

[to appear in the *International Journal of Corpus Linguistics* 15 (2010)]

Believe-type raising-to-object and raising-to-subject verbs in English and Dutch: A contrastive investigation in diachronic construction grammar.

Dirk Noël, *The University of Hong Kong* & Timothy Coleman, *Ghent University*

Dirk Noël
School of English
The University of Hong Kong
Pokfulam Road
Hong Kong S.A.R.
P.R. China
dnoel@hkucc.hku.hk
tel. +852 2859 1922
fax +852 2559 7139

Timothy Coleman
Vakgroep Nederlandse Taalkunde
Universiteit Gent
Blandijnberg 2
B-9000 Gent
Belgium
timothy.colleman@UGent.be
tel. +32 9 264 40 74
fax +32 9 264 41 70

Believe-type raising-to-object and raising-to-subject verbs in English and Dutch: A contrastive investigation in diachronic construction grammar.

The so-called “raising-to-subject” pattern that verbs of the type *believe* can occur in is usually treated as the passive alternative for the so-called “raising-to-object” pattern. In addition to broadening the empirical basis for the opposite claim that the English and Dutch raising-to-subject (or “nominative and infinitive”) patterns have a special functionality which is different from that of the passive construction, this paper specifically examines the stronger proposition that this has always been the case. It empirically investigates whether this proposition holds equally well for English and Dutch through a comparison of the frequencies of *believe-type* raising-to-object and raising-to-subject patterns in two diachronic corpora. The methodology makes use of Distinctive Collexeme Analysis.

Keywords: raising, nominative and infinitive, evidentiality, diachronic construction grammar, distinctive collexeme analysis, English/Dutch

1. Introduction

This paper¹ reports on a contrastive, diachronic, constructionist investigation of two morphosyntactic configurations that a few centuries ago were shared by the grammars of English and Dutch but which only remained productive in the first of these two languages. The first is a pattern that consists of an active perception, cognition or utterance verb (which Givón 1990 has grouped as “P-C-U verbs”) complemented by a *to*-infinitive that has its own explicit subject, as illustrated by the bits in italics in (1) and (2). Both in Chomskyan and post-Chomskyan linguistics this pattern is usually referred to as “raising to object”, also outside formalist paradigms (see, e.g., Givón 1993). An older, non-transformational, term for it is “accusative and infinitive” (or “ACI”, short for “accusativus cum infinitivo”). We have used this term in previous work (Noël 2003, 2008; Noël & Coleman 2009a) and will continue to do so here for reasons that will become clear below, having to do with the origin of the pattern.²

- (1) The former Lord Chancellor, Lord Hailsham, said he *believed the scientific community to be* a responsible one. (BNC A96 259)
- (2) Elk *meent zijn uil een valk te zijn*. (Dutch proverb)
‘Everyone considers his owl to be a falcon.’

The second pattern consists of a passive P-C-U verb complemented by a *to/te*-infinitive, as in (3) and (4), and is often referred to as “raising to subject”. The alternative term is “nominative and infinitive” (or “NCI”, short for “nominativus cum infinitivo”).

- (3) BOMB DISPOSAL specialists were yesterday called out to beaches on the south-west coast of Scotland after *Second World War phosphorus canisters believed to be* from an undersea dump were washed ashore. (BNC K5D 9706)
- (4) *De boten worden geacht* over een dag of vier Kaap Hoorn *te rond*. (ConDiv *De Telegraaf* NIE_S2)
‘The boats are predicted to round Cape Horn in about four days.’

The functionality of the ACI has received a considerable amount of attention in the linguistics of English, where it is usually contrasted with that of a matrix + *that*-clause construction, both in a synchronic (see Noël 1997, 2003; and the references there) and a diachronic perspective (see Fischer 1989, 1992; Los 2005; and the references there). Especially in historical English linguistics, the terms “accusative and infinitive” and “ACI” are not unusual. “Nominative and infinitive” and “NCI”, on the other hand, are not commonly used in Anglophone linguistics, since the NCI pattern is usually perceived as merely the passive counterpart of the ACI and consequently is not given independent attention. Largely the same is true of Dutch linguistics, where the ACI has mainly been treated in historical accounts (e.g., Overdiep 1935; Duinhoven 1991; Fischer 1994), Zajicek (1970) being the only dedicated study of the ACI in Present-day Dutch. We have proposed in Noël (2008) and Noël and Coleman (2009a), however, that both in English and in Dutch NCI patterns usually have a symbolic value which is quite different from that of the passive and that therefore the NCI cannot in most cases be reduced to a combination of an ACI construction and the general passive construction. In construction grammar terms, the NCI is not merely the passive version of the ACI, but qualifies as a construction — or rather, a cluster of constructions — in its own right. We have also argued, specifically about the English NCI, that it has *always* been more than a mere passive, i.e. that the English NCI pattern was not first a passive before certain specific NCI constructions developed, particularly what we have called the “evidential NCI construction” (Noël 2008).

Our objective in the present paper is to strengthen the empirical basis for both of these claims, and to establish whether the second one applies to Dutch as well, through a comparison of English and Dutch historical frequency data on the ACI and the NCI. We will start the paper with a summary presentation of the semantic potential of the English and Dutch NCI and of the evidence adduced so far in support of the claim that at least one specific NCI construction, the evidential NCI, has always been there in English (section 2). We will then formulate and motivate the research questions addressed in the present paper (section 3) and describe in detail where we went looking for answers and how we went about it (section 4), followed by a presentation and discussion of our findings (section 5).

2. Background

2.1 The plain passive NCI and three NCI constructions

In English, both the ACI and the NCI patterns are productive morphosyntactic configurations that allow a great variety of P-C-U verbs. However, not only is the variety greater in the case of the NCI, the NCI is also generally more frequent than the ACI. In the 100-million-word British National Corpus (BNC), the overall ratio is almost three to one, but for some representatives of the P-C-U class figures are much more dramatic, to the point that some verbs do not display the active pattern at all (e.g. *say*, *rumour*, *repute*; see Noël 2001 for details). This is the observation that led us to suspect that the NCI is not just a passive, because passives are not normally more frequent than actives (see, e.g., Biber *et al.* 1999). We have proposed in Noël (2008) that most NCI patterns have at least three uses and that a distinction should be made between a plain passive NCI (as in (5)), an evidential NCI (6) and a descriptive NCI (7). Only the first of these constitutes a mere combination of an ACI construction and the general passive construction. The other two are constructions in their own right, with specific semantic properties. The discussion of the examples will help to make clear this distinction.

