Informalization in Dutch journalistic subgenres over time

Abstract: This paper investigates the claim that a trend of informalization can be found in public discourse. This hypothesis is studied in the context of Dutch news reports: recent news reports show a greater amount of subjective elements than older reports. To test the hypothesis, a corpus linguistic analysis of five major Dutch newspapers from 1950/1 and 2002 was carried out, in which the frequencies of subjectivity indicators were established in three different journalistic subgenres: hard news, opinion, and background news. On the basis of a pilot study, a distinction was made between reporter’s text and quoted text. This proved central to our findings: overall, the more recent news reports do not display a higher degree of informalization; reporter’s text even shows a decrease of subjective elements. In contrast, quoted text shows a higher degree of informalization across the board. Although subgenre was involved in various interactions with year of publication, the pattern was too irregular to reach firm conclusions about the role of subgenre in the informalization process.

Keywords: Biber; conversationalization; corpus analysis; diachronic change; Dutch national newspapers; informalization; journalism; journalistic subgenres; logistic regression; personalization; quantitative analysis; quotation; subjectivity, linguistic features of

1 Introduction

It has been suggested that “a major change in discursive practices affecting many public institutions in contemporary society is the ‘conversationalization’ of public discourse […]” (Fairclough and Wodak 1997, p. 265). This change,

1 We thank three anonymous reviewers for their valuable comments on a previous version of this paper, and Lieke Verheijen for correcting our English. We thank Onno Huber (VU University Amsterdam) for his assistance in assembling and analysing the corpus, and Gerben Mulder (VU University Amsterdam) for his help in the statistical analyses.

K. de Haan-Vis, Dedicon, Grave
W. Spooren, Centre for Language Studies, Radboud University Nijmegen
also indicated as ‘informalization’, involves “the modelling of public discourse upon the discursive practices of ordinary life, ‘conversational’ practices in a broad sense” (Fairclough 1994, p. 253). It entails a change in the style and register of many genres, among which journalistic prose, since World War II.

Journalistic prose consists of many subgenres, such as commentaries, news reports, and interviews. In this chapter, we take genres and subgenres to be experience-based, conventionalized conceptualizations of fixed combinations of language or code, context and text, much along the lines of Steen (2011). We regard genre knowledge as schematic knowledge: becoming experienced in a discourse culture means learning the combinations of language, goals and forms of that culture, to perform certain communicative tasks. Typically, a western news report brings a more or less objective report of what happened, how it happened, who was involved, and what the consequences were (context). Such a report is typically expressed in fairly objective language, with a strong focus on the ideational function (Halliday, 1974) (code). The typical discourse form to express such a news event has an inverted structure, with the main news occurring first, followed by additional details and background information (text; Knobloch et al. 2004). Knowledge of the genre combinations helps us to process information quickly and efficiently. Such knowledge is typically culture dependent: for example, Iranian journalists frequently use chronological ordering in delivering news reports (Rafiee, in preparation).

This chapter focuses on genre change at the level of the language, or code. From a linguistic viewpoint, what the changes toward informalization have in common is a shift in attention from the object, i.e. the news, to the subject, i.e. the journalist and sources he or she quotes in the news report. In other words, the informalization hypothesis presumes that the journalist and the news sources have become more central as speakers. This centrality of speakers is linguistically reflected in the expression of self: the perspective of the writer or another subject of consciousness mentioned in the text. These are all aspects of what in linguistic theorizing has become known as subjectivity (Lyons 1982; Langacker 1990; Traugott 1989). Note that studying the informalization hypothesis in terms of changes in expression of linguistic subjectivity is by no means the only way to study the process of informalization. Others have looked at, for example, changes in lexis (such as the use of metaphors; Pasma 2011) or grammatical developments (e.g. the use of contractions and questions; Pearce 2005). Our focus is on subjectivity because it is a linguistic phenomenon that is particularly apt at reflecting changes with respect to informalization: as will be discussed below, there seems to be a strong link between informal and subjective language.

Hundt and Mair (1999) described journalistic prose as an ‘agile’ genre compared to academic prose, in that journalistic prose is more open to innovations
such as ‘colloquialization’, as observed in increases of first and second person pronouns, contractions, etc. Since informalization would imply an increase over time of “conversation” between two voices, the journalist and his/her reader, a crucial question is whether these voices have become more prominent over time.

Another question is whether the process of informalization is homogeneous across subgenres. Journalistic texts can be written in various subgenres; some of them are generally considered to be objective reports of news events, whereas others allow for an opinionated voice from the journalist (news commentaries, for example). An in-between subgenre is the background article, which typically expands on news reports that were published earlier, and allows the writer more freedom.

A journalistic text is multivoiced: when describing a news event, a journalist often quotes sources for reasons of vividness (quotes are often seen as a means of enlivening a text) and distancing (the journalist may not want to take responsibility for a particular piece of information and may therefore attribute it to a specific source). To investigate the multivoicedness of journalistic texts, a distinction was made in this paper between quoted text and journalist’s text (or: ‘reporter’s text’).

The linguistic analysis of journalistic subgenres over time is a particularly interesting case in point for the present volume’s goal for several reasons. Firstly, such a diachronic investigation of changes in linguistic features presents a view on ‘genres-in-development’: how is genre change expressed linguistically? Secondly, the scale of the analysis, in terms of both corpus size and the number of linguistic features analyzed, plus the rigid statistical comparison of the patterns over time, newspapers and subgenres, makes this an enterprise that could be a model for future linguistic analyses of genre change. Hence, our research also has strong methodological value.

2 Theoretical background

For the analysis of subjective elements, we followed a very pragmatic approach by selecting operationalizations mentioned in the literature (see also Vis, Sanders and Spooren 2012). More specifically, we selected features described by Bekker (2006), Biber (2004), Scheibman (2002), Wiebe (1994) and Wiebe et al. (2005). The list was completed by the descriptions of the grammatical categories of stance and modality in the grammars of Biber et al. (1999) and Haeseryn et al. (1997). The appendix gives a detailed operationalization of our subjectivity lexicon. Here we list the general categories.
1. First and second person pronouns explicitly refer to the writer or the addressee, or, in case of a quotation, to a character mentioned in the text. That is why many authors take them as clear indicators of subjectivity (a.o. Scheibman 2002).

2. Modal elements (such as adverbs [possibly], verbs [must] and particles [only]) reflect the writer's attitude toward what is expressed in the text. As a consequence, they have been mentioned as indicators of subjectivity (a.o. Biber, Johansson, Leech, Conrad, and Finegan 1999; Haeseryn et al. 1997; Precht 2000).

