage sources. Recent years have seen a great increase in the availability of Dutch language materials, which makes antedatings possible and even likely as compared to the linguistic sources mentioned above. For that reason, I checked results from our metalinguistic investigation in Delpher, an online repository containing over 120 million pages from newspapers, books, and magazines.⁶

After dating the words, I had to evaluate for each usage item whether the date of its origins can indeed be called recent, as compared to the date of the prescriptive recency statement. Clearly, the scope of what can be considered 'recent' depends very much on context. Set against the development of the faculty of language in humans, for example, any change that happened since the standardisation of Dutch in the 16th century is extremely recent. However, this is presumably not what language users mean when they say that a form is 'new' or 'recent'. Zwicky gave the following example of what is *not* recent for him:

Charles Hockett wrote in 1958 (*A Course in Modern Linguistics*, p. 428) about "the recent colloquial pattern *I'm going home and eat*" (...). But Hockett's belief that the construction was recent in 1958 is just wrong; David Denison, at Manchester, has collected examples from roughly 30 years before that. (Zwicky 2005)

For Zwicky, then, *recent* had to be well within 30 years, although he gives neither an explanation for this judgement nor a more exact timeframe. The fact that he stated Hockett's belief is 'just wrong' implies that it was even substantially less than 30 years. However, in the absence of any operationalisations of recency within linguistics that I was aware of, and in light of the fact that the rest of our research is also built upon Zwicky's definitions, I used this threshold value. Moreover, when we take into account that all prescriptive authors are over 30 years old, this meant that recency implies a word or construction originated within the lifetime of an author. If my data, then, showed the origins of a word or phrase to be more than 30 years before the prescriptive utterance, I considered this to be a case of the Recency Illusion.

2.3. Establishing Frequency

The results for the frequency statements presented us with similar issues as the recency statements. These included the aforementioned usage items that discussed new meanings for existing wordforms. Additionally, several syntactic types of ellipsis, such as the omission of the complementizer *om* ('for to') proved hard to study quantitatively. An additional selection criterion resulted from Zwicky's formulation of the Frequency Illusion as the belief that certain linguistic variants occur *often*. In other words, although low and middle frequency statements, such as *sometimes* or *rarely*, are statements about frequency, they do

not qualify as possible Frequency Illusions in the Zwickian sense, and thus were left out of the subsequent analysis.

Within these parameters, for lexis I again randomly extracted two usage items per decade. I matched the variant targeted by the frequency statement with its proposed counterpart or counterparts, mapping out the raw and relative frequency of both variants in a usage corpus (with relative frequency being the proportion with which a variant occurs as compared to all variants combined for a variable, expressed as a percentage). I used different corpora for different periods, depending on availability and corpus quality. For lexical usage items, I initially searched C-CLAMP, a corpus containing ± 200 million words from various magazines and newspapers published between 1837 and 1999 (Piersoul et al. n.d.). When this corpus yielded fewer than 100 combined hits for both variants for any usage item, I again used Delpher. It is important to note that this dataset unfortunately has some serious limitations with regard to quantitative research, as its OCR quality is quite low for certain time periods, and its representativeness is questionable (see van der Sijs 2019). Still, as I base my conclusions on relative frequency, this becomes less of an issue, although it has to be taken into account when interpreting the results. As with the recency statements, I again use a time window of 30 years prior to the statement. So, when Moortgat calls *duistering* ('darkness') the "usual" variant as opposed to *duisternis* in 1925 (Moortgat 1925, p. 48), I look at occurrences of both variants in the 1896–1925 segment of C-CLAMP.

For grammatical usage items, establishing the frequency of two variants in a certain time period of corpus data presented us with problems. The polysemous nature of many grammatical words, coupled with their high occurrence, would require more cleaning-up effort than I could achieve given the overall scope of the present paper, even after sampling. To still be able to test some grammatical usage frequency statements, I based my study on data as reported in (Van der Meulen n.d.). For this paper, spoken and written usage data were collected and cleaned-up for nine morphosyntactic variables, in order to test the distribution of variants against prescriptive evaluations (i.e., whether a variant was acceptable) and attitudinal data. I used either written, spoken, or all data from this dataset, depending on the scope of the statement. The usage data comprise the period 1995–2004, and thus I looked at frequency statements from roughly that period (i.e., 1990–2015).

To evaluate whether any result constituted a Frequency Illusion, I turned to a survey of Dutch probability and frequency terms and phrases, as reported in Willems et al. (2020). Conveniently, they asked participants to give numerical interpretations of their target phrases in percentages. Thus, for example, *vaak* ('often') has a mean numerical interpretation of 73%, although there is considerable variability in the answers (Willems et al. 2020, p. 9). However, as very few of the high frequency terms found in our data were included by Willems, Albers, and Smeets, I only used the result for *vaak*, arguing that higher frequency terms such as *heel vaak* ('very often') can be expected to at least entail this meaning. In other words, taking this relatively low threshold, I am being lenient towards prescriptivists' judgments. In terms of the Frequency Illusion, this approach meant that if less than 73% of the combined variants for a given usage item constitute the one that the prescriptive author targets, this constitutes a case of the Frequency Illusion. Taking our example from above, the disapproved variant *duistering* occurs 8 times in our data, the approved variant *duisternis* 909 times. Relatively speaking, then, *duistering* only occurred in 0.9% of all 917 combined variants. This is clearly far below our threshold of 73%, and we can thus conclude that this is an example of the Frequency Illusion.

3. Results

First, I discuss the results of our prescriptive survey for recency terms (Section 3.1) and frequency terms (Section 3.2). After that, I zoom in on the comparison with usage, looking at lexis and grammar separately, again first for recency statements (Section 3.3) and then frequency statements (Section 3.4).

3.1. Prescriptive Statements about Recency

Our data showed 238 entries that contain at least one recency statement. Most entries (214, 89.9%) were aimed at a single usage item; 24 entries discussed more than one. For example, Charivarius listed eight *bruikbare nieuwe Nederlandse woorden* ('useable new Dutch words', Charivarius 1940, p. 73). By far the most extreme cases were two series of *modewoorden* ('fashionable words'), both from the same work by Haje (1932), which contained 40 and 24 words, respectively. In total then, this leads to 341 usage items being targeted. The majority of these pertained to lexis (230 or 67.6%). Grammar comprises a much smaller proportion (98 or 28.8%); the final twelve items concerned pronunciation, stylistics, or spelling. This distribution was particularly striking when we take into account that the database as a whole focused more on morphosyntax. It seems then, that lexical items are more likely to attract recency statements.