- (5) In this book authorities *are said to be* limited also by the kinds of reasons on which they may or may not rely in making decisions and issuing directives, and by the kind of reasons their decisions can pre-empt. (BNC ANH 148)
- (6) AMERICAN ring doughnuts from The Delicious Donut Co are made from a flour which *is said to give* them a light, fluffy, and non-greasy consistency. (BNC A0C 1141)
- (7) A market can *be said to be* a place where buyers and sellers meet to make an exchange of goods (or services). (BNC K8W 508)

We call the NCI in (5) a “plain passive” (even though no active “equivalent”, i.e. an ACI, with *say* is possible in Present-day English) because the meaning of the matrix verb (*said*) is very much part of the propositional meaning of the sentence, which refers to a spatiotemporally locatable utterance act: a statement made “in this book”. In this example the meaning of *are said to* is very much “on-stage” (cf. Langacker 1987): if you remove it, the sentence becomes nonsensical, or at the very least conveys an altogether different meaning (?*In this book authorities are limited also by...*). The NCI is used here for the same information packaging reasons that motivate the use of the passive generally (compare: *The authors of this book say that authorities...*). In (6) and (7), on the other hand, the meaning of the *be said to* pattern is “qualificational” (cf. Aijmer 1972:39; Nuyts 2001:113). These sentences do not report on a specific utterance act. The meaning of *be said to* is “off-stage” here and you can safely remove the pattern, so that *said* can be argued to no longer be a matrix verb, but to be part of an auxiliary-like construction that modifies the meaning of the infinitive. (6) and (7) have in common that the NCI is not used for information structural reasons (since it can be left out: *...a flour which gives them..., A market is a place...*),³ but they differ in that they illustrate two different form-meaning pairings. In (6) the modifying construction has an evidential function, i.e. its writer uses the pattern to indicate that s/he has a *source* for the information s/he is conveying, so that s/he is not the (sole) judge of the factuality of the statement that American ring doughnuts are light, fluffy, and non-greasy (see Noël 2008 for further elaboration). In (7) the modifying construction connects a description with a descriptum (see Goossens 1991).

Other frequent instantiations of the “evidential NCI construction” in Present-day English are *be alleged to, be assumed to, be believed to, be claimed to, be considered to, be deemed to, be estimated to, be expected to, be felt to, be found to, be held to, be known to, be reported to, be seen to, be shown to, be supposed to, be taken to, be thought to* and *be understood to* (Noël 2008). Whether these can all realize the “descriptive NCI construction” as well remains to be seen. Two of these patterns can carry the additional meaning illustrated in (8) and (9), however.

- (8) Pupils *are expected to use* their Maths to solve problems. (BNC K9X 434)
- (9) You *were supposed to do* six and you only did four! (BNC KST 788)

In these examples *be expected to* and *be supposed to* instantiate a “deontic NCI construction”.⁴ The following examples illustrate that the same patterns can also be the realization of a plain passive (10-11) and the evidential NCI construction (12-13), and they can serve as further illustrations of the distinction between these two NCI uses. Note that in (12-13), *be expected to* and *be supposed to* have the same, somewhat paradoxical, function as *be said to* in (6): the writer has a source, which diminishes

his/her responsibility for the information conveyed and at the same time adds to its reliability.

- (10) Standards for exposure to benzene *are expected* by the UK government *to be set* in 1993: a level of 3 ppb is under consideration, although according to the World Health Organization there is “no known safe threshold dose”. (BNC JC3 608)
- (11) The mechanism *supposed* by Miller (1948) *to underlie* acquired equivalence is that introduced by Hull (1939) with his notion of secondary generalization. (BNC APH 1337)
- (12) The Japanese economy *is expected to grow* by only 3.8% during fiscal 1991 compared with this year’s 5.2%. (BNC ABD 953)
- (13) Napoleon *is supposed to have said* “An army marches on its stomach.” (BNC A77 422)

In Dutch the ACI and NCI are much less “part of the grammar” than in English. To illustrate the ACI construction in Present-day Dutch we resorted to presenting an archaism in the introduction (2), obviously a relic from a time when the pattern was more common in Dutch than it is now, because modern examples of the pattern are very hard to come by. In historical accounts the ACI is said to have had two usage peaks (see, e.g., Duinhoven 1991). It occurred more than sporadically at the start of the Middle Dutch period (generally taken to span the time from 1200 to 1500), but rapidly grew out of fashion again during that period. Later it enjoyed a brief moment of popularity in the early stages of Modern Dutch (late 16th and 17th century), mainly in the formal writings of authors who had had a classical training (Van Leuvensteijn 1997). Subsequently it disappeared from the language, barring a few relics (e.g., Duinhoven 1991; Fischer 1994). Zajicek (1970:203) claims that the ACI still occurs with a limited set of P-C-U verbs in very formal administrative and didactic texts, but Duinhoven (1991:425) questions the grammaticality of most of the examples he offers, though he recognizes that the ACI is not altogether impossible in relative clauses. The Dutch NCI, on the other hand, has survived better than the ACI, but it is a far less prolific pattern there than in Present-day English. Corpus research of our own revealed that the only patterns that still occur today with any frequency are *geacht worden te* (‘be considered/supposed to’), *verondersteld worden te* (‘be supposed to’) and, to a lesser extent, *verwacht worden te* (‘be expected to’). While the 20 most frequent NCI patterns in Present-day English occur with a frequency of 340.47 tokens per million words in the entire BNC, and with a frequency of 433.45 and 710.53 tokens per million words respectively in a newspaper and a natural sciences sub-corpus of it, these three Dutch patterns together only occur with a frequency of 4.4 tokens per million words in a fragment of about 12.5 million words of the newspaper component of the Dutch ConDiv corpus (Noël & Coleman 2009a).

Like the English NCI, the Dutch NCI is not simply a passive, however, if at all. Most often it instantiates a deontic NCI construction (14), less often an evidential NCI construction (15) and occasionally also a descriptive NCI construction (16). We have exemplified them here for *geacht worden te* (for examples of the other patterns, see Noël & Coleman 2009a).

- (14) Chefs van afdelingen *worden geacht* excessief of nutteloos surfen *te voorkomen*. (ConDiv NRC_NIEUWS07)

- ‘Heads of departments are supposed to prevent excessive and useless surfing.’
- (15) Ze werden populair bij atleten en wielrenners omdat meer rode bloedlichaampjes *worden geacht* een “zuurstofvoorsprong” *te geven*. (ConDiv NRC_VARIA01)
 ‘They [EPO hormones] became popular with athletes and cyclists because more red blood cells are thought to give an “oxygen advantage”.’
- (16) Kan de rechterlijke macht, nu in ons constitutioneel staatsbestel alle machten van de natie uitgaan, *geacht worden* de natie *te vertegenwoordigen*? (ConDiv DS961216)
 ‘Since in our constitutional system all powers are vested in the nation, can the judicial power be supposed to represent the nation?’

Zooming in on the evidential NCI construction, we will now summarize our evidence so far for the claim that the arrival of this construction in English (and Dutch) cannot be separated in time from the NCI’s introduction into these languages as a morphosyntactic possibility.

2.2 A concise and selective history of the evidential NCI

Though there is no complete consensus on the origin of the ACI and the NCI in the two languages considered here, the two late-20th-century authorities on these structures in English historical linguistics, Warner (1982) and Fischer (1989, 1992, 1994), agree that these patterns are calques from Latin which became thoroughly entrenched in the grammar of English in the 15th century. In Dutch historical linguistics the debate on whether they are native or of Latin origin seems not to have been settled yet (for a summary and references, see Noël & Coleman 2009a:166) but there is agreement at least that when these patterns acquired a certain popularity in Early Modern Dutch (between 1500 and 1650) this was due to the influence of Latin. With regard to the NCI this raises the question of whether what was copied was simply a passive of the ACI — which went on to develop a number of “qualificational” functions through grammaticalization processes internal to English and Dutch — or whether there already was an evidential NCI in Latin. Comments in text-analytical work on the use of *dicitur* (‘be said’) and *creditur* (‘be thought/believed/supposed’) by Roman authors seem to suggest that the latter was the case (Noël 2008:323-324).⁵ One cannot therefore avoid the conclusion that, if the NCI was a borrowing from Latin, so must have been the evidential NCI construction. In other words, it is unlikely that the English and Dutch evidential NCI are the result of a grammaticalization of the passive NCI owing to the routinization resulting from frequent use that is typical of central cases of grammaticalization. The fact that many early examples of the NCI in English are clearly of an evidential nature supports this (Noël 2008:324-325). It is a claim for which further support is needed, however, and it especially needs to be confirmed for Dutch.