3. Intensifiers or degree adverbs express a writer's estimate, on an intensity scale, of the degree to which an utterance is true (Bolinger 1972) and are often tied up to evaluations by a subject of consciousness (Thompson and Hunston 2000).

4. Cognitive verbs like know, doubt and answer “can be interpreted as expressing author/speaker attitude and assessment of likelihood” (Biber 2004). As a result, they reflect a subject’s point of view.

5. Exclamations and questions also reflect a subject’s point of view, as they suggest that a subject expresses his/her mood or addresses a communicational partner. As such, they have been suggested as elements reflecting subjectivity (e.g. Wiebe 1994).

6. Deixis positions an utterance with respect to the here and now of the current writer/speaker. As a consequence, it reflects the presence of a subject of consciousness, which is why it has been mentioned as reflecting epistemological stance (e.g. Bekker 2006).

Several of these lexicogrammatical indicators of subjectivity have been included in one or more diachronic studies. This section discusses the most prominent of those studies. As will be seen, many indicators show an increase over time, suggesting a shift in written discourse toward more conversational and informal norms. Table 1 presents a summary of these findings.

A number of these studies focus on news genres. These are discussed in the remainder of this section. Westin and Geisler (2002) used Biber's (1988) multi-feature/multi-dimensional model and the concomitant characteristic features for a study of informalization in British newspaper discourse. In an analysis of 554 editorials from three up-market British newspapers published between 1900 and 1993, they found that the editorials became less abstract (dimension 5) and less dependent on referential elaboration (dimension 3, marked by such features as an increase of temporal and spatial adverbials), which resulted in more informal language. Similar to Biber and Finegan's (2001) findings for news
reportage, no regular pattern was observed with respect to dimension 1 (distinguishing between involved and informational discourse, marked by such features as private verbs, first and second person pronouns, hedges, amplifiers, ‘wh’-questions, and possibility modals). Westin and Geisler’s analyses indicate that the changes mainly took place during the latter part of the twentieth century, since the last period included in their study – representing the years 1960 up to 1993 – differs from the preceding two periods on three of the dimensions. Steen (2003) provided a study similar to Westin and Geisler’s: he, too, used Biber’s (1988) multi-feature/multi-dimensional model to study the stylistic changes in editorials from the British press, in this case The Times. The period of publication of the 120 editorials (85,000 words) overlaps with the last period from Westin and Geisler’s study: 1950 to 2000. Steen investigated four styles, labeled as “involved”, “informational”, “narrative” and “persuasive”, representing Biber’s dimension 1, 2, and 4. In an analysis of a limited number of features

Table 1: Summary of findings in previous diachronological studies

<table>
<thead>
<tr>
<th>↑</th>
<th>↓</th>
<th>↔</th>
</tr>
</thead>
<tbody>
<tr>
<td>intensifiers</td>
<td></td>
<td>Pearce (2005); Westin and Geisler (2002)</td>
</tr>
<tr>
<td>semi-modals</td>
<td>Biber (2004); Leech et al. (2009)</td>
<td></td>
</tr>
</tbody>
</table>

Note: ‘↑’: increase; ‘↓’: decrease; ‘↔’: no change; studies of news reports are printed in boldface.
per style (including the subjectivity indicators first and second person pronouns for the “involved” style, and modal and semi-modal verbs for the “persuasive” style), Steen found an increase in involvement and persuasion and a decrease in what he, following Biber (1988), marks as narrativity: past tense, perfect aspect and third person. This can be interpreted as a stylistic change toward conversations, as these are typically involved and persuasive in nature and show few markings of narrativity.

Similar to Biber and Finegan (2001), Hundt and Mair (1999) also observed a difference in the shift toward a more involved and oral style between several written genres. In a study of news and academic prose (320,000 words) from the Brown corpora, Hundt and Mair (1999) found evidence of a “colloquialization” of the norms of written English, that is, a narrowing of the gap between spoken and written norms. However, they also observed that the two genres differ in their openness to innovation: news is relatively receptive to innovation (“agile”), whereas academic prose is more reluctant to change (“uptight”). This openness to innovation in journalistic prose is shown by an increase in the use of first and second person pronouns, contractions, sentence-initial conjunctions, phrasal verbs, and progressive aspect.

Using the same corpora, Leech et al. (2009) described the diachronic change of a wide range of grammatical topics, including modal and semi-modal verbs, progressives, subjunctives, passives, genitives, and relative clauses in contemporary written English. Leech et al. (2009) found evidence for colloquialization in several features, such as an increase of contractions, ‘not’-negation versus ‘no’-negation, progressives, ‘get’-passives, a decrease of ‘be’-passives, a decline of ‘wh’-relative clauses and an increase of ‘that’- and zero relative clauses. There was also an increase of first and second person pronouns in the American English corpora, an increase of questions in the corpus overall (although they decreased in the press section), an increase in the use of a representative set of so-called semi-modals (e.g. ‘be going to’, ‘(had) better’, ‘need to’), and a substantial decline in the frequency of so-called core modals (e.g. ‘will’, ‘would’, ‘must’, ‘might’). However, Leech et al. (2009: 245) observed that colloquialization does not occur in all cases: for example, they were surprised to find that the use of exclamation marks has declined over time in both American and British English, and that the use of first person pronouns in British English “shows a mysterious anti-colloquial trend” (Leech et al. 2009: 247).

Biber (2004) observed that in news, drama, personal letters, and medical prose speakers and writers have become more inclined to express their opinions, attitudes and emotions, especially over the last fifty years of the study. In an analysis of 560 texts from the ARCHER corpus from a period from 1650 to 1990,
he investigated three major structural types of stance marking from Biber et al. (1999): modal and semi-modal verbs, stance adverbials (e.g. ‘hopefully’, ‘undoubtedly’, ‘evidently’), and stance complement clause constructions (e.g. ‘suggest that’, ‘expect that’, ‘possible that’, ‘hope to’, ‘believe to’, ‘easy to’). The results differed between the four genres: modal verbs declined in all genres, whereas semi-modal verbs showed an increase, but this was restricted primarily to drama and personal letters. Stance adverbials increased steadily in use across all genres in the nineteenth century, but in the most recent period this increase did not continue in news and medical prose. Finally, the complement clause constructions also underwent a steady increase in frequency across the periods, the shift toward greater use of stance verb + complement being especially characteristic of drama and personal letters, and the stance adjective/noun + complement of newspapers and to a lesser extent medical prose.