Looking at which particular usage items were mentioned, we saw that the 341 usage items belonged to 307 types (TTR = 0.9). A total of 283 items were mentioned only once (83.3%); the remaining 58 belonged to 22 types (see Table 1). The two types mentioned most often (five occurrences each) were both morphosyntactic: the relative pronouns *dat/wat* ('that/which') and number agreement *een aantal N is/zijn* ('a number of N is/are'). The usage items that were mentioned more than once usually stemmed from different years. The largest time difference, 64 years, was found with regard to *N dat/wat* ('N that/which'), for which the first recency statement stemmed from 1932, and the last from 1996. This difference in itself makes Recency Illusions inevitable: if particular variation already existed in 1932, it cannot be a recent phenomenon in the 1990s.

Table 1. Usage items which are mentioned more than once with recency statements.

Usage Item	Translation	Level	No. of Recency Statements	Year(s) of Recency Statements
<i>aantal N is/zijn</i>	'a number of N is/are'	grammar	5	1979, 1994, 2000, 2001, 2002
<i>N dat/wat</i>	'N that/which'	grammar	5	1932, 1963, 1990, 1996, 1996
<i>als/dan</i>	'as/than'	grammar	4	1984, 1994, 1994, 2006
<i>trappen van vergelijking</i>	comparative/superlative formation	grammar	3	1940, 1946, 1994
<i>omdat/doordat</i>	causal conjunctions	grammar	3	1962, 1994, 1996
<i>nieuwbouw</i>	'new construction'	lexis	3	1932, 1942, 1964
<i>grootstad</i>	'big city'	lexis	3	1932, 1942, 1964
<i>belevenis</i>	'experience'	lexis	3	1932, 1964, 1978
<i>woordgeslacht</i>	noun gender	grammar	3	1993, 1993, 1994
11 items	N/A	lexis	2	N/A
3 items	N/A	grammar	2	N/A

To see whether there was any development in the use of recency statements, I looked at the distribution of recency judgements over time (see Table 2). Looking first at usage items ($n = 341$), the highest number of statements was found in the 1930s, which was largely caused by the two wordlists mentioned before. The 1990s was the second-highest scoring decade, as it had 50 usage items attracting recency statements. However, as the number of prescriptive publications varied greatly over time in our database, we cannot draw conclusions based on these raw usage items frequencies. Instead, we should look at relative occurrences per decade as compared to the total database. Unfortunately, as we do not know the exact number of usage items for every entry in the whole database, I had to base our study here on the number of entries, which we do know. Looking at entries containing recency statements ($n = 238$) gives a somewhat different view. Now, the 1970s show the highest proportion of entries containing at least one recency statement (11.3%). The lowest percentage is in the 2010s (1%). Furthermore, there seems to be a declining tendency to use recency statements, as earlier decades show higher proportions than later. This tendency is not linear, but the largest outlier, the 1970s, is due to the predilection of one particular author to use such statements.⁷ The same goes for the 1990s, albeit to a lesser extent. Although the preferences of particular authors influence the presence of statements, we can conclude that the use of recency statements does proportionally decline over time.

Table 2. Recency statements per decade as raw frequency and percentage of total amount of entries.

Decade	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
No. of usage items with recency statements ($n = 341$)	0	18	97	42	29	35	44	8	50	15	4
No. of entries with recency statements ($n = 238$)	0	16	38	32	26	22	38	6	43	13	4
Total no. of entries in database ($n = 5678$)	216	193	563	396	531	544	336	524	1082	887	406
% of entries that contain recency statements compared to total no. of entries in database	0.0	8.3	6.7	8.1	4.9	4.0	11.3	1.1	4.0	1.6	1.0

Lastly, the division of recency statements into origins and recent development showed that the former, strictly Zwickyian remarks comprise the majority of statements (80.9%, 276). Almost all statements that comment on a development in usage signal an increase; only in a few isolated cases do writers indicate that a variant is losing ground. This preference points towards the rhetorical nature of such statements: after all, no-one needs to be ‘warned’ about a disapproved variant disappearing.

3.2. Prescriptive Statements about Frequency

Our material showed that 1249 entries contain at least one mention to frequency. Frequency statements are thus far more prevalent in Dutch prescriptive publications than recency statements. In total, these 1249 entries contained 1786 frequency statements. A difference with the recency statements was that the majority of frequency statements pertained to grammar (1064, 59.5%), while 36.8% of statements was found in relation to lexis. Again, a small number of 67 cases referred to other levels, such as spelling and pronunciation. This distribution seems to be more in line with the general contents of the database. Another notable difference between the recency and frequency utterances was that the latter were distributed over a proportionally much smaller number of different usage types, namely 601 (TTR = 0.33). Moreover, for the frequency items, just 400 (66.6%) of the types were found only once.

Just as with the recency statements, those particular items that attract the most frequency statements are grammatical in nature (see Table 3). However, for the frequency statements this effect is much stronger. In fact, when we look at all 33 types that occur more than ten times, only one type, verbal formation, refers to lexis, and then only partly. Another noteworthy property of the items in Table 4 is the fact that the top five most frequently found usage items in the 20th and 21st century (*hen/hun, als/dan, dat/wat, tante betje* and congruence, see [van der Meulen 2021](#), p. 170) are also among those items that attract the most frequency statements. This may show that frequency statements are at least partially related to the prescriptive canon. By contrast, however, two of the other types in this top-10, namely *zinsbouw* and *beknopte bijzin*, relate to quite complex syntactic phenomena, which are generally rare in prescriptive publications (cf. [Chapman 2021](#)).

Table 3. Linguistic items mentioned more than once with recency statements.

	Usage Item	English Translation	Level	No. of Times Mentioned
1.	woordgeslacht	noun gender	grammar	88
2.	hen/hun	3rd person personal pronoun direct and indirect object	grammar	77
3.	als/dan	comparative conjunctions	grammar	66
4.	dat/vat	relative pronouns	grammar	61
5.	zinsbouw	sentence structure	grammar	39
6.	tante betje	word order in subordinate clauses	grammar	37
7.	congruentie	agreement	grammar	36
8.	beknopte bijzin	participle clauses	grammar	30
9.	trappen van vergelijking	degrees of comparison	grammar	27
10.	omdat/doorat	causal conjunctions	grammar	26

Table 4. Frequency statements per decade as raw frequency and percentage of total amount of entries.