3. Problem: have the English and Dutch NCI always been more than passives?

The research reported on in the present paper was designed to provide a type of evidence for the past constructional status of the NCI in both English and Dutch which does not rely on the analyst’s interpretation of individual examples but which provides claims about this with a more objective, empirical basis. If the evidential NCI was borrowed from Latin — and hence was there from the start of the introduction of the NCI pattern in English/Dutch — and if the presence of this construction helps to explain

the higher frequency of the NCI pattern relative to the ACI in Present-day English and Dutch, one might expect that the NCI has always been more frequent than the ACI in these languages. Research on the occurrence of these patterns in Middle English does not support this expectation (Warner 1982, Fischer 1992), nor does research on Middle Dutch, Fischer (1994:111) reporting having found “no examples” of the NCI in the Middle Dutch sources she consulted.⁶ However, while there is no relevant research on Early Modern English, Fischer (1994:113) did find a “large number” of NCIs in a corpus of 17th-century Dutch she put together. The difference between these two observations on Dutch could be highly significant, also for English, because the influence of Latin on Dutch as well as English of course came in two completely different waves. There first was the influence of biblical and ecclesiastical Latin during the Middle Ages and later the influence of classical texts during the Renaissance. One could speculate that there was less need for an evidential construction in the religious texts that were influential for Middle English and Middle Dutch than in the scholarly texts that were important during the 17th century. Therefore, starting where earlier research suggests it might be meaningful to do so, we will trace back the history of the English and Dutch ACI and NCI to the 17th century, asking whether the NCI has consistently been more frequent than the ACI from then till now in both languages. If this turns out to be the case, one may assume the NCI to have had a different symbolic value from the passive from the moment the frequency of this morphosyntactic pattern was boosted under the influence of classical texts.

Next, inspired by the work by Gries and Stefanowitsch (Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004), we will try to find supplementary evidence for the special symbolic value of the NCI by asking whether the ACI and NCI have consistently displayed a different preference for certain verbs in the two languages. If indeed Latin *dicitur* and *creditur* served as a model for the evidential NCI, one might expect for instance that English verbs like *say* and *believe* and Dutch verbs like *zeggen* and *geloven*, and/or semantically closely related verbs, will have consistently preferred the NCI over the ACI. In more general terms, if from the moment the ACI and NCI patterns were introduced in English and Dutch the NCI had a number of specific semantic properties not shared by the active construction, most notably the evidential function, this should be evident from the kinds of verbs frequently attested in both patterns.

4. Methodology

The decision to start this contrastive diachronic investigation in the 17th century was also partly determined by the availability of two comparable corpora that were large enough to produce reliable frequency data on the phenomenon under investigation. The first one is the Corpus of Late Modern English Texts (CLMET) compiled at the University of Leuven by Hendrik De Smet, with texts drawn from the *Project Gutenberg* and the *Oxford Text Archive*. At the time the CLMET was downloaded for use in the present research (i.e. in the spring of 2006), it spanned a period from 1640 to 1920, divided into four sub-periods of 70 years each, ranging from 1.9 to 6.1 million words of running text (for an account of the principles behind the compilation of the corpus, see De Smet 2005). The comparable Dutch corpus is one that was compiled along the same principles as the CLMET for the purpose of the study reported on in Noël and Coleman (2009a). It consists of extracts from texts available online from the *Digitale Bibliotheek voor de Nederlandse Letteren* and the *Project Gutenberg* and spans

the same period as the CLMET, the size of the four 70 year-subcorpora ranging from 1.2 to 3.5 million words of running text.

English NCI patterns were identified in the CLMET by looking for past participle forms of P-C-U verbs immediately followed by *to*. The ACI patterns were identified by searching for any form of P-C-U verbs separated from *to* by 0 to 6 words. The verbs searched for constitute the union of three collections of verbs: a) the verbs that were investigated in Noël (2001), i.e. the ones referred to there as “the sixty-odd *believe*-type verbs that Postal (1974:297-317) lists in his chapter on ‘The scope of raising in clause domains’”, b) the verbs that were included in Noël (2008) and Noël and Coleman (2009a), i.e. the ones that were identified as NCI verbs in the BNC by manually sifting the results of a query for any form of the verb *be*, followed by “any word”, followed by the infinitival particle *to*, and c) the verbs that were identified in the CLMET as NCI verbs by manually sifting the results of a query for any form of the verb *be*, followed by “any word”, followed by *to be*. This produced a list of 138 English verbs, 74 of which turned out to occur at least once as an ACI and/or an NCI verb in the CLMET (i.e. the ones included in Table 1).

The Dutch ACI and NCI patterns were identified in the Dutch corpus by looking for any form of P-C-U verbs separated from *te* by 0 to 10 intervening words (the larger span as compared to the English queries being motivated by the relatively freer word order in Dutch). The verbs searched for were those that have either been said in the literature to still occur occasionally in ACI and/or NCI patterns today (Duinhoven 1991; Zajicek 1970, Noël & Coleman 2009a) or to have occurred in these patterns during the 17th century (Overdiep 1935; Duinhoven 1991; Fischer 1994). These selection criteria resulted in a list of 62 Dutch verbs, 42 of which were attested at least once in ACI and/or NCI patterns in the diachronic corpus (i.e. the ones listed in Table 2).

For each language, the absolute frequencies of the patterns subsequently formed the input of four separate *distinctive collexeme analyses*, i.e. one for each of the subcorpora. For a full explanation and justification of this methodology we would like to refer to Gries and Stefanowitsch (2004), but very briefly what a distinctive collexeme analysis does is test the degree of association between two or more “competing” constructions C_1 , C_2 , etc. and the various lexemes occurring in a particular slot of these constructions, on the basis of the co-occurrence frequencies of lexeme x and C_1 , x and C_2 , etc. and the overall frequencies of C_1 , C_2 , etc. in the corpus (or the sub-corpus in our case). A lexeme is revealed by this test to be significantly attracted to one of the constructions under investigation if its observed frequency in that construction significantly exceeds the frequency expected on the basis of the overall distributions. If this procedure is repeated for all lexemes occurring in the investigated slot of either of the constructions in the corpus, the outcome is a list of so-called *distinctive collexemes* for each of the examined constructions, i.e. the lexemes which significantly prefer that construction over the other construction(s) (in this case, the verbs with a significantly above-average preference for the ACI over the NCI or vice versa). The analyses were carried out using version 3 of Stefan Gries’ R-script for collostructional analysis (Gries 2004).

5. Results and discussion

5.1 General comments

The results of the frequency counts in each of the four parts of the two corpora are provided in Table 1 and Table 2, respectively. To allow comparison, the bottom row in both tables presents the totals as frequencies per million words.