In sum, the discussion above shows that there are several lexicogrammatical indicators of subjectivity that have been included in one or more diachronic studies. Many of the indicators have seen an increase over time, which suggests a shift in written discourse toward more conversational and informal norms.

None of the studies mentioned have investigated the informalization hypothesis for Dutch, and none have distinguished between subgenres in newspaper texts. Our present study aims to fill that gap.

3 Method

3.1 Corpus selection

To answer our research questions, a diachronic corpus of Dutch written news data was compiled, representing different subgenres and different time periods. News reports were selected from five Dutch national newspapers from the periods 1950/1951 and 2002. This interval reflects two pivotal instances in societal development: the first point is directly after World War II and before the presumed informalization development, while the second point is after the presumed informalization. The year 2002 was chosen because it fell between two relevant developments within journalism: after a period of professionalization, but before the broad spread of digitization of journalism. 2002 was also a relevant year because in a different study, the data from the news corpus have been compared with those from conversation data in the so-called Corpus of Spoken Dutch, which was collected around 2002 (see Vis 2011).
The newspapers selected for the corpus were national newspapers published in both periods of investigation. Our choice of newspapers reflects the distinction between quality newspapers and more tabloid-like newspapers (although the difference is much smaller in the Netherlands compared to, for example, the United Kingdom; Deuze 2005). *NRC Handelsblad*, *Volkskrant* and *Trouw* are generally considered quality newspapers, whereas *De Telegraaf* and *Algemeen Dagblad* are more tabloid-like.

The data for 1950/1 were collected from prints in the libraries of VU University Amsterdam and the University of Amsterdam. The data for one newspaper (*Algemeen Dagblad*) came from high-quality scans made available by the National Library (Koninklijke Bibliotheek). The articles were selected semi-randomly: for each newspaper, three weeks were chosen; all editions of that week were included in the corpus until a total of 50 pages was reached. The weeks were more or less evenly distributed over the year. If a page contained too many advertisements, it was disregarded and a new page was selected. The pages were digitized using OCR. Post-processing was applied to eliminate recognition errors.

The data for 2002 were collected from the LexisNexis database. For each newspaper, a week was selected semi-randomly. From each specific week, all articles of that week were included in the corpus. The weeks were more or less evenly distributed over the year.

The articles selected from the corpus have been aggregated into three subgenres: hard news, opinion news and background news (following Bell 1991). Hard news encompasses front page news, domestic news, foreign news, and business news; opinion news comprises editorials, opinion news, and arts (reviews); background news contains the categories sports, and science, and the section on 'showbiz' news in *De Telegraaf* from 2002 and the articles in the 1950s newspapers categorized as miscellaneous.

Both corpora were annotated automatically for part of speech and lemma information using the Tadpole tagger (Van den Bosch et al., 2007). During this phase, the texts were also transformed from a text format to an XML format. Because of its unique character, the resulting corpus had been made available through a so-called Clarin action as the VU-DNC corpus (cf. http://tst-centrale.org/nl/producten/corpora/vu-dnc-corpus/6-79).

The corpus consists of approximately 1,900,000 words in 6,618 news reports from three different subgenres (hard news, opinion news, background news). Table 2 presents an overview of the corpus.
Table 2: Corpus size (in words)

<table>
<thead>
<tr>
<th></th>
<th>1950/1951</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hard news</td>
<td>Opinion</td>
<td>Background</td>
<td>Total</td>
</tr>
<tr>
<td>Algemeen Dagblad</td>
<td>90,247</td>
<td>34,798</td>
<td>60,790</td>
<td>185,835</td>
</tr>
<tr>
<td>NRC (Handelsblad)</td>
<td>117,574</td>
<td>36,375</td>
<td>31,041</td>
<td>184,990</td>
</tr>
<tr>
<td>De Telegraaf</td>
<td>105,811</td>
<td>30,516</td>
<td>42,287</td>
<td>178,614</td>
</tr>
<tr>
<td>Trouw</td>
<td>135,761</td>
<td>13,484</td>
<td>32,410</td>
<td>181,655</td>
</tr>
<tr>
<td>De Volkskrant</td>
<td>115,648</td>
<td>23,954</td>
<td>60,878</td>
<td>200,480</td>
</tr>
<tr>
<td>Total</td>
<td>565,041</td>
<td>139,127</td>
<td>227,406</td>
<td>931,574</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2002</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Hard news</td>
<td>Opinion</td>
<td>Background</td>
</tr>
<tr>
<td>Algemeen Dagblad</td>
<td>117,083</td>
<td>28,046</td>
<td>75,951</td>
<td>221,080</td>
</tr>
<tr>
<td>NRC (Handelsblad)</td>
<td>164,416</td>
<td>53,846</td>
<td>41,767</td>
<td>260,029</td>
</tr>
<tr>
<td>De Telegraaf</td>
<td>76,658</td>
<td>20,755</td>
<td>34,700</td>
<td>132,113</td>
</tr>
<tr>
<td>Trouw</td>
<td>86,270</td>
<td>40,551</td>
<td>22,570</td>
<td>149,391</td>
</tr>
<tr>
<td>De Volkskrant</td>
<td>114,484</td>
<td>46,687</td>
<td>47,275</td>
<td>208,446</td>
</tr>
<tr>
<td>Total</td>
<td>558,911</td>
<td>189,885</td>
<td>222,263</td>
<td>971,059</td>
</tr>
</tbody>
</table>

3.2 Subjectivity analysis

An automatic analysis was made of indicators of subjectivity at the lexicographical level. Drawing on previous research on subjectivity and stance, a list of lexicographical indicators of subjectivity was compiled. Examples are linguistic features such as first and second person pronouns, modal elements (verbs, adverbs, and particles) and intensifiers. An overview of the list of features used for this analysis can be found in the Appendix.

On the basis of the literature and previous analyses, a distinction was made between lexicographical indicators that reflected reporter subjectivity (found in the reporter’s words) and lexicographical indicators that reflected source subjectivity (found in direct quotations).

3.3 Identifying quoted words

In order to be able to distinguish between the reporter’s text and the quoted text, the news reports were marked for direct quotations on the basis of quotation marks. The choice for direct quotation over, for example, indirect and free indirect discourse was made deliberately: the quotation marks are an explicit indication that the journalist is not responsible for the exact wording of the quote, so that responsibility, including the choice for subjective elements, lies with the quoted source. We realize that relying on quotation marks is a fairly
rough measure to indentify direct quotations. However, the size of the corpus did not allow us to identify the quotations manually. To ensure that this method was reliable, samples of quotation annotations were checked manually. In a total of 70,000 words, only two percent was annotated incorrectly. This was considered sufficient for present purposes.