Decade	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
No. of usage items with recency statements	0	66	74	211	153	283	155	121	340	277	107
No. of entries with recency statements	0	53	59	132	82	195	108	94	233	214	79
Total no. of entries in database (<i>n</i> = 5678)	216	193	563	396	531	544	336	524	1082	887	406
% of entries that contain recency statements compared to total no. of entries in database	0.0	27.5	10.5	33.3	15.4	35.8	32.1	17.9	21.5	24.1	19.5

The patterning for the development of frequency statements over time, again looking at entries rather than specific items (*n* = 1249), is more haphazard than for recency statements (see Table 4). There is another cut-off point between the 1970s and 1980s, but the decline in the last decades is less pronounced, and the presence of frequency terms remains higher than for recency terms. Finally, the influence from specific authors is more marginal. One exception is B. Cees Damsteegt, who uses frequency statements in 102 entries in his *In de Doolhof van Het Nederlands* (Damsteegt 1964). These comprise 52.3% of all mentions in this decade, but this percentage is actually in line with Damsteegt’s overall presence in that time period: the entries from his detailed guide comprise 51.5% of all the 544 entries for the 1960s.

As I was interested in the way prescriptivists evaluate usage, I next looked at the specific frequency terms, and the distribution among categories of frequency term and degree (Table 5). As for type, the top-10 contained seven absolute frequency terms and three diachronic ones. This was an underrepresentation of absolute frequency terms, as they comprise 85.3% (1524 tokens) of all terms. I also found seven high degree frequency terms, which is an overrepresentation as compared to the presence of such terms in total, as high degree frequency terms comprise 59% (1054 tokens), as opposed to middle (331, 18.5%), and low terms (402, 22.5%). Still, it was clear that prescriptive authors favour high frequency terms. This may be another sign of the rhetorical nature of such terms: the sense of urgency is lower for low frequency terms, which defies the purpose of prescriptive publications.

Table 5. Top-10 most frequently used frequency terms.

	Frequency Term	English Translation	Category	Degree	No. of Times Mentioned
1.	<i>vaak</i>	'often'	absolute	high	317
2.	<i>soms</i>	'sometimes'	absolute	middle	107
3.	<i>steeds vaker</i>	'increasingly more often'	diachronic	high	51
4.	<i>wel eens</i>	'occasionally'	absolute	low	51
5.	<i>veel</i>	'much/many'	absolute	high	47
6.	<i>gebruikt</i>	'used'	absolute	middle	36
7.	<i>vooral</i>	'mostly'	absolute	high	36
8.	<i>steeds meer</i>	'increasingly'	diachronic	high	35
9.	<i>vaker</i>	'more often'	diachronic	high	33
10.	<i>dikwijls</i>	'often'	absolute	high	32

3.3. Recency and Usage

Of the 23 recency statements I investigated, 17 lexical usage items are instances of the Recency Illusion, which is by far the majority of statements (20, 87%). Just three recency statements are not Illusions (see Table 6). These exceptions are *mond-aan-mondreclame* ('mouth-on-mouth marketing'), for which there was only six years between the recency statement and the first observed occurrence in usage; *hoogspanning* ('high voltage'), for which the time gap was 25 years (a later statement claiming the recency of this word from 1978 is a Recency Illusion according to our definition), and *mentaliteit* ('mentality'), for which the time gap is exactly 30 years. However, these results highlight the arbitrariness of our definition: *mentaliteit* just makes the cut, while *belevenis* ('experience'), for which the time gap is 34 years, is only an illusion by a small margin. However, such occurrences are actually quite rare: for 17 of the statements, the earliest occurrence of a word or phrase predates the recency statement by more than 50 years.

Table 6. Lexical usage items with year(s) of prescriptive utterance, earliest occurrence in usage, and evaluation with regard to Recency Illusion.

Usage Item	English Translation	Year(s) of Recency Statement	Source ⁸	Illusion?
<i>toonkunstenaar</i>	'musician'	1925	1756	yes
<i>mentaliteit</i>	'mentality'	1932	1902	no
<i>vandaag de dag</i>	'today'	1940	1895	yes
<i>nieuwbouw</i>	'new housing estate'	1932, 1942, 1964	1859	1932: yes; 1942: yes; 1964: yes;
<i>halfbroeder</i>	'half-brother'	1932, 1964	1748	1932: yes; 1964: yes
<i>grootmacht</i>	'superpower'	1932, 1964	1871	1932: yes 1964: yes
<i>hoogspanning</i>	'high-voltage'	1932, 1978	1907	1932: no 1978: yes
<i>belevenis</i>	'experience'	1932, 1964, 1978	1898	1932: yes; 1964: yes; 1978: yes
<i>witlof</i>	'chicory'	1959	1854	yes
<i>festival</i>	'festival'	1959	1872	yes
<i>emballage</i>	'packaging'	1980	1745	yes
<i>uitonderhandelen</i>	'negotiate'	1989	1955	yes
<i>mond-aan-mondreclame</i>	'word-of-mouth advertising'	1995	1989	no
<i>met behulp van + person</i>	'with the help of' + person	2001	1857	yes
<i>in de loop der tijd</i>	'over time'	2008	1946	yes
<i>uitboeien</i>	'booing'	2012	1966	yes

Within our selection criteria, there were few recency statements that commented on the origins of a particular grammatical variant, and that could be checked. Still, for those

13 that did occur for eight usage items, the picture is much the same as for lexical items: they are almost all Recency Illusions (see Table 7). The only possible exception in our data is relative pronoun *wat*, which, according to Haje, is *tegenwoordig* ('at present') frequently used in written language following neuter nouns (Haje 1932, p. 118). Although the general development of d-forms to w-forms is well-established (see, for example, Schoonenboom 2000), relatively little is known about the particular progression of this phenomenon with regard to neuter nouns (cf. Rutten 2020). However, as van der Horst and van der Horst posit that the use of the disapproved variant *wat* must have been limited to spoken language until approximately 1900 (van der Horst and van der Horst 1999, p. 170), I evaluate this case then as *maybe* constituting a Recency Illusion.

An interesting difference with the lexical items is that the time gaps are much larger for grammatical items. Several of the usage items go back to the early days of Dutch normative writing. For example, when Heldring complained in 1993 that many neuter proper nouns for countries and cities were increasingly being referred to by using feminine possessive pronouns (for example, *Amsterdam en haar grachten*, 'Amsterdam and her canals'), this echoes a very similar comment made by Van Hoogstraten in 1700 (quoted in van der Sijs 2021, p. 402). On the one hand, it seems unlikely that Heldring, a journalist, would have read a grammar book from 300 years ago. On the other hand, the referential practice of using feminine possessive pronouns is a staple of prescriptive publications, dating back to at least the 1930s.⁹ With regard to these grammatical cases, then, Zwicky's complaint that "if only they'd thought to consult some standard sources or look at some facts" seems warranted (Zwicky 2005).