[Insert Table 1 about here]

[Insert Table 2 about here]

A brief glance at these normalized total frequencies suffices to show that both the ACI and the NCI consistently occurred far more frequently in English than in Dutch. Even at their respective frequency peaks — 100.93 ACI instances per million words in the first period and 21.44 NCI instances per million words in the second period — the Dutch frequencies are nowhere near the English ones, which range between 181.51 (Period 4) and 335.7 (Period 1) per million words for the ACI and between 185.43 (Period 4) and 278.62 (Period 1) for the NCI. In the last period, the Dutch frequencies have dropped to 5.12 instances per million words for the ACI and 7.96 for the NCI. Moreover, the majority of these instances is made up of just four verbs, viz. *achten* ‘consider’, *menen* ‘be of the opinion’, *veronderstellen* ‘suppose’ and *zeggen* ‘say’, corroborating earlier observations that the ACI and NCI are infrequent and lexically restricted patterns in the later stages of Modern Dutch (Fischer 1994, Noël & Coleman 2009a). In English, by contrast, both patterns are very much part of the grammar throughout the investigated period and many individual verbs attain respectable ACI and/or NCI frequencies, allowing for statistical comparison. In the next two subsections, we will first zoom in on these English data and then compare the observed trends to the situation in Dutch, to the extent that the data enable us to do so.

5.2 English

A comparison of the normalized frequencies in the bottom row of Table 1 reveals that the overall higher frequency of the English NCI relative to the ACI only came about in the 18th century.⁷ Counter to the expectations formulated in Section 3, the ACI was more frequent than the NCI in the 17th century. However, unlike in Dutch (see below), the NCI was already well established at the start of the investigated timespan, the proportion of ACI to NCI instances in the first period (1640-1710) being about six to five. In addition, if we start comparing the frequencies of the NCI and ACI patterning of individual verbs, we notice that there are a number of verbs that already preferred the NCI over the ACI in the 17th century. This is confirmed by the results of the distinctive collexeme tests for each of the four sub-corpora, summarized in Table 3. This table lists the distinctive ACI and NCI collexemes in each period, in order of diminishing “collostructional strength” (i.e. degree of association).⁸ For instance, in the first period, *find* is the verb with the strongest preference for the ACI over the NCI, and *say* displays the strongest preference for the NCI over the ACI.

[Insert Table 3 about here]

Table 3 shows that six verbs display a statistically significant above-average preference for the NCI in the 17th century, viz. *say*, *suppose*, *understand*, *think*, *note* and *presume*. Three of these continue to display a significant preference for the NCI in subsequent centuries, viz. *say*, *suppose* and *think*, while one of them, *presume*, continues to be more frequent as an NCI verb but not in a statistically significant way. Other verbs that consistently prefer the NCI over the ACI but not in a statistically significant above-average way in all four sub-corpora are *see*, *report* and *repute*. Conversely, there are also verbs that display a consistent preference for the ACI, viz. *affirm*, *assert*, *conclude*, *consider*, *declare*, *deny*, *discover*, *fancy*, *guess*, *imagine*, *judge*, *perceive*, *profess*, *prove* and *show*, and four verbs do so statistically significantly in all four periods, viz. *believe*, *take*, *conceive* and *know*. For several of the other verbs just mentioned the ACI preference is statistically significant in three of the four periods; instances include *conclude*, *declare*, *perceive* and *prove*, all of which display a significantly above average preference for the ACI in the first three subperiods (and a non-significant ACI preference in Period IV). The results therefore clearly point toward the existence of two subclasses of P-C-U verbs, distinguished by their constructional preferences, “ACI verbs” and “NCI verbs”.⁹ Moreover, such a distinction was already present at the start of the timespan covered in this investigation. There is only one verb that changes sides, as it were, in a statistically significant way, viz. *find*, which demonstrates a significant above-average preference for the ACI in the first two time periods, but then a significant above-average preference for the NCI in the last two periods.

We think all of this is congruent with the hypothesis that the special symbolic value of the NCI is not a recent development and that, moreover, the history of the English evidential NCI goes back further than the 18th century. Though the NCI only overtook the ACI in frequency during that century, this does not mean that the NCI only developed its special functionality around then. It is merely the confirmation of a stylistic change that has been independently established to have taken its course during the 18th century and as a result of which a construction we would argue was already available increased in usefulness. Adamson (1998:662), for instance, has written that

During the eighteenth century, an objectivising, generalising style had become the goal in most forms of public discourse, prompting writers to look for ways of eliminating, minimising or conventionalising the use of subjective features. In scientific writing, for instance, the impulse towards objective description led to the gradual emergence of the passive, which linguistically emancipates an experiment from its author’s personal experience (by converting, for instance, ‘I saw the liquid boil’ to ‘the liquid was seen to boil’).

In our analysis the example of the “passive” given here is an evidential NCI.

The claim that the evidential NCI was already around before then is supported by the semantic nature of the verbs that already displayed a significant preference for the NCI in the 17th century. It cannot be a coincidence that most of these verbs are semantically closely related to either of the two Latin verbs *dicere* (‘say’) or *credere* (‘believe, think’), which led to the Latin evidential constructions *dicitur* and *creditur*. *Be said to* is obviously equivalent to the former and *be thought/supposed/presumed to* are all equivalent to the latter. In other words, the semantics of these NCI verbs is compatible with the hypothesis that the evidential NCI is a calque from Latin.

The fact that there is a sizeable group of verbs that has consistently preferred the ACI above the NCI and a group that has consistently preferred the NCI above the ACI

can be taken to confirm that the two patterns have always had a different symbolic value (at least for the period covered by the CLMET). It is not easy to detect what the members of each group have in common that distinguishes them from the other group, but one could argue that most of the verbs that consistently prefer the ACI over the NCI have a more specific meaning than the average NCI verb. *Affirm, assert, declare, deny* and *profess*, for instance, cannot be characterized as “basic linguistic action verbs” (Verschueren 1985), unlike *say*, the most typical NCI verb. Likewise *conceive, conclude, consider, fancy, guess, imagine, judge* and *take* are less basic cognition verbs than *think* and *suppose*. The same cannot be said of the typical ACI verbs *believe* and *know*, but they perhaps differ from the typical NCI verbs *think* and *suppose* in strength or commitment. *Believe* can be a near synonym of *think*, but it can also have the stronger meaning glossed by Dixon (2005:140) as “think of something as true (when in fact it may not be, but the Cogitator will not accept that it may not be)”. Similarly, *know* implies more commitment to the truth of a proposition than both *think* and *suppose*. One could hypothesize, therefore, that the specificity and/or the commitment entailed by the verbs typically occurring in the ACI pattern requires these verbs to have an expressed, topical “agent”, which makes them less compatible with the NCI pattern, because this is usually agentless. It would also make these verbs less compatible with the typical “off-stage” meaning of the NCI.

If we compare the results of our distinctive collexeme analysis with Gries and Stefanowitsch’ (2004) distinctive collexeme analysis for the active vs. the passive construction, displayed in Table 4, a few additional observations can be made with relation to the symbolic value of both patterns.

[Insert Table 4 about here]

A first observation is that none of the verbs we considered are part of Gries and Stefanowitsch’ list of the 20 most distinctive collexemes of the passive. Our verbs are all P-C-U verbs, but there are no P-C-U verbs among the verbs most typically associated with the passive. This alone might already be an indication that NCIs are not first and foremost passives. Gries and Stefanowitsch (2004:110) conclude from their results that

the distinctive-collocate collexeme analysis shows that passive voice is a construction in its own right with its own specific semantics. It encodes a situation where the referent of the passive-voice subject has come to be in some relatively stable end state as a result of someone acting on it. The distinctive collexemes of the active-voice construction are either action verbs that do not lead to such end states, or they are states that are not brought about by someone acting on the entity in this state.