3.4 Statistical analysis

Three subsets of materials were distinguished: the complete texts (the text “overall”), the reporter’s text, and the direct quotations. For each of the three subsets and for each of the subjectivity markers, a logit analysis was carried out, in which for each text the number of observed occurrences of subjectivity markers was related to the number of words in the text. A logit analysis is a type of logistic regression in which the probability of frequencies is calculated using the logit function: the logit of the number of words logit(F) is defined as log(F/(N-F)), where F is the frequency of occurrence of the subjectivity markers and N is the number of words in the text/reporter’s text/direct quotations in the text. A logit analysis was chosen because it can deal with samples that differ greatly in the frequency of the phenomenon under investigation. Moreover, such an analysis allows one to study interactions between predictors. In that sense, it resembles an analysis of variance.

In the “overall” analysis, the number of occurrences was related to the total number of words in the text; in the analysis of reporter’s words, it was related to the total number of reporter words; and in the analysis of quoted words, it was related to the total number of quoted words. The predictor variables in the analysis were Year (1950/1951, 2002), Newspaper (Algemeen Dagblad, NRC (Handelsblad), De Telegraaf, Trouw, and De Volkskrant) and Subgenre (hard news, opinion, background news). Texts were treated as a random factor. For the analysis the MLwiN software (version 2.18) was used.

The analysis was carried out using backward elimination. Only those terms were kept that contributed significantly to the model. If the main effect of Year was in the final model, a separate χ²-analysis of Year was carried out.

In the following section, we give an overall overview of the results, one for each subset of materials (overall, reporter’s text, quoted text). Space does not allow us to give all results, so we restrict ourselves to findings in which the factor Year is involved. That implies that main effects of Subgenre and Newspaper and interactions between Newspaper and Subgenre were disregarded. For a complete overview of the results and statistical tests, we refer to Vis (2011), especially chapter 5 and appendix B.
4 Results

4.1 ‘Overall’ analysis

Table 3 gives a summary of the results for the overall texts. An increase from 1950/1951 to 2002 is indicated by an upward arrow (‘↑’), a decrease is indicated by a downward arrow (‘↓’), and a lack of change is indicated by a horizontal bidirectional arrow (‘↔’). Arrows that are underlined indicate statistical significance. For each of the indicators, the effects in which the year of publication is involved are listed; the corresponding development is displayed in the adjacent column.

For example, for first person singular pronouns, there is a main effect of year (Y) which shows an increase (the arrow in the adjacent column), whereas for second person singular pronouns, there are two two-way interactions in which year of publication is involved: one interaction with newspaper (N) and one with Subgenre (S). The corresponding developments are listed in alphabetical order of the name of the newspaper or subgenre: for newspaper, this is *Algemeen Dagblad*, *NRC (Handelsblad)*, *De Telegraaf*, *Trouw*, and *De Volkskrant*, so the first arrow in the column next to Y*N* for second person singular pronouns is for *Algemeen Dagblad* (a ‘↑’ for increase), the second arrow for *NRC (Handelsblad)* (another ‘↑’), and so on. For the subgenres, the order is ‘hard news’, ‘opinion news’, ‘background news’; hence, for the interaction between year and subgenre the first arrow is for hard news (also a ‘↑’). In case of a three-way interaction, there are three times five arrows; the first set of five is for the five newspapers in alphabetical order in hard news (‘h’), the second set of five for the newspapers in opinion news (‘o’), and the final set for the newspapers in background news (‘b’). The arrows are not based on significance, but rather on direction of the development.

The results of the subjective indicators in the text “overall” can be summarized as follows. For all elements, the year of publication has a significant effect, either as a main effect or in interaction with newspaper and/or subgenre. The only exception is the exclamation mark, which is not used differently over time. The newspapers are mutually divergent in how they use exclamation marks, as are the subgenres, yet these differences do not change over time.

The overview in Table 3 demonstrates that many elements show increases in all (or nearly all) newspapers and subgenres, i.e. first person singular pronouns, second person singular pronouns, cognitive verbs, intensifiers, and modal particles. For the first person singular pronouns, there is a main effect of year; for the second person singular pronouns, the frequency in all newspapers and subgenres is higher in 2002 than in 1950/1; cognitive verbs are used more in 2002...
by all newspapers in hard news, by one newspaper in opinion, and by all but one newspaper in background news; all newspapers and all subgenres contain more intensifiers in 2002; and modal particles are used more in 2002 in all newspapers in all subgenres. These increases are in line with the hypothesis that newspapers have undergone a process of informalization.

Table 3: Overview of developments of the indicators of subjectivity in the texts, overall

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Rel. freq. 1950/1</th>
<th>Rel. freq. 2002</th>
<th>Effect</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pronoun 1p. sg.</td>
<td>19.92</td>
<td>61.29</td>
<td>Y</td>
<td>↑</td>
</tr>
<tr>
<td>2. Pronoun 2p. sg.</td>
<td>10.37</td>
<td>29.91</td>
<td>YxN</td>
<td>↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>3. Pronoun 1p. pl.</td>
<td>31.25</td>
<td>28.49</td>
<td>YxN</td>
<td>↓↓↑↓↑</td>
</tr>
<tr>
<td>4. Deictic element</td>
<td>41.18</td>
<td>40.21</td>
<td>YxNxS</td>
<td>↓↑↑↑↑</td>
</tr>
<tr>
<td>5. Cognitive verb</td>
<td>87.66</td>
<td>111.76</td>
<td>YxNxS</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>6. Exclamation mark</td>
<td>4.20</td>
<td>3.49</td>
<td>–</td>
<td>↔</td>
</tr>
<tr>
<td>7. Question mark</td>
<td>10.11</td>
<td>14.36</td>
<td>YxNxS</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>8. Intensifier</td>
<td>49.02</td>
<td>65.97</td>
<td>YxN</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>9. Modal adverb</td>
<td>25.11</td>
<td>24.32</td>
<td>YxN</td>
<td>↓↓↑↑↑</td>
</tr>
<tr>
<td>10. Modal particle</td>
<td>98.35</td>
<td>143.83</td>
<td>YxNxS</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>11. Modal verb</td>
<td>169.06</td>
<td>165.85</td>
<td>YxNxR</td>
<td>↓↓↔↑↑</td>
</tr>
</tbody>
</table>

Y = year of publication, N = newspaper, S = subgenre; h = hard news, o = opinion news, b = background news; ↑ = increase, ↓ = decrease, ↔ = no effect of year, so no change. Significant changes (p < .05) are underlined. Relative frequency: frequency per 10,000 words. Note that these relative frequencies generalize over newspapers and subgenres.
There are no elements for which all newspapers and subgenres uniformly decrease their use over time. The most convincing case of decrease are modal verbs: in hard news, two newspapers use significantly fewer modal verbs in 2002 than in 1950/1, and in opinion news, even more newspapers show a decrease in the number of modal verbs used. However, in background news, four out of five newspaper show an increase in their use of modal verbs.