Table 7. Grammatical usage items with year(s) of prescriptive utterance, earliest occurrence in usage, and evaluation with regard to Recency Illusion.

Usage Item	Recency Statement	Source ¹⁰	Illusion?
relative pronouns <i>dat/wat</i>	1932, 1990	1930s	1932: maybe; 1990: yes
periphrastic comparative/superlative complementizer <i>om</i>	1940, 1994	1890	1940: yes; 1994: yes
article ellipsis before certain nouns (e.g., <i>ondergetekende</i> 'undersigned')	1948	1898	yes
causal conjunctions <i>omdat/door</i>	1959	1900	yes
comparative conjunctions <i>als/dan</i>	1962, 1994, 1996	1860	1962: yes; 1994: yes; 1996: yes
noun gender	1984	17th century	yes
<i>hun</i> -subject	1993, 1994	1700	1993: yes; 1994: yes
	2000	1911	yes

3.4. Frequency and Usage

Within our parameters, I found two checkable lexical examples in all decades except the 1920s and 1950s, for a total of 18 (see Table 8). Our data from the lexical variants showed that all but one of our examples constituted Frequency Illusions. The only exception was *meer of mindere mate* ('a greater or lesser extent'), for which the condemned variant comprised 76.7% of all variants. In seven of the cases, the proportion of targeted variants comprises less than 5% of all variants. Noteworthy is the case of compounds ending in *-toename* (increase), which is targeted by Grauls (1957, p. 197) in favour of the proposed alternatives ending in *-toeneming*. In our data, the disapproved variants occur in almost half of all cases (47.2%), making this a Frequency Illusion. However, data from the present day showed that the disapproved variant has all but disappeared, as I found 64 hits for *-toeneming*, against 8767 for *-toename*. So, while the statement was not correct at the time, Grauls was perhaps aware of a change in progress.

Table 8. Lexical usage items with year(s) of prescriptive utterance, absolute and relative frequency counts of condemned variant in usage, and evaluation with regard to Frequency Illusion.

Condemned Variant	Proposed Alternative(s)	English Translation	Year of Meta Utterance	Absolute Frequency of Condemned Variant ¹¹	Relative Frequency in % of Condemned Variant	<73%
<i>duistering</i>	<i>duisternis</i>	‘darkness’	1925	8	0.9	yes
<i>(on)toelaatbaar</i>	<i>(on)geoorloofd</i>	‘impermissible’	1932	73	17.9	yes
<i>jongeman</i>	<i>jonkman</i>	‘young man’	1935	81	38.8	yes
<i>onmeedogenloos</i>	<i>meedogenloos</i>	‘merciless’	1940	49	3.5	yes
<i>zodoende</i>	<i>dus, daarom</i>	‘thusly’	1941	20	0.1	yes
<i>-toename</i>	<i>-toeneming</i>	‘increase’	1957	50	47.2	yes
<i>vanwege</i>	<i>wegens</i>	‘because of’	1962	186	34.0	yes
<i>zwempool</i>	<i>zwembad</i>	‘swimming pool’	1964	69	0.1	yes
<i>ons inziens</i>	<i>onzes inziens</i>	‘in our opinion’	1975	7791	65.7	yes
<i>begeesterd</i>	<i>verrukt, geestdriftig</i>	‘enraptured’	1978	4	1.2	yes
<i>verplichtend</i>	<i>verplicht</i>	‘obliged’	1980	16	2.5	yes
<i>ik mankeer</i>	<i>mij mankeert</i>	‘I lack’	1986	205	58.6	yes
<i>middels</i>	<i>door middel van</i>	‘by means of’	1991	328	25.5	yes
<i>behartenswaardig</i>	<i>behartigenswaardig</i>	‘worthy of consideration’	1997	146	56.6	yes
<i>overnieuw</i>	<i>opnieuw</i>	‘all over again’	2000	552	0.1	yes
<i>meer of mindere mate</i>	<i>meerdere of mindere mate</i>	‘a greater or lesser degree’	2005	1457	76.7	no
<i>scherpst van de snede</i>	<i>scherp van de snede</i>	‘sharpest part of the knife’ (idiom)	2013	297	47.9	yes
<i>zei af</i>	<i>zegde af</i>	‘cancelled’	2013	25	27.5	yes

Of the nine grammatical usage items under investigation, seven occurred in our data with high frequency statements, sometimes multiple times (see Table 9). For two of these usage items, *u heeft* (‘you-FORM have’) and *hele mooie auto* (‘very-INFL nice car’), the proportion of condemned variants is above our threshold of 73%. Thus, the three frequency statements I find for these usage items are not Frequency Illusions. However, these correct assessments are the minority, as the other 21 statements, distributed over five variants, show relative frequencies below the threshold. Thus, as with the lexical variants, by far the majority of statements can be classified as Frequency Illusions.

Table 9. Grammatical usage items with year(s) of prescriptive utterance, absolute and relative frequency counts of condemned variant in usage, and evaluation with regard to Frequency Illusion.

Condemned Variant	Proposed Alternative(s)	Year of Meta Utterance	Absolute Frequency of Condemned Variant ¹²	Relative Frequency in % of Condemned Variant	<73%
<i>person waarvan</i>	<i>person van wie</i>	1999, 2000, 2001, 2001	120	18.6	yes
<i>Neuter noun wat</i>	<i>Neuter noun dat</i>	1996, 1998, 2001, 2007, 2009, 2013	339	7.4	yes
<i>subject-hun</i>	<i>subject-zij</i>	1994, 2000, 2008, 2011	197	2.7	yes
<i>een aantal mensen zijn</i>	<i>een aantal mensen is</i>	1994, 1996, 2000, 2005, 2007, 2011	465	60.7	yes
<i>je kan</i>	<i>je kunt</i>	1993	1792	28.6	yes
<i>u heeft</i>	<i>u hebt</i>	2009	827	74.1	no
<i>hele mooie auto</i>	<i>heel mooie auto</i>	1993, 2011	2482	79.7	no

Of particular interest was one of the frequency statements made for relative pronoun *wat*. Houët (2000, p. 249) claimed that in spoken language this variant is used *vrijwel uitsluitend* (‘almost exclusively’). When I extracted the spoken part out of the data used,

I observed that, although with 35.8% it was much higher than the combined written and spoken data reported above, it was far from used exclusively. In fact, it was not even the dominant variant. This label *uitsluitend* ('exclusively') was quite rare in our data, and I found only two other instances. In 1962, Heidebuchel claimed that *omdat* ('because') is used *vrijwel uitsluitend*, to the detriment of alternative *doordat* (Heidebuchel 1962, p. 144). However, in C-CLAMP (1932–1962) I found 832 examples of *doordat* (8.8% of all variants). Similarly, Weverink stated that *vrijwel iedereen* ('almost everybody') used *ze* ('them') in spoken language to indicate direct or indirect object, rather than *hen* or *hun* (Weverink 2012, p. 52). In this case, I found over 2300 instances of *hen* and *hun* in these functions in the Corpus Gesproken Nederlands ('Corpus of Spoken Dutch', see Oostdijk 2000). Both statements were thus clearly illusions. Arguably, these examples embodied a 'worse' Frequency Illusion than the use of *vaak*, as *uitsluitend* is much more categorical. This was also shown in the research by Willems et al.: the interpretation of more categorical words, such as *always*, had a much smaller range of answers than that of high frequency words such as *often* (Willems et al. 2020, p. 10). These examples highlight the rhetorical dimension of the use of such extreme frequency statements, but also the risk of their use.