Utterance acts and acts of cognition and perception do not, of course, involve entities that are acted on, which predisposes P-C-U verbs to favouring the active voice. Yet still many of them have for a long time preferred the ostensibly “passive” NCI pattern over the active ACI.

A second and more crucial observation is that there are quite a few P-C-U verbs in the active column of Table 4 and four of them are verbs that consistently prefer the NCI above the ACI: *say, think, suppose* and *see*. This is an important fact: if these verbs both typically occur in the active voice construction *and* in the NCI pattern, the latter must be more than just a passive. Translated in constructionist terms, there exist NCI

constructions that do not “inherit” the semantics of the passive construction (on “inheritance” in construction grammars, see Lakoff 1987:508, Goldberg 1995:73-74, Kay & Fillmore 1999:7-8, 30-31).

A third and final observation is that two verbs that show an above-average preference for the active over the passive also display a consistent above-average preference for the ACI over the NCI, viz. *believe* and *know*. This would be consistent with a situation in a constructional network where the ACI is a daughter construction of the active construction. In other words, ACIs are actives, but NCIs are usually not passives, at least not semantically.

How to explain the fact that *find* changes from being a typical ACI verb to being a typical NCI verb? We would contend that this, too, can be related to the historical stylistic change affecting expository genres which we already referred to above. Additional support for it is supplied by Montgomery (1996), who has observed that 17th-century science writing offered “detailed descriptions of experiments, as a kind of ‘historical’ writing of what happened, told directly through the ‘I’ as a personal narrator of events, such that the reader would be brought as close to these events as possible” (Montgomery 1996:93), whereas “[b]y the early 19th century, [...], the actor-I had begun to disappear” (Montgomery 1996:106). In an academic culture where the most reliable underpinning of information offered is observational or experimental evidence which is somehow “found” *find* is of course the quintessential evidential verb. It is consequently little surprising that it was first frequently used in the active voice and that it was subsequently recruited for use in the evidential NCI when changing evidentiality requirements boosted the frequency of this construction. The fact that *find* was first predominantly associated with the ACI and later with the NCI therefore substantiates the symbolic value of the latter.¹⁰

5.3 Contrasting English and Dutch

If we compare the normalized frequencies in the bottom row of Table 2 with the matching information in Table 1, we notice that the frequencies are not only consistently much lower in Dutch than in English, both for the ACI and the NCI, but the proportional relationship between the two patterns in the two languages is different as well. Whereas in English the NCI was only mildly less frequent than the ACI in the 17th century and then became the more frequent pattern, in Dutch the NCI started out being considerably less frequent than the ACI, remained much less frequent in the 18th century in spite of a rise in its frequency and a drop in the frequency of the ACI, and only in the second half of the 19th century became the more frequent pattern, though a very rare one compared to the English NCI. As was stated in Section 2.1 above, earlier research into the presence of both patterns in *Present-day* Dutch has shown that the ACI is (virtually) extinct and that only three verbs are attested in the NCI pattern with any frequency, viz. *achten* (‘consider’), *veronderstellen* (‘suppose’) and *verwachten* (‘expect’) (Noël & Coleman 2009a). Overall, therefore, it seems clear that Dutch has moved from a situation where only the ACI was a true option, if not a very frequently taken one, to a situation where only the NCI is possible, though again not frequent (neither in type nor token frequency).

To corroborate this we computed a gamma coefficient of the relation between the numbers of observed ACI and NCI instances over time. This revealed that, in statistical terms, there is a significant linear increase in the proportion of the NCI in the combined total of ACI and NCI instances from the first period in our corpus to the last (effect of NCI versus ACI uses: $\gamma = 0.5970491$, ASE = 0.05621057).¹¹ In other words,

in Dutch, the proportion of the NCI displays a significant linear increase as time goes by. There is no such trend in English: the figures in Table 1 do show a slight increase of the NCI proportion from period I to period II, but this is followed by a status quo. A conclusion can be that the Dutch NCI developed its special functionality later than the English one.¹²

As for the ACI and NCI frequencies of the individual verbs in the Dutch table, many of them are too low to allow meaningful comparison. It is technically possible to conduct collexeme analyses of these data, since the method of distinctive collexeme analysis can handle small frequencies, but, unsurprisingly, the results of these tests are far from spectacular. Only in Period II (i.e. the period with the largest absolute ACI and NCI frequencies) does the test reveal a number of statistically significant contrasts: *geloven* ('believe'), *denken* ('think') and *menen* ('be of the opinion') are significant ACI collexemes while *zeggen* ('say'), *(ver)onderstellen* ('suppose') and *rekenen* ('estimate') are significant NCI collexemes. Overall, however, and leaving aside statistical significance, the numbers that seem large enough to be indicative of constructional preferences suggest a preference for the ACI before the 19th century (i.e. in the first and second periods of the corpus), most notably in the cases of *achten* ('consider'), *bevinden* ('find'), *oordelen* ('judge'), *denken* ('think'), *geloven* ('believe') and *menen* ('be of the opinion'). At least two of these verbs seem later to have developed a preference for the NCI, viz. *achten* and *bevinden*, but the numbers are small and only NCI *achten* has survived to this day (Noël & Coleman 2009a). *Zeggen* ('say') preferred the ACI in the first period only and "already" developed a preference for the NCI in the 18th century, which is interesting in the light of a dedicated study of NCI *zeggen* in a large 19th-century Dutch corpus (Coleman & Noël 2009), which showed that *gezegd worden te* ('be said to') at one time used to be quite common as an evidential in formal varieties of written Dutch. Note, however, that the preference for NCI *zeggen* in Dutch is less marked and comes later than the preference for NCI *say* in English.

In sum, though we cannot underpin this claim with statistical evidence of the kind offered in the previous subsection, the fact that *achten*, *bevinden* and *zeggen* seem to have changed from preferring the ACI to favouring the NCI does constitute evidence that the Dutch NCI did not simply inherit the general semantic properties of the passive construction, unless one would wish to argue that the meaning of the passive has changed. The special functionality of the Dutch NCI surfaced later than that of the English NCI, however, and was never exploited to the same extent.

6. Conclusion

This paper has demonstrated differences between English and Dutch in the historical evolution of the entrenchment of their ACI and NCI patterns. Our main aim, however, was to test the hypothesis that the NCI has always had a symbolic value different from (or on top of) that of the passive construction. The empirical, non-interpretative evidence supplied here was inspired by and based in part on Gries and Stefanowitsch' (2004) distinctive collexeme analysis methodology. Distinctive collexeme analysis can provide evidence for the symbolic value of constructions through a measure of the degree of association between (alternating) constructions and the lexemes that fill them. We have shown that, in English, there is a subclass of P-C-U verbs which have for a long time (starting at least in the 17th century, possibly earlier) displayed a statistically significant above-average preference for the NCI. Several of these verbs display the same kind of preference for the active voice and this makes it very unlikely that the NCI, though "passive" in form, has a mainly 'passive' symbolic value, or that this has ever

been the case even. Our findings therefore corroborate the hypothesis that right from the start of the investigated period the English ACI and NCI were not just perspectival variants distinguished primarily by their active vs. passive information-structural properties, but had quite different symbolic values. In Dutch, unlike in English, the NCI was decidedly less frequent than the ACI during the first half of the investigated timespan and became the more frequent pattern only in the second half of the 19th century. Our data reveal a significant linear change in the preference of the examined verbs from the ACI in the 17th century to the NCI in the 20th century. The extent of the change is such that the ACI has all but disappeared from Dutch. The verbs that remain in the NCI today, *achten* ('consider'), *veronderstellen* ('suppose') and *verwachten* ('expect'), are all cognition verbs. Unless one can argue that the Dutch passive attracts an altogether different category of verbs than the English passive, the extremely close association between these three verbs and the NCI in Present-day Dutch, combined with the fact that one of them used to be closely associated with the ACI, is an indication that in Dutch, too, the NCI has a symbolic value different from the passive. However, the Dutch NCI seems to have developed this special functionality more recently than the English pattern.