The remaining elements show an even less straightforward pattern: the incidence of first person plural pronouns is higher in 2002 than in 1950/1 in some newspapers, whereas it is higher in 1950/1 in others. The same is true for the three subgenres: in one subgenre, more first person plural pronouns are used in 2002 than in 1950/1, whereas in the other two subgenres, more are used in 1950/1. Similarly, the newspapers show no uniform pattern for deictic elements: in both hard news and background news, some newspapers increase their use of deictic elements over time, whereas others decrease or show no change at all. The incidence of question marks increases in most newspapers in hard news and opinion news, but the results are more ambiguous in background news: in two newspapers, the incidence increases, in two others, it decreases, and in one, it remains unchanged. Finally, the modal adverbs show increases and decreases both in newspapers and in subgenres. These ambiguous results do not meet the expectations that the indicators of subjectivity would show an increase over time.

In sum, it seems that in the text “overall”, some indicators provide evidence in favour of informalization, whereas other indicators reveal mixed results: only some newspapers and subgenres show informalization and others do not.

4.2 Analysis of reporter’s text

In the following two sections, a distinction is made between words from the journalist/reporter and words from news sources. This section reports the results for reporter’s text. Table 4 summarizes the results.

The results for the indicators in the words of the reporter show that all elements change in frequency, either increasing or decreasing over time, with the exception of first person singular pronouns. In contrast with the changes in the analysis “overall”, all changes in the words of the reporter are rather uniform across newspapers and subgenres.

Five indicators show a rather uniform increase over time for (almost) all newspapers and subgenres: second person singular pronouns, cognitive verbs, question marks, intensifiers and modal particles. The previous section has shown that these indicators also increase in the text “overall”. In the reporter's text, the
second person singular pronouns are used more in 2002 than in 1950/1 by all newspapers but *De Telegraaf* and in all subgenres except for background news.

Similarly, the incidence of cognitive verbs increases in all newspapers in hard and background news (except for *Trouw* and *De Telegraaf*, in the latter subgenre), although the changes in opinion news are slightly more ambiguous.

---

**Table 4:** Overview of developments of the 11 indicators in the reporter’s text

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Rel. freq. 1950/1</th>
<th>Rel. freq. 2002</th>
<th>Effect</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun 1p. sg.</td>
<td>13.89</td>
<td>14.50</td>
<td>–</td>
<td>↔</td>
</tr>
<tr>
<td>Pronoun 2p. sg.</td>
<td>7.58</td>
<td>13.57</td>
<td>YxN</td>
<td>↑↑↓↑↑</td>
</tr>
<tr>
<td>Pronoun 1p. pl.</td>
<td>27.05</td>
<td>9.12</td>
<td>YxNxS</td>
<td>h ↓↓↓↓↑</td>
</tr>
<tr>
<td>Deictic element</td>
<td>41.32</td>
<td>37.15</td>
<td>YxNxS</td>
<td>h ↓↑↓↑↑</td>
</tr>
<tr>
<td>Cognitive verb</td>
<td>85.95</td>
<td>102.45</td>
<td>YxNxS</td>
<td>h ↑↑↑↑↑</td>
</tr>
<tr>
<td>Exclamation mark</td>
<td>3.53</td>
<td>2.04</td>
<td>Y</td>
<td>↓</td>
</tr>
<tr>
<td>Question mark</td>
<td>8.94</td>
<td>11.57</td>
<td>YxNxS</td>
<td>h ↑↑↓↑↑</td>
</tr>
<tr>
<td>Intensifier</td>
<td>48.03</td>
<td>59.84</td>
<td>YxN</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>Modal adverb</td>
<td>24.48</td>
<td>19.99</td>
<td>YxN</td>
<td>↓↓↔↓↓↓</td>
</tr>
<tr>
<td>Modal particle</td>
<td>97.06</td>
<td>131.81</td>
<td>YxNxS</td>
<td>h ↑↑↑↑↑</td>
</tr>
<tr>
<td>Modal verb</td>
<td>165.86</td>
<td>148.72</td>
<td>YxN</td>
<td>↓↓↓↓↓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YxS</td>
<td>↓↓↓↓</td>
</tr>
</tbody>
</table>

Y = year of publication, N = newspaper, S = subgenre; h = hard news, o = opinion news, b = background news; ↑ = increase, ↓ = decrease, ↔ = no effect of year, so no change.

Significant changes (p < .05) are underlined. Note that these relative frequencies generalize over newspapers and subgenres.
Question marks also increase in all but one newspaper, in hard news and opinion news. Furthermore, in all newspapers and subgenres, intensifiers are used more in 2002 than in 1950/1. Finally, the incidence of modal particles increases uniformly across all newspapers in all subgenres. These increases correspond to the informalization hypothesis. It seems that over the course of the second half of the twentieth century, reporters have become more inclined to express interaction with the addressee (in second person singular pronouns and questions) and interpretation through opinions, attitudes and beliefs through some of the subjectivity indicators (cognitive verbs, intensifiers, and modal particles).

However, contrary to the analysis of the text “overall”, which showed a mixed result for the remaining indicators, there is a uniform decrease between 1950/1 and 2002 in five elements in the reporter’s text: first person plural pronouns, deictic elements, exclamation marks, modal adverbs, and modal verbs. First person plural pronouns decrease uniformly in all newspapers in all subgenres, although the rates of decrease are not the same. For deictic elements, the decrease is not as consistent: in hard news, all but two newspapers (NRC (Handelsblad) and Trouw) use less deictic elements over time, in opinion news, this is the case in all newspapers but De Telegraaf, and in background news, in all but two newspapers (NRC(Handelsblad) and De Telegraaf) the use of deictic elements decreases. For exclamation marks, there is a general pattern of decrease, which does not differ between the newspapers or subgenres. Modal adverbs show a decline in frequency in all newspaper but De Telegraaf, which does not change its use over time. Finally, modal verbs occur less in 2002 than in 1950/1 in all newspapers and all subgenres. The only element which does not manifest a change in frequency of occurrence is the first person singular pronoun.