4. Discussion

Our research shows that recency and frequency statements are part of the modern Dutch prescriptive tradition, but in different ways, and to different degrees. Recency statements occur mostly in relation to lexis, and seem to be falling out of fashion over the course of the twentieth century. Although, as I mentioned, this may partly be due to the design of our database, it may also be an effect of the solidification of the prescriptive canon. Tieken-Boon van Ostade notes this development for the English usage guide tradition when she says that "many usage problems came to have a remarkably stable presence" (Tieken-Boon van Ostade 2020); the same has been shown for Dutch (van der Meulen 2021). This lack of innovation in the genre presumably diminishes the need and even the possibility for variation to be perceived as recent. Thus, when recency statements are used, they are often illusions, but the fact that prescriptivists use fewer of them over time points to some awareness on their part with regard to 'canonised' usage items.

Frequency statements differ in three ways from recency statements in our data: they are far more widespread, are mostly used for grammatical usage items, and do not show a decline over time. It is noteworthy that frequency statements regularly occurred with the 'old chestnuts' of Dutch prescriptivism, those cases of (grammatical) variation that have been targeted for decades or even centuries (van der Meulen 2021). This implies that we are seeing a rhetorical dimension of prescriptivism: frequency statements *belong* to the prescriptive genre, partly because of tradition, partly to keep on creating a sense of urgency. The fact that there is a tendency to use high frequency terms further supports this hypothesis. More in-depth qualitative research should investigate to what extent prescriptive authors genuinely believe their frequency statements, and to what extent they use them, consciously or not, as rhetorical devices.

For the recency statements, the fact that most of them target unique linguistic variants initially pointed towards these statements possibly correctly identifying new variation. However, as our investigation showed, this was not the case: most of them constituted Zwickian illusions. Why the type:token ratio was so high remains unclear. One explanation for the statements that did occur more than once could again be intertextuality. The example of *grootmacht*, which was correctly identified as recent in 1932, according to our definition of recency, was inaccurately stated in 1964. Perhaps this later claim was not a genuine observation, but simply copied from the earlier work. As for the scope of our investigation, our results could be skewed by the fact that I did not look at recency statements targeting new meanings. However, it seems unlikely that prescriptivists would be able to detect sometimes very subtle shifts more accurately than more straightforward wordform occurrences. The same applies to statements observing recent developments in frequency: there is no reason to assume that this arguably more difficult type of observation

would yield a higher degree of accuracy. Still, such observations would be worthwhile exploring further, given the availability of adequate usage corpora.

Almost all frequency statements, both for lexis and grammar, were also illusions. Although for grammatical usage items, our research was admittedly limited in scope, both for type and temporal coverage, the results were similar across the board. Moreover, while we may have expected earlier prescriptivists to be inaccurate in their assumptions because they lacked available data, this becomes less of an excuse for later periods, when increasing numbers of data sources become available. Nevertheless, the absence of a panoptic view, as Zwicky calls it, or more specifically the extrapolation of the personal linguistic experience to a language as a whole, could be the main factor explaining these results. One way of further testing this is by attempting to better match the linguistic reality of specific prescriptivists to the statement. An example of this can be found in the aforementioned [van der Meulen and Rutten \(2022\)](#), who, in their investigation of frequency terms used in a Dutch normative work from the nineteenth century, are able to approximate the type of language targeted by a particular prescriptivist, Matthijs Siegenbeek. Their results deviate from the present conclusions, in that the prescriptivist work under scrutiny does show broadly correct assumptions about frequency. Whether such an approach would yield different insights for the prescriptivists under investigation here remains the question. The difference in results may also be an effect of the time period: it seems possible that, with linguistic communities being smaller in the nineteenth century, the personal exposure to language of Siegenbeek was more representative of the language as a whole.

The present research provides solid empirical evidence for the existence of the Recency and Frequency Illusions. Of course, this by no means should be taken as evidence that all prescriptivists, let alone all language users, are always ‘wrong’ in their assessments of recency and frequency. Much more research should and can be conducted to further our understanding of the circumstances in which people make such evaluations. From a methodological standpoint, we consider our approach to be viable, and well worth extending further. Care should be taken with using the present research design however, as the interpretation of seemingly similar frequency terms may differ between different languages ([Willems et al. 2020](#), p. 11). More generally, as much depends on the definitions of recency and frequency, it is important to further delve into what language users actually mean when they use these terms. We saw that sometimes a recency statement explicitly noted a recent increase, rather than a recent origin. Perhaps this is implicitly meant by other recency statements as well. Similarly, both types of statements may accurately observe the emergence or occurrence of a variant in a new genre, even when this is not mentioned explicitly. Perhaps more fine-grained qualitative work would show that prescriptivists do have better antennas for evaluating usage, given different parameters. However, when we take prescriptive statements at face-value, as I did in the present paper, and as language users perusing the prescriptive publications cannot help but do, we have to conclude that Zwicky was right: we are indeed illuded.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Most usage data are freely available online; this is mentioned in the paper for different datasets. The Corpus Gesproken Nederlands ‘Corpus of Spoken Dutch’ is available through the OpenSoNaR interface ([OpenSonar 2020](#)). The prescriptive data cannot be made available for copyright reasons.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Sources for Lexical Variants Mentioned with Recency Statements

All newspapers were approached through the Delpher interface ([Delpher 2021](#)). All magazines and books were found using the *Digitale Bibliotheek voor de Nederlandse Letteren*

‘Digital Library for Dutch Literature’ (DBNL—Digitale Bibliotheek voor de Nederlandse Letteren 2021).

Table A1. Lexical usage items with source for usage dating and type of source.