Notes

¹ We are grateful for the comments of two anonymous referees and for the financial support of the University of Hong Kong Seed Funding Programme for Basic Research (contract no. 200611159021), the Research Fund of the University of Ghent (for the project "Meaning in between structure and the lexicon", contract no. GOA B/05971/01) and the Belgian Federal Science Policy Office (for the project "Grammaticalization and (Inter)Subjectification", contract no. IUAP P6/44). We also owe gratitude to Chan Tsz Ying and Fung So Hing for their help with sifting the data that went into Table 1.

² The description and the examples of the ACI offered here should make clear that what we will be dealing with is the pattern which is sometimes called the "genuine", "learned" or "Latin-type" ACI (e.g., see Fischer 1989, 1992, 1994), i.e. the pattern containing what Postal (1974) termed "B-element R-triggers" (verbs of the type of *believe* that "trigger Raising"). The term "ACI" has also been used to refer to perception verbs and causative verbs (*make*, *let*) followed by "accusatives" and bare infinitives, and mandative verbs (e.g. *order*) followed by accusatives and *to*-infinitives, but these patterns do not constitute the object of investigation in this paper.

³ An anonymous referee has rightly pointed out that our paraphrase of (7) not only leaves out the NCI pattern but also the modal *can*, which might therefore be part of the descriptive NCI construction. However, we have noted in Noël and Colleman (2009a:153) that modals (not necessarily *can*) are only an optional (though frequently occurring) part of this construction. An example without a modal cited there is:

- (i) When two grammatical items occur together in a specified syntagmatic relation, they *are said to* colligate and the combination is a colligation (as opposed to a collocation). (BNC HOY 1159)

⁴ The reference to "additional" meanings should not be taken to mean that the deontic NCI construction has developed from another NCI construction, particularly the evidential one, though this has recently been suggested in the literature (for references and a paper-length discussion of this point, see Noël & van der Auwera 2009).

⁵ Moles (1991:553), for instance, offers the following comment about the use of *dicitur* in the lines *risissi Cupido / dicitur atque unum surripuisse pedem* ("Cupid is said to have laughed and to have stealthily removed one foot") from Ovid's *Amores*: "It is said', 'they say', 'there is a story' etc. are often used as 'distancing' formulae whereby the writer does not commit himself to the veracity of certain material, particularly when it is of a supernatural character." Space prevents a more elaborate discussion, but obviously, this shedding of responsibility for the truthfulness of a proposition through the invocation of an unspecified source is strikingly reminiscent of the effect of the NCI in some of the Present-day English and Dutch examples discussed above.

⁶ Fischer (1994) terms NCI patterns "second passives", following Warner (1982), who in turn borrowed the term from Lees (1960).

⁷ The general drop in the frequencies of the ACI and NCI is less relevant here and there is no space to discuss it. Suffice it to say regarding the frequency of the NCI that we have argued elsewhere that a comparison with its present-day frequency, established on the basis of the British National Corpus, suggests that the drop in its frequency in the genres that are well-represented in the CLMET (fiction and philosophy) must have been matched by a frequency rise in genres that are not represented there (journalism and science) (Noël 2008). We have no data on whether such genre distinctions are relevant to the ACI as well and on whether the frequency of the ACI continued to drop in the 20th century.

⁸ Only collexemes which are significant at the 95% level of statistical confidence ($p < 0.05$) are included in the table.

⁹ There are also a number of verbs which are not revealed by the tests to be significantly attracted to either the ACI or the NCI in any of the four investigated sub-corpora. Most of these are low-frequency verbs with too few attested ACI and NCI instances to allow for conclusions about their constructional behaviour (e.g. *avow*, *conjecture*, *surmise*). *Pretend* is the only verb with over 20 relevant instances which is attracted to neither of the two constructions in none of the four periods and which might hence be considered as truly “neutral” with regard to the ACI/NCI distinction (note, however, that the majority of *pretend*'s relevant instances date from the first period, after which *both* its ACI and NCI uses seem to have dwindled).

¹⁰ For a paper-length elaboration of this point, see Noël and Coleman (2009b).

¹¹ The gamma coefficient characterizes the strength of the association between two variables of which at least one is ordinal (in this case the period variable is inherently ordered, from period I to period IV). Values range from -1 (perfect negative linear association) to +1 (perfect positive linear association), with a value of zero indicating the absence of association. A .95 confidence interval (CI) is computed around the gamma coefficient as follows: $CI = \text{gamma} \pm 1.96 * ASE$ (=Asymptotic Standard Error). For the association of NCI to ACI frequencies in Dutch, the gamma coefficient is 0.5970491 and the ASE is 0.05621057. This means that the .95 confidence interval is [0,486 ; 0,707]. Since this interval excludes the zero value, we can be 95% certain that there is a positive linear association: the number of NCI as opposed to ACI instances significantly increases from period I to period IV.

¹² Our findings for Period I seem to be at odds with the claim in Fischer (1994:113) that she observed a “large number” of NCIs in 17th-century Dutch. Since Fischer does not report exact ACI and NCI frequencies, it is impossible to compare her findings on this score with ours. However, her statement about the large number of NCIs in Renaissance Dutch must probably be seen in relation to the observed absence of such patterns in Middle Dutch, rather than as a claim about the synchronic relation between the ACI and the NCI in the 17th century.

References

- Adamson, S. 1998. “Literary language”. In S. Romaine (Ed.), *The Cambridge History of the English Language, IV, 1776-1997*. Cambridge: Cambridge University Press, 589-692.
- Aijmer, K. 1972. *Some Aspects of Psychological Predicates in English*. Stockholm: Almqvist & Wiksell.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finnegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Coleman, Timothy & Dirk Noël. 2009. “*Gezegd worden + te*-infinitief: een verouderde evidentiële constructie”. [Dutch *gezegd worden* (‘be said to’) + *te*-infinitive: an archaic evidential construction.] *Tijdschrift voor Nederlandse Taal en Letterkunde*, 125.
- De Smet, H. 2005. “A corpus of Late Modern English texts”. *ICAME Journal*, 29, 69-82.
- Dixon, R. M. W. 2005. *A Semantic Approach to English Grammar*. Oxford: Oxford University Press.
- Duinhoven, A. M. 1991. “Dat siet men wit ende reine wesen. A.c.i.-constructies in het Nederlands”. *Nieuwe Taalgids*, 84, 409-30.
- Fischer, O. 1989. “The origin and spread of the accusative and infinitive construction in English. *Folia Linguistica Historica*, 8, 143-217.