The decreases found in this analysis for first person plural pronouns, deictic elements, and modal adverbs and the stability over time of first person singular pronouns were unexpected on the basis of the previous diachronic studies. However, those studies did not examine the reporter’s text separately. Our results suggest that, contrary to the informalization hypothesis, reporters have become more reluctant to refer to themselves and to the here-and-now. It is possible that this development reflects the professionalization process in Dutch journalism (Van Vree et al. 2003). Students of journalistic programs are explicitly trained to use an objective style.

4.3 Analysis of direct quotations

The final set of results relates to the parts of the texts that come from quoted news sources, which were presented as direct quotations. In a separate paper, we have discussed the development of quoted discourse over time in more detail.
(Vis, Sanders and Spooren 2012). There it is argued that the relative amount of quoted news increases over time (1950/1: 1153 words per 10,000 words; 2002: 2101/10,000 words). Table 5 gives a summary of the results.

Table 5: Overview of developments of the 11 indicators in direct quotations

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Rel. freq. 1950/1</th>
<th>Rel. freq. 2002</th>
<th>Effect</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pronoun 1p. sg.</td>
<td>78.18</td>
<td>287.91</td>
<td>YxN</td>
<td>↑↑↑↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YxS</td>
<td>↑↑↑↑</td>
</tr>
<tr>
<td>2. Pronoun 2p. sg.</td>
<td>37.31</td>
<td>100.78</td>
<td>YxN</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td>3. Pronoun 1p. pl.</td>
<td>71.77</td>
<td>130.57</td>
<td>YxNxS</td>
<td>↑↑↑↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td>↑↑↑↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>↑↑↑↑↑↑</td>
</tr>
<tr>
<td>4. Deictic element</td>
<td>39.83</td>
<td>55.05</td>
<td>Y</td>
<td>↑</td>
</tr>
<tr>
<td>5. Cognitive verb</td>
<td>104.16</td>
<td>156.86</td>
<td>Y</td>
<td>↑</td>
</tr>
<tr>
<td>6. Exclamation mark</td>
<td>10.64</td>
<td>10.53</td>
<td>YxN</td>
<td>↓↑↓↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YxS</td>
<td>↓↑</td>
</tr>
<tr>
<td>7. Question mark</td>
<td>21.40</td>
<td>27.86</td>
<td>YxN</td>
<td>↑↑↓↓</td>
</tr>
<tr>
<td>8. Intensifier</td>
<td>58.60</td>
<td>95.67</td>
<td>Y</td>
<td>↑</td>
</tr>
<tr>
<td>9. Modal adverb</td>
<td>31.13</td>
<td>45.31</td>
<td>YxS</td>
<td>↑↓↑</td>
</tr>
<tr>
<td>10. Modal particle</td>
<td>110.80</td>
<td>196.21</td>
<td>Y</td>
<td>↑</td>
</tr>
<tr>
<td>11. Modal verb</td>
<td>199.96</td>
<td>248.80</td>
<td>YxNxs</td>
<td>↑↑↑↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o</td>
<td>↑↑↑↑↑</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>↑↑↑↑↑</td>
</tr>
</tbody>
</table>

Y = year of publication, N = newspaper, S = subgenre; h = hard news, o = opinion news, b = background news; ↑ = increase, ↓ = decrease, ↔ = no effect of year, so no change.

Significant changes (p < .05) are underlined. Note that these relative frequencies generalize over newspapers and subgenres.

With respect to the research question concerning the development over time, the results of the analysis of subjective elements within quotations are relatively straightforward. For all elements, the year of publication has a significant effect on the frequency of occurrence, and this effect is always one of increase over time, except for the exclamation mark, for which the result is more ambiguous. For most elements, the increase is clear and uniform across all newspapers and subgenres; only for a few elements, the increase is not as consistent. Table 5 gives an overview of the developments for the 11 indicators.
Table 5 demonstrates that eight indicators show an increase for (almost) all newspapers and subgenres: first person singular pronouns, second person singular pronouns, first person plural pronouns (with the exception of Algemeen Dagblad and Trouw in opinion news), deictic elements, cognitive verbs, intensifiers, modal particles, and modal verbs (again with the exception of Trouw in opinion news). Only three elements display more variation in their changes over time: exclamation marks, question marks, and modal adverbs. The use of exclamation marks increases slightly over time in two of the newspapers, but in the others it decreases much more strongly (Trouw and De Telegraaf). In background news and hard news, the incidence of exclamation marks decreases, as well. The incidence of question marks increases in three newspapers, but decreases in two others (again, Trouw and De Telegraaf). Finally, the use of modal adverbs decreases in one of the subgenres, opinion news.

The conclusion is, then, that the expected informalization patterns does occur in direct quotations.

4.4 Summary of the results

In this paper, the developments of the proportions of subjectivity indicators in the text “overall” have been compared to the developments of subjectivity indicators in two subsets of the data, namely reporter’s text and direct quotations of news sources. This comparison shows that separating the different voices in an analysis of subjectivity is, in fact, crucial. Table 6 gives a summary of the developments over time of the eleven indicators in the three text types (“overall”, reporter’s text, quoted text), disregarding differences between subgenres and newspapers.

With respect to the informalization hypothesis, Table 6 shows that there is a clear difference between the developments over time in the voice of the reporter on the one hand, and the developments in the voice of quoted news sources on the other: the subjectivity of the latter, as expressed in direct quotations, has increased across all indicators, with the only exception of exclamation marks. The subjectivity of the reporter, however, has increased in only five of the eleven indicators (second person singular pronouns, cognitive verbs, question marks, intensifiers, and modal particles). Moreover, it has decreased in four others (first person plural pronouns, exclamation marks, modal adverbs, and modal verbs). In one indicator, there is no change at all (first person singular pronouns). The results for the deictic elements are ambiguous, because the newspapers and/or subgenres do not concur in the direction of the change.
5 Making statistics concrete: The case of modal adverbs

The following trend can be seen in our data: when looking at entire texts, there is no clear indication of informalization; some features show an increase over time, whereas others stay the same. In the reporter’s text, the trend is even more mixed, with some features increasing and other features decreasing. In the direct quotations, we see an overall increase of the subjectivity features.