Usage Item	Recency Statement	Source
<i>toonkunstenaar</i>	WNT, s.v. <i>toonkunstenaar</i>	secondary, dictionary
<i>mentaliteit</i>	<i>Dagblad van Zuidholland en 's Gravenhage</i> , 23 June 1902	primary, newspaper
<i>vandaag de dag</i>	<i>Algemeen Handelsblad</i> , 9 October 1895	primary, newspaper
<i>nieuwbouw</i>	<i>Nieuw Amsterdamsch handels-en effectenblad</i> , 23 February 1859	primary, newspaper
<i>halfbroeder</i>	<i>Leydse courant</i> , 3 October 1748	primary, newspaper
<i>grootmacht</i>	<i>Bataviaasch handelsblad</i> , 17 July 1871	primary, newspaper
<i>hoogspanning</i>	WNT, s.v. <i>vereffenen</i>	secondary, dictionary
<i>belevenis</i>	<i>Provinciale Drentsche en Asser courant</i> , 12 January 1898	primary, newspaper
<i>witlof</i>	WNT, s.v. <i>lof</i>	secondary, dictionary
<i>festival</i>	WNT, s.v. <i>festival</i>	secondary, dictionary
<i>emballage</i>	Etymologiebank, s.v. <i>emballage</i>	secondary, etymological dictionary
<i>uitonderhandelen</i>	<i>Algemeen Indisch dagblad: de Preangerbode</i> 22 October 1955	primary, newspaper
<i>mond-aan-mondreclame</i>	Vonk, 1989	primary, magazine
<i>met behulp van + persoon</i>	<i>Dietsche Warande</i> , 1857, p. 121	primary, magazine
<i>in de loop der tijd</i>	Niko Tinbergen (1946) <i>Inleiding tot de diersociologie</i> . p. 139	primary, non-fiction book
<i>uitboeën</i>	<i>Het vrije volk: democratisch-socialistisch dagblad</i> 7 October 1966	primary, newspaper

Appendix B

Table A2. Grammatical usage items with source used for evaluation against Recency Illusion and type of source.

Usage Item	Source	Source Type
relative pronouns <i>dat/wat</i>	(van der Horst and van der Horst 1999, p. 170)	secondary, linguistic work
periphrastic comparative/superlative	<i>Bataviaasch Handelsblad</i> , 9 April 1890	primary, newspaper
complementizer <i>om</i>	<i>Dagblad van Zuidholland en 's Gravenhage</i> , 14 March 1898	primary, newspaper
article ellipsis before certain nouns (e.g., <i>ondergetekende</i> ‘undersigned’)	<i>Leeuwarder Courant</i> , 1 January 1900	primary, newspaper
causal conjunctions <i>omdat/doordat</i>	H. Kern (1860), <i>Handleiding bij het onderwijs der Nederlandsche taal</i> , quoted in (van der Sijs 2021, p. 485)	secondary, educational work
<i>als/dan</i>	(van der Sijs 2021, pp. 468–71)	secondary, linguistic work
noun gender	David van Hoogstraten (1700), <i>Aenmerkingen over de geslachten der zelfstandige naamwoorden</i> , quoted in (van der Sijs 2021, p. 402)	secondary, grammar
<i>hun</i> -subject	(vor der Hake 1911, p. 20)	secondary, linguistic work

Appendix C

Table A3. Lexical usage items with year(s) of prescriptive utterance, absolute and relative frequency counts of condemned and approved variants in usage, and source of the usage data.

Condemned Variant	Proposed Alternative (s)	Absolute Frequency of Condemned Variant	Absolute Frequency of Proposed Alternative	Relative Frequency in % of Condemned Variant	Relative Frequency in % of Proposed Alternative	Source Usage Data
<i>duistering</i>	<i>duisternis</i>	8	909	0.9	99.1	C-CLAMP 1896–1925
<i>(on)toelaatbaar</i>	<i>(on)geoorloofd</i>	73	334	17.9	82.1	C-CLAMP 1903–1932
<i>jongeman</i>	<i>jonkman</i>	81	128	38.8	61.2	C-CLAMP 1906–1935
<i>onmeedogenloos</i>	<i>meedogenloos</i>	49	1346	3.5	96.5	Delpher 1910–1940
<i>zodoende</i>	<i>dus, daarom</i>	20	18,639	0.1	99.9	C-CLAMP 1912–1941
<i>-toename</i>	<i>-toeneming</i>	50	56	47.2	52.8	C-CLAMP 1928–1957
<i>vanwege</i>	<i>wegens</i>	186	361	34.0	66.0	C-CLAMP 1933–1962
<i>zwempeel</i>	<i>zwembad</i>	69	91,854	0.1	99.9	Delpher 1935–1964
<i>ons inziens</i>	<i>onzes inziens</i>	7791	4074	65.7	34.3	Delpher 1946–1975
<i>begeesterd</i>	<i>verrukt, geestdriftig</i>	4	330	1.2	98.8	C-CLAMP 1949–1978
<i>verplichtend</i>	<i>verplicht</i>	16	615	2.5	97.5	C-CLAMP 1951–1980
<i>ik mankeer</i>	<i>mij mankeert</i>	205	145	58.6	41.4	Delpher 1957–1986
<i>middels</i>	<i>door middel van</i>	328	957	25.5	74.5	C-CLAMP 1972–1991
<i>behartenswaardig</i>	<i>behartigenswaardig</i>	146	112	56.6	43.4	Delpher 1978–1997
<i>overnieuw</i>	<i>opnieuw</i>	552	554,358	0.1	99.9	Delpher 1981–2000
<i>meer of mindere mate</i>	<i>meerdere of mindere mate</i>	1457	443	76.7	23.3	Delpher 1986–2005
<i>scherpst van de snede</i>	<i>scherp van de snede</i>	297	323	47.9	52.1	Delpher1994–2013
<i>zei af</i>	<i>zegde af</i>	25	66	27.5	72.5	Delpher1994–2013

Appendix D

Reported usage data combine results for each variable from the Corpus Gesproken Nederlands and the 1995–1999 section of C-CLAMP.

Table A4. Grammatical usage items with absolute and relative frequency counts of condemned and approved variants in usage.