- Fischer, O. 1992. "Syntactic change and borrowing: the case of the accusative and infinitive construction in English". In D. Stein & M. Gerritsen (Eds.), *Internal and External Factors in Syntactic Change*. Berlin: Mouton de Gruyter, 77-88.
- Fischer, O. 1994. "The fortunes of the Latin-type accusative and infinitive construction in Dutch and English compared". In T. Swan, E. Mørck & O. Jansen Westvik (Eds.), *Language Change and Language Structure: Older Germanic Languages in a Comparative Perspective*. Berlin: Mouton de Gruyter, 91-133.
- Givón, T. 1990. *Syntax: A Functional-typological Introduction*, vol. 2. Amsterdam: John Benjamins.
- Givón, T. 1993. *English Grammar: A Function-based Introduction*, vol. 2. Amsterdam: John Benjamins.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goossens, L. 1991. "FG reflections on 'Tobacco is said to be harmful'". *Cahiers de l'Institut de Linguistique de Louvain*, 17 (1-3), 65-74.
- Gries, S. Th. 2004. Coll.analysis 3, R-script for colostruational analysis.
- Gries, S. Th. & Anatol Stefanowitsch. 2004. "Extending colostruational analysis: A corpus-based perspective on alternations". *International Journal of Corpus Linguistics*, 9 (1), 97-129.
- Kay, P. & C. J. Fillmore. 1999. "Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction". *Language*, 75, 1-33.
- Lakoff, G. 1987. *Women, Fire and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar*, vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.
- Lees, R. B. 1960. *The Grammar of English Nominalizations*. Bloomington: Indiana University Press.
- Los, B. 2005. *The Rise of the To-Infinitive*. Oxford: Oxford University Press.
- Moles, J. 1991. "The dramatic coherence of Ovid, *Amores* 1.1 and 1.2". *The Classical Quarterly*, 41 (2), 551-554.
- Montgomery, S. L. 1996. *The Scientific Voice*. New York: The Guilford Press.
- Noël, D. 1997. "The choice between infinitives and *that*-clauses after *believe*". *English Language and Linguistics*, 1 (2), 271-284.
- Noël, D. 2001. "The passive matrices of English infinitival complement clauses. Evidentials on the road to auxiliarihood". *Studies in Language*, 25 (2), 255-296.
- Noël, D. 2003. "Is there semantics in all syntax? The case of accusative and infinitive constructions vs. *that*-clauses". In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of Grammatical Variation in English*. (=Topics in English Linguistics, 43). Berlin: Mouton de Gruyter, 347-377.
- Noël, D. 2008. "The Nominative and Infinitive in Late Modern English: a diachronic constructionist approach". *Journal of English Linguistics*, 36 (4), 314-340.
- Noël, D. & T. Colleman. 2009a. "The nominative and infinitive in English and Dutch: an exercise in contrastive diachronic construction grammar". *Languages in Contrast*, 9 (1), 144-181.
- Noël, D. & T. Colleman. 2009b. "ACI and NCI *find* and the development of evidentiality in Modern English, with a note on Dutch (*be*)*vinden*". In S. Slembrouck, M. Taverniers & M. Van Herreweghe (Eds.), *From Will to Well: Studies in Linguistics Offered to Anne-Marie Simon-Vandenberghe*. Ghent: Academia Press, 337-346.

- Noël, D. & J. van der Auwera. 2009. Revisiting *be supposed to* from a diachronic constructionist perspective. *English Studies*, 90 (5), 599-623.
- Nuyts, J. 2001. *Epistemic Modality, Language and Conceptualization: A Cognitive-pragmatic Perspective*. Amsterdam: John Benjamins.
- Overdiep, G. S. 1935. *Zeventiende-eeuwsche Syntaxis (Derde Stuk)*. Groningen/Batavia: Wolters.
- Postal, P. M. 1974. *On Raising: One Rule of English Grammar and Its Theoretical Implications*. Cambridge, MA: MIT Press.
- Stefanowitsch, A. & S. Th. Gries. 2003. "Collostructions: Investigating the interaction of words and constructions". *International Journal of Corpus Linguistics*, 8 (2), 209-243.
- Van Leuvensteijn, J. A. 1997. "Vroegnieuwlands (circa 1550-1650)". In M. C. van den Toorn et al., *Geschiedenis van de Nederlandse Taal*. Amsterdam: Amsterdam University Press, 227-272.
- Verschuere, J. 1985. *What People Say They Do With Words: Prolegomena to an Empirical-Conceptual Approach to Linguistic Action*. Norwood, NJ: Ablex.
- Warner, A. R. 1982. *Complementation in Middle English and the Methodology of Historical Syntax: A Study of the Wyclifite Sermons*. London: Croom Helm.
- Zajicek, J. 1970. "Réflexions sur l'accusativus cum infinitivo". *Nieuwe Taalgids*, 63, 198-208.

PCU verb	Period I: 1640-1710 (1,978,050 wds)		Period II: 1710-1780 (3,036,325 wds)		Period III: 1780-1850 (5,777,348 wds)		Period IV: 1850-1920 (6,103,660 wds)	
	ACI	NCI	ACI	NCI	ACI	NCI	ACI	NCI
acknowledge	17	9	12	15	21	10	6	6
admit	2	0	6	2	22	14	8	16
affirm	7	1	8	1	8	3	7	0
allege	1	0	0	1	2	3	0	7
announce	0	0	0	0	1	2	1	5
apprehend	4	0	6	1	1	1	0	0
approve	0	0	1	3	0	0	0	0
ascertain	0	0	0	0	5	2	2	2
assert	2	0	11	1	5	5	3	1
assume	0	0	0	0	1	3	18	11
avow	0	0	0	1	2	1	0	0
believe	36	2	47	6	161	24	142	31
calculate	0	0	1	1	1	0	2	7
certify	0	0	0	0	0	0	0	1
claim	1	0	0	0	0	1	1	4
compute	1	3	2	2	0	2	0	0
conceive	24	8	16	6	57	9	19	6
conclude	17	1	12	2	15	0	2	0
confess	20	2	4	7	6	1	2	0
conjecture	0	0	1	0	3	1	2	1
consider	5	1	6	3	40	9	29	21
declare	16	4	11	2	29	10	19	14
decree	0	0	0	0	0	1	0	0
deem	1	0	0	0	9	5	2	1
demonstrate	1	2	0	0	1	3	1	4
deny	17	0	7	5	1	0	2	0
describe	2	0	1	0	6	5	2	0
detect	0	0	0	1	0	0	0	0
determine	0	0	1	2	2	0	0	0
discern	1	2	0	0	1	0	1	0
discover	3	2	20	6	17	5	12	6
establish	0	0	1	0	0	0	0	0
esteem	4	1	5	2	2	0	1	0
estimate	0	0	1	0	1	0	0	2
expect	10	2	24	20	99	118	208	139
experience	0	0	4	5	0	0	0	0
fancy	2	0	6	0	5	0	8	1
feel	1	1	4	0	40	6	69	9
find	108	19	132	76	52	137	37	92
grant	7	5	2	0	3	0	2	0
guarantee	0	0	0	0	0	1	0	1
guess	2	1	3	0	9	0	6	0
hold	9	1	8	5	25	15	31	38
imagine	23	5	23	13	39	2	33	5
judge	19	9	28	2	8	4	10	3
know	63	37	73	48	169	128	118	103
note	0	7	0	0	1	1	1	1
notice	0	0	0	0	0	0	2	1
observe	24	27	23	12	9	13	5	8