A case in point are the modal adverbs. In the texts “overall,” the results of the analyses are mixed: some newspapers and subgenres show an increase, some a decrease. In the reporter’s text, we see a decrease in most newspapers. In the direct quotations, the use of modal adverbs shows an increase in two out of three subgenres. Looking at a particular modal adverb (natuurlijk, meaning ‘of course’), we see the same pattern: a small increase overall, but a very strong increase in the direct quotations.

This means that examples like (1), where natuurlijk occurs in the reporter’s text, are used slightly more often in 1950/1 than in 2002, whereas examples like (2) (natuurlijk in quoted text) are much more common in 2002 than in 1950/1.

| Table 6: Summary of the developments of the 11 indicators in the text overall, in the reporter’s text, and in direct quotations |
|----------------------|------------------|------------------|------------------|
| Indicator             | Overall          | In reporter’s text | In direct quotations |
| 1. First person singular pronouns | ↑ | ↔ | ↑ |
| 2. Second person singular pronouns | ↑ | ↑ | ↑ |
| 3. First person plural pronouns | ~ | ↓ | ↑ |
| 4. Deictic elements | ~ | ~ | ↑ |
| 5. Cognitive verbs | ↑ | ~ | ↑ |
| 6. Exclamation marks | ↔ | ↓ | ~ |
| 7. Question marks | ~↑ | ~↑ | ~↑ |
| 8. Intensifiers | ↑ | ↑ | ↑ |
| 9. Modal adverbs | ~ | ↓ | ↑ |
| 10. Modal particles | ↑ | ↑ | ↑ |
| 11. Modal verbs | ~ | ↓ | ↑ |

↑ = increase, ↓ = decrease, ↔ = no effect of year, so no change, ~ = ambiguous result, ~↑ = ambiguous result but tendency to increase.
Informalization in Dutch journalistic subgenres over time

6 Conclusion

Informalization in the text “overall”

The results show that analyzing the expression of subjectivity in journalistic discourse without accounting for the distinction between reporter’s subjectivity and source subjectivity can lead to misleading results. Indeed, from the analysis of the indicators in the text “overall”, the conclusion would be that there are some indicators that provide evidence in favour of informalization, whereas for other indicators only some newspapers and subgenres show informalization and others do not. Table 6 shows that for the “overall” text, six indicators demonstrate an increase (first and second person singular pronouns, cognitive verbs,
question marks, intensifiers, and modal particles). These are all indicators for
which the results show an increase in both the reporter’s text and the direct quo-
tations. However, for four other indicators in the text “overall,” the results are
inconclusive: there is an effect of the year of publication, but the results for
the five newspapers and three subgenres are so divergent that they cannot be
summarized as a development in either direction, or they cancel each other out
statistically (first person plural pronouns, deictic elements, modal adverbs, and
modal verbs). Table 6 demonstrates that the findings for three of these inconclu-
sive indicators are in fact based on an increase in the direct quotations and a
decrease in the reporter’s words. This means that the conclusion based on the
results for the indicators “overall” fails to recognize what is truly happening:
there are two distinct developments; one in the direct quotations, which already
were more subjective than the words of the reporter in 1950/1 and have con-
tinued to become even more subjective; and a different development in reporter’s
words, which have informalized only to a degree. These separate developments
in the direct quotations and the reporter’s words are obscured in the overall
analysis, because they cancel each other out.

**Informalization in reporter’s text and direct quotations**

Our results show that the informalization hypothesis does not hold equally for
the two subsets of our data. The hypothesis states that in general, lexicogram-
matical indicators of subjectivity show an increase over time.

In the direct quotations, all these expected increases are observed, although
for the question marks and the modal adverbs, this rise in frequency does not
take place in all newspapers (for question marks) and subgenres (for modal
adverbs). In accordance with the previous findings, exclamation marks have
not increased over time and have even decreased in the newspapers that used
them most in 1950/1 (*De Telegraaf* and *Trouw*). This decrease in the use of exclama-
tion marks might be caused by increased prescriptivism, that may have come
up with the professionalization of journalism and the start of professional train-
ing for journalists between 1950 and 2002 (see Vis 2011, Chapter 1, for a discus-
sion of these historical developments in journalism). The use of exclamation
marks in journalistic prose may be deemed too informal to be acceptable in
written text now, whereas this was not the case in 1950/1. Indeed, the present-
day style guide for *de Volkskrant* prescribes to use exclamation marks sparingly
(Van Gessel et al. 1992). Additionally, it states that exclamation marks may be
used for exclamations in quotations, but that as a rule the text should speak
for itself.
The modal verbs, which were not expected to show a rise in frequency in light of the previous findings, did, however, predominantly increase in frequency in the direct quotations, although in hard news in newspapers, this change is not so clear. The difference between this finding and the findings of previous studies with respect to the diachronic change in use of modal verbs might be related to the difference between English and Dutch in the options of expressing modality: in English, there are two candidates for expressing modality in verbs, core modal verbs and semi-modal verbs, whereas in Dutch only the first option is available. As discussed in section 2, studies on changes in English found a decrease in modal verbs and an increase in semi-modal verbs. As the latter option is not available in Dutch, this might (partly) explain the increase in the use of core modal verbs. Another explanation might be found in the fact that the previous diachronic studies, on which our expectations were based, did not systematically account for the distinction between the reporter’s words and the direct quotations.

With respect to the reporter’s words, the expected increases were only observed in second person singular pronouns, cognitive verbs, question marks, intensifiers, and modal particles. First person pronouns (both singular and plural), deictic elements, and modal adverbs did not meet the expectation. In light of the previous findings, exclamation marks and modal verbs were not expected to show an increase, and, in fact, decreased over time. In sum, some indicators in the reporter’s words provided evidence in favour of the informalization hypothesis, whereas others display a significant decrease, suggesting that the reporter’s words have informalized only to a degree. The increases in second person singular pronouns and question marks suggest that newspapers have become more inclined to express interactive indicators over time. In contrast, the decreasing occurrence of deictic elements and first person plural pronouns, and the stable frequency of first person singular pronouns might mean that newspapers have become more reluctant over time to explicitly refer to the here-and-now and to themselves as institutions or to the community or nation that they are part of. Similar to the results for the direct quotations, the differences between the results of this study for the reporter’s text and the findings of previous studies might be caused by the fact that the earlier findings were based on analyses that did not systematically take the distinction between reporter’s text and direct quotations into account.