Condemned Variant	Proposed Alternative (s)	Absolute Frequency of Condemned Variant	Absolute Frequency of Proposed Alternative	Relative Frequency in % of Condemned Variant	Relative Frequency in % of Proposed Alternative
person + <i>waarvan</i>	person + <i>van wie</i>	120	524	18.6	71.4
Neuter noun <i>wat</i>	Neuter noun <i>dat</i>	339	4216	7.4	82.6
<i>hun hebben</i>	<i>zij hebben</i>	197	7,26	2.7	97.3
<i>een aantal mensen zijn</i>	<i>een aantal mensen is</i>	465	301	60.7	39.3
<i>je kan</i>	<i>je kunt</i>	1792	4481	28.6	71.4
<i>u heeft</i>	<i>u hebt</i>	827	289	74.1	25.9
<i>hele mooie auto</i>	<i>heel mooie auto</i>	2482	634	79.7	20.3
person + <i>waarvan</i>	person + <i>van wie</i>	120	524	18.6	81.4

Notes

- There is, however, a problematic gap between the frequency statement, which was made in 1992 (van Gessel et al. 1992, p. 62), and the usage data, which stem from 2011. As such, the claim that this frequency statement is correct should be approached critically. We will return to this particular case later on in the paper.
- As the database is still being processed, the exact distribution between the different linguistic levels cannot be given at the time of writing. Similarly, we do not know how many usage items all the entries contain.
- The complete list of recency queries was: *tegenwoordig, nieuw*, sinds, recent*, laatste tijd, modern*, hedendaag*, jong**. After these, a sample check of other frequency terms did not produce new results.

- ⁴ The complete list of frequency queries was: *gebruik*, komt *voor*, komt voor, tref*, vaak, vaker, steeds, soms, af en toe, geregeld, regelmatig, zelden, nooit, veelvoorkomend*, wel eens, weleens, geijkt, zwang, vroeger, tegenwoordig, spreektaal, schrijftaal, volkstaal, taalwerkelijkheid. After these, a sample check of other frequency terms did not produce new results.
- ⁵ Available at <https://gtb.ivdnt.org/search/?owner=wnt> (accessed on 23 November 2021).
- ⁶ For more information see <https://www.delpher.nl/over-delpher/wat-zit-er-in-delpher/wat-zit-er-in-delpher#7b8c9> (accessed on 23 November 2021).
- ⁷ In fact, the work of *Germanismen in het Nederlands* (1978) by Siegfried Theissen is quite an outlier in terms of prescriptive publications in general. This book is a reworking aimed at a general audience of Theissen's PhD thesis, for which he investigated how dictionaries dealt with Germanisms, comparing their treatment to usage.
- ⁸ See Appendix A for references to the specific sources, as well as the type of source (e.g., primary or secondary, newspaper, and dictionary etc.). Moreover, it is very well possible that earlier examples can be found for many of these cases. However, as finding the earliest occurrence was not the goal of the current research, we stopped searching when we found an example in general usage that invalidated the recency statement. These dates can thus best be interpreted as *at least as early as*, as is normal in lexicography (van der Sijs 2001, p. 41).
- ⁹ Ironically, the earliest prescriptive publication in our database that references this usage item is a style guide published by national newspaper NRC in 1935. Heldring's prescriptive publication was a collection of columns he wrote for this very newspaper.
- ¹⁰ See Appendix B for references to the specific sources, as well as the type of source (e.g., primary or secondary, newspaper, dictionary etc.). The same caveat as mentioned in footnote 8 applies.
- ¹¹ See Appendix C for raw frequency data for the proposed alternative, as well as specifications of the particular usage data on which the counts are based.
- ¹² All cases targeted spoken and written language, with the exception of the frequency statement about *hun*-subject. Subsequently, for *hun*-subject we used only the spoken data. See Appendix D for raw frequency data for the proposed alternative, as well as specifications of the particular usage data on which the counts are based.

References

- Algemeen Nederlandsch Verbond. 1925. *Lijst van Nederlandsche Woorden Ter Vervanging van Op Bureaux Gebruikelijke Vreemde Termen, Voorgesteld En Aanbevolen Door Het Algemeen Nederlandsch Verbond*, 6th ed. Dordrecht: Algemeen Nederlandsch Verbond.
- Arnold, Vicky, Philip A. Collier, Stewart A. Leech, and Steve G. Sutton. 2000. The Effect of Experience and Complexity on Order Andrency Bias in Decision Making by Professional Accountants. *Accounting and Finance* 40: 109–34. [CrossRef]
- Ayres-Bennet, Wendy. 2020. From Haugen's codification to Thomas's purism: Assessing the role of description and prescription, prescriptivism and purism in linguistic standardisation. *Language Policy* 19: 183–213. [CrossRef]
- Brezina, Vaclav, Pierre Weill-Tessier, and Tony McEnery. 2020. #LancsBox, Version 5.1.2; Available online: <http://corpora.lancs.ac.uk/lancsbox> (accessed on 3 May 2020).
- Brysbaert, Marc, Paweł Mandera, and Emmanuel Keuleers. 2017. The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science* 27: 45–50. [CrossRef]
- Cameron, Deborah. 1995. *Verbal Hygiene. The Politics of Language*. London and New York: Routledge.
- Chapman, Don. 2021. 'Not Just a Few Dozen Trouble Spots'. Tallying the Rules in English Usage Guides. Presented at 6th Prescriptivism Conference, Vigo, Spain, September 25.
- Charivarius, G. Nolst Trenité. 1940. *Is Dat Goed Nederlands?* Amsterdam: De Spieghel.
- Damsteegt, B. Cees. 1964. *In de Doolhof van Het Nederlands/Aanwijzingen Voor Een Zuiver Taalgebruik*. 2 vols. Zwolle: W.E.J. Tjeenk Willink, vol. 1.
- DBNL—Digitale Bibliotheek voor de Nederlandse Letteren. 2021. Available online: <https://www.dbnl.org/> (accessed on 23 October 2021).
- de Raat, Friederike. 2012. *Hoe Bereidt Je Een Paard? & Andere Onuitroeibare Taalfouten*. Amsterdam: Nieuw Amsterdam.
- Delpher. 2021. Available online: <https://www.delpher.nl/> (accessed on 25 October 2021).
- Grauls, Jan. 1957. *Hoe Het Werd En Hoe Het Moet Zijn: Opstellen over Oorsprong En Betekenis van Woorden En Gezegden*. Leuven: Davidsfonds.
- Haje, Ch. F. 1932. *Taalschut. Schrijf Weer Nederlands*. Leiden: N.V. Leidsche Uitgeversmaatschappij.
- vor der Hake, Jan Arend. 1911. Is de Beleefdheidsvorm U 'n Verbastering van UEd. *De Nieuwe Taalgids* 5: 16–24.
- Heidbuchel, Hendrik. 1962. *Hoe Zeg En Schrijf Ik Het?* Hasselt: HeideLand.
- Hogeweg, Lotte, Stefanie Ramachers, and Helen de Hoop. 2018. Singular Agreement in Special Partitive Constructions in Dutch. *Journal of Germanic Linguistics* 30: 335–70. [CrossRef]
- Houët, Henriëtte. 2000. *Prisma Handboek van de Nederlandse Taal*. Utrecht: Het Spectrum.
- Kostadinova, Viktorija. 2018. Language Prescriptivism: Attitudes to Usage vs. Actual Language Use in American English. Unpublished Ph.D. thesis, Leiden University, Leiden, The Netherlands. Available online: <https://openaccess.leidenuniv.nl/handle/1887/68226> (accessed on 2 October 2021).
- Liberman, Mark. 2010. So New? *LanguageLog*. August 22. Available online: <https://languagelog ldc.upenn.edu/nll/?p=2570> (accessed on 5 May 2020).