perceive	17	5	15	0	18	6	5	4
presume	5	14	0	2	3	6	2	3
presuppose	0	0	0	0	0	0	1	0
pretend	10	10	1	2	4	7	1	0
proclaim	2	1	0	0	3	0	1	0
profess	6	0	1	0	5	0	1	0
pronounce	0	1	8	3	29	9	15	5
propound	0	2	0	0	0	0	0	0
prove	17	6	16	6	34	22	21	15
reckon	0	1	3	4	1	1	0	2
recognise	0	0	3	0	3	1	2	0
report	2	5	0	17	2	12	2	19
repute	0	3	0	1	0	3	0	6
reveal	0	0	0	1	0	0	1	0
rumour	0	1	0	0	0	0	0	2
say	4	157	0	243	0	318	0	196
see	12	17	10	14	3	31	19	45
show	7	3	13	0	13	7	36	18
state	1	0	1	2	11	14	5	6
suppose	33	70	120	224	137	218	101	196
surmise	0	0	0	0	1	0	0	0
suspect	4	5	16	3	16	1	9	1
take	58	15	76	3	26	2	33	15
think	31	56	11	46	16	43	18	36
understand	5	27	5	17	12	25	21	16
TOTAL (absolute)	664	551	809	839	1216	1271	1108	1132
<i>(normalized per million words)</i>	<i>335.7</i>	<i>278.62</i>	<i>266.44</i>	<i>276.38</i>	<i>210.49</i>	<i>220</i>	<i>181.51</i>	<i>185.43</i>

Table 1: observed frequency of the ACI and the NCI in the four CLMET sub-corpora

PCU verb	Period I: 1640-1710 (1,188,932 wds)		Period II: 1710-1780 (2,471,692 wds)		Period III: 1780-1850 (2,625,226 wds)		Period IV: 1850-1920 (3,518,374 wds)	
	ACI	NCI	ACI	NCI	ACI	NCI	ACI	NCI
achten 'consider'	10	0	2	1	7	5	4	9
bedenken 'think of'	0	1	0	0	0	0	0	0
begrijpen 'understand'	0	0	9	1	0	0	0	0
bekennen 'confess'	0	0	2	0	1	0	0	0
beschouwen 'consider'	0	0	0	0	0	1	0	1
betonen 'show, prove'	0	0	0	1	0	0	0	0
betuigen 'express'	0	0	1	0	0	0	0	0
bevinden 'find'	16	0	15	4	2	6	0	2
beweren 'state, claim'	1	0	4	0	0	0	1	0
bewijzen 'prove'	0	1	4	0	2	1	0	0
considereren 'consider'	0	3	0	0	0	0	0	0
denken 'think'	4	0	12	0	2	0	1	0
erkennen 'recognize'	1	0	3	0	1	1	0	0
geloven 'believe'	5	0	15	0	2	0	1	0
gevoelen 'feel'	1	0	0	0	0	0	0	0
hopen 'hope'	5	0	2	0	0	0	0	0
houden 'hold, consider'	1	0	1	1	1	3	0	1
kennen 'know'	1	0	2	0	0	0	0	0
menen 'be of the opinion'	5	0	15	0	4	0	5	0
merken 'perceive'	3	0	0	0	0	0	0	0
noemen 'call'	0	0	2	0	0	0	0	0
ondervinden 'experience'	0	0	3	0	0	0	0	0
ontkennen 'deny'	0	0	1	0	0	0	0	0
oordelen 'judge'	32	1	33	5	8	1	0	0
rekenen 'estimate'	0	0	3	5	4	6	0	1
schatten 'estimate'	0	1	0	0	0	0	0	0
schrijven 'write'	1	0	0	0	0	0	0	0
stellen 'state'	1	0	1	0	1	0	0	0
sustineren 'assume'	0	0	1	0	0	1	0	0
verhalen 'tell'	1	0	0	0	0	0	0	0
verklaren 'declare'	0	0	4	0	5	3	0	0
vermoeden 'suppose'	0	0	1	0	1	0	1	0
vernemen 'be told'	1	0	0	0	0	0	0	0
(ver)onderstellen 'suppose'	0	1	8	9	2	1	3	5
vinden 'find'	4	0	2	1	2	1	0	0
voorgeven 'profess'	0	0	1	2	1	0	0	0
voorstellen 'present, imagine'	0	0	1	1	0	0	0	0
voorwenden 'pretend'	0	0	0	2	0	0	0	0
verstaan 'understand'	0	1	0	0	0	0	0	0
wanen 'presume'	2	0	5	0	1	0	0	0
weten 'know'	5	0	5	0	0	1	0	0
zeggen 'say'	20	2	14	20	6	12	2	9
TOTAL (absolute)	120	11	172	53	53	43	18	28
<i>(normalized per million words)</i>	<i>100.93</i>	<i>9.25</i>	<i>69.59</i>	<i>21.44</i>	<i>20.19</i>	<i>16.38</i>	<i>5.12</i>	<i>7.96</i>

Table 2: observed frequency of the ACI and the NCI in the four sub-corpora of a diachronic corpus of Dutch literary texts

Period I (1640-1710)		Period II (1710-1780)		Period III (1780-1850)		Period IV (1850-1920)	
ACI	NCI	ACI	NCI	ACI	NCI	ACI	NCI
find	say	take	say	believe	say	believe	say
believe	suppose	believe	suppose	conceive	find	feel	suppose
take	understand	judge	think	imagine	see	imagine	find
deny	think	find	report	feel	suppose	expect	report
confess	note	perceive	understand	take	think	take	see
conclude	presume	show		consider	report	conceive	allege
imagine		suspect		conclude	understand	affirm	think
conceive		assert		suspect		show	repute
declare		discover		pronounce		suspect	
hold		conclude		declare		guess	
perceive		know		guess		fancy	
profess		declare		know		pronounce	
expect		fancy		discover		judge	
prove		affirm		perceive		know	
know		conceive		acknowledge			
		prove		fancy			
		observe		profess			
				prove			

Table 3: significant distinctive collexemes of the English ACI and NCI in the four CLMET sub-corpora, in order of diminishing collocation strength

active collexemes	passive collexemes
<i>have</i>	<i>base</i>
<i>think</i>	<i>concern</i>
<i>get</i>	<i>use</i>
<i>say</i>	<i>involve</i>
<i>want</i>	<i>publish</i>
<i>do</i>	<i>associate</i>
<i>know</i>	<i>bear</i>
<i>see</i>	<i>engage</i>
<i>mean</i>	<i>design</i>
<i>like</i>	<i>confine</i>
<i>try</i>	<i>entitle</i>
<i>hope</i>	<i>relate</i>
<i>believe</i>	<i>deposit</i>
<i>remember</i>	<i>compare</i>
<i>feel</i>	<i>derive</i>
<i>suppose</i>	<i>deal</i>
<i>wish</i>	<i>aim</i>
<i>thank</i>	<i>release</i>
<i>enjoy</i>	<i>attach</i>
<i>ensure</i>	<i>store</i>

Table 4: Collexemes distinguishing between active and passive (culled from Table 3 in Gries & Stefanowitsch 2004:109)