- Meijers, J. A. 1959. *Allerlei Taalkwesties. Op de Slingerpaden van onze Taaltuin*. Amsterdam: Moussault's Uitgeverij.
- Moortgat, Alfons. 1925. *Germanismen in Het Nederlandsch*. Gent: Vanderpoorten.
- Oostdijk, Nelleke. 2000. The Spoken Dutch Corpus. Overview and First Evaluation. Presented at LREC-2000 (Second International Conference on Language Resources and Evaluation) Proceedings, Athens, Greece, May 31–June 2; Edited by Maria Gravididou, George Carayannis, Stella Markantantou, Stelios Piperidis and Gregory Stainhauer. Vol. 2, pp. 887–94.
- OpenSonar. 2020. Available online: opensonar.ivdnt.org (accessed on 6 May 2020).
- Piersoul, Jozefien, Robbert De Troij, and Freek Van De Velde. n.d. De Compilatie van Het Dutch C-CLAMP Corpus (Dutch Corpus of Contemporary & Late Modern Periodicals). Unpublished Work.
- Rutten, Gijsbert. 2020. Een Onderwerp Wat Onderzoek Behoeft. N.a.v. De Schutter & Kloots 2000. *Nederlandse Taalkunde* 25: 363–69.
- Schoonenboom, Judith. 2000. Analyse, Norm En Gebruik Als Factoren van Taalverandering: Een Studie Naar Veranderingen in Het Nederlands Onzijdig Relativum. Ph.D. thesis, Universiteit van Amsterdam, Amsterdam, The Netherlands.
- Sundby, Bertil, Anne Kari Bjørge, and Kari E. Haugland. 1991. *A Dictionary of English Normative Grammar 1700–1800 (DENG)*. Studies in the History of the Language Sciences 63. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Theissen, Siegfried. 1978. *Germanismen in Het Nederlands*. Hasselt: HeideLand-Orbis N.V.
- Tieken-Boon van Ostade, Ingrid. 2020. *Describing Prescriptivism*. London and New York: Routledge.
- Tiggeler, Eric. 2001. *Vraagbaak Nederlands*. 's-Gravenhage: Sdu Uitgevers.
- Trask, Robert Lawrence. 1999. *Key Concepts in Language and Linguistics*. London: Routledge.
- van der Horst, Joop, and Kees van der Horst. 1999. *Geschiedenis van Het Nederlands in de Twintigste Eeuw*. Den Haag and Antwerpen: Sdu/Standgaard.
- van der Meulen, Marten. 2020. Language Should Be Pure and Grammatical: Values in Prescriptivism in the Netherlands 1917–2016. In *Language Prescription Values, Ideologies and Identity*. Multilingual Matters 170. Edited by Don Chapman and Jacob D. Rawlins. Bristol: Clevedon, pp. 121–44.
- van der Meulen, Marten. 2021. De Canon van Taaladvies. In *Wat Gebeurt Er in Het Nederlands?! Edited by Nicoline van der Sijs, Lauren Fonteyn and Marten van der Meulen*. Utrecht: Sterck & de Vreese, pp. 169–73.
- Van der Meulen, Marten. n.d. The Feedback Loop of Prescriptivism. Relations between Prescript, Attitude and Usage in Dutch Morphosyntax. Unpublished Work.
- van der Meulen, Marten, and Gijsbert Rutten. 2022. Prescriptivism on Its Own Terms. Perceptions and Realities of Usage in Siegenbeek's Lijst (1847). *Language & History*, 1–23. [CrossRef]
- van der Sijs, Nicoline. 2001. *Chronologisch Woordenboek. De Ouderdom En Herkomst van Onze Woorden En Betekenissen*. Amsterdam and Antwerpen: L.J. Veen.
- Nicoline van der Sijs, comp. 2010. Etymologiebank.nl. Available online: <http://etymologiebank.ivdnt.org> (accessed on 25 October 2021).
- van der Sijs, Nicoline. 2019. Historische Taalkunde En Digital Humanities: Samen Naar Een Mooie Toekomst. *Tijdschrift Voor Nederlandse Taal-En Letterkunde* 135: 384–405.
- van der Sijs, Nicoline. 2021. *Taalwetten Maken En Vinden. Het Ontstaan van Het Standaardnederlands*. Utrecht: Sterck & De Vreese.
- van Gessel, Han, Bas van Kleef, and Jan Tromp. 1992. *Het Nieuwe Stijlboek*. 's-Gravenhage: Sdu.
- Weverink, Dolf. 2012. *Het Nieuwe Schrijven*. Amstelveen: Leporello Uitgevers.
- Wilke, Andreas, and Rui Mata. 2012. Cognitive Bias. In *The Encyclopedia of Human Behavior*. Edited by Vilayanur Ramachandran. Boston: Academic Press, vol. 1, pp. 531–35.
- Willems, Sanne, Casper Albers, and Ionica Smeets. 2020. Variability in the Interpretation of Probability Phrases Used in Dutch News Articles—A Risk for Miscommunication. *Journal of Science Communication* 19: 1–26. [CrossRef]
- Zimmer, Benjamin. 2019. 'Frequency Illusion' in the OED. *LanguageLog*. March 19. Available online: <https://languagelog.ldc.upenn.edu/nll/?p=42180> (accessed on 1 May 2020).
- Zwicky, Arnold. 2005. Just between Dr. Language and I. *LanguageLog*. August 7. Available online: <http://itre.cis.upenn.edu/~myl/languagelog/archives/002386.html> (accessed on 1 May 2020).
- Zwicky, Arnold. 2006. Why Are We so Illuded? *Stanford University*. Available online: <https://web.stanford.edu/~zwicky/LSA07illude.abst.pdf> (accessed on 1 May 2020).
- Zwicky, Arnold. 2008. Recency. *LanguageLog*. August 9. Available online: <https://languagelog.ldc.upenn.edu/nll/?p=463> (accessed on 2 May 2020).