7 Discussion

To what extent is informalization dependent on subgenre and on different newspapers? It is noteworthy that the newspapers and subgenres are involved in a
two- or three-way interaction with the year of publication in a markedly high
number of cases. Newspaper is involved in an interaction with year of publica-
tion in 25 out of 33 times (11 indicators times 3 for text “overall”, reporter’s text,
and direct quotations) and subgenre in 22 out of 33 cases. This means that for
the majority of indicators, there are significant differences between the newspa-
pers and/or between the subgenres. This is not necessarily a difference with
respect to the direction of the development; in fact, it is striking how often the
newspapers and subgenres concur in the direction of the change, both in the
reporter’s text and the direct quotations. Rather, they vary in the extent to which
they change over time.

In some cases, the newspapers and subgenres are not involved in an inter-
action with year; for example, in first person singular pronouns in the text
“overall” and in the reporter’s words, there are no significant differences
between newspapers or between subgenres with respect to their develop-
ment over time. However, even if there is no interaction with year, the newspapers
and subgenres usually still have an effect, either in a two-way interaction with
each other or separately as a main effect. This means that in those cases, there
is a difference between newspapers and/or subgenres, but this difference does
not change over time.

For only four indicators, the newspapers and/or subgenres show no differ-
ence at all. Interestingly, for each of these four indicators, this concerns the
incidences in direct quotations: there is no effect of newspaper on deictic ele-
ments and cognitive verbs in direct quotations, and no effect of subgenre on
deictic elements, question marks, and intensifiers in direct quotations.

Although this shows that there are significant differences between newspa-
pers and subgenres in their behaviour with respect to the use of subjective
elements, it is difficult to make general statements about the character of these
differences, since there are so many different indicators and interactions in-
volved in the analyses. There is no newspaper that shows a clear consistent
development across all indicators and types of voice (reporter or direct quota-
tion). Note that there is also no clear difference between tabloids (De Telegraaf,
Algemeen Dagblad) versus quality papers (NRC Handelsblad, de Volkskrant,
Trouw). As mentioned before, trends toward tabloidization are less strong in
the Netherlands compared to, for example, the United Kingdom. Similarly,
although subgenre was involved in various interactions with year of publication,
the pattern was too irregular to reach firm conclusions about the role of sub-
genre in the informalization process.

In conclusion, this study provides evidence that it is crucial to distinguish
between reporter’s subjectivity and subjectivity of news sources quoted in the
text. With respect to the informalization hypothesis, it seems that the reporter’s
text has informalized only to a degree: some indicators show evidence in favour
of the informalization hypothesis, whereas others appear to be evidence against
it. As suggested above, the results for the reporter’s text might mean that newspa-
papers tend to express more interactive indicators over time, whereas they
became more reluctant to explicitly refer to themselves and to the here-and-now.
Direct quotations, however, show strong evidence of informalization. This devel-
opment seems to be relatively independent of the subgenre that is investigated.

This study contributes to our understanding of developments in journalistic
language use. News reports have often been treated as a single register. We have
shown how such a limitation can be overcome using corpus analysis. By using a
sophisticated type of statistics, we were able to paint a detailed picture of a com-
plex diachronic development using a large set of linguistic features. We also
believe to have added to the existing body of knowledge which mainly focuses
on English language use, by describing patterns in Dutch journalistic language
use, occurring in the major national newspapers of the Netherlands.

References

Bekker, Birgit. 2006. De feiten verdraaid. Over tekstvolgorde, talige markering en sprekerbe-
trokkenheid [Twisted facts. On discourse order, linguistic marking and speaker involve-
ment], PhD dissertation/ Tilburg: University of Tilburg.
Press.
Biber, Douglas. 2004. Historical patterns for the grammatical marking of stance: A cross-
three genres. Language 65. 487–517.
Biber, Douglas & Edward Finegan. 2001. Diachronic relations among speech-based and written
registers in English. In Susan Conrad & Douglas Biber (eds.), Variation in English: Multi-
1994. The design and analysis of the ARCHER Corpus: A progress report. In Merja Kytö,
Matti Rissanen & Susan Wright (eds.), Corpora across the centuries (Proceedings of the
Deuze, Mark. 2005. Popular journalism and professional ideology: Tabloid reporters and editors
Fairclough, Norman. 1994. Conversationalization of public discourse and the authority of the
consumer. In Russel Keat, Nigel Whiteley & Nicholas Abercrombie (eds.), The Authority of
Appendix: Subjectivity lexicon

For the analysis, a lexicon of subjectivity elements was created, based on insights from the literature.

1. First and second person pronouns [a.o. Scheibman 2002]
   All lexical items tagged as first or second person pronoun, either personal or possessive, were coded as subjective. Additional codes were set to distinguish between person (first or second person) and number (singular or plural).

   33 frequent modal adverbs were counted per text. Some of these items can only be used as adverbs, e.g. allicht (‘of course’), helaas (‘unfortunately’) and misschien (‘maybe’). However, some of the items are in fact adjectives that can be used either adjectively or adverbially, such as natuurlijk (‘natural’/‘of course’) and mogelijk (‘possible’/‘possibly’). The annotation scheme only identified uninflected forms of these items as modal adverb. The instances matching the identification patterns were tagged as subjective category ‘modal’ and subcategory ‘adverb’.

   Modal particles are uninflected words that reflect the attitude of the speaker. Many of the words functioning as a modal particles have other functions as
well. In the set used in this study, only those modal particles were included
that could automatically be recognized as such on the basis of its part of
speech tags in combination with (the part of speech tags of) the immediate
context. For example, maar ('but') as an adverb was annotated as modal
particle, whereas instances of maar as a conjunctive were not; the adverb
al ('already') was only annotated as a modal particle when it was not
followed by an adjective, adverb, numeral/quantifier or pronoun; and the
lexical item pas ('only') tagged as an adverb was annotated as a modal par-
ticle except when it was preceded by van ('of') or te ('to'), which happened
rather frequently in the 1950s corpus.

4. Intensifiers [Haeseryn et al. 1997]
The lexicon contains three sets of intensifiers, grouped with respect to the
conditions necessary for the identification of the intensifier. The items in
the first set were words like bijna ('almost'), echt ('really'), and nogal
('rather'). The second set consists of prefixes that all intensify the sense of
the words they are added to. For instance, sterk means 'strong', but oersterk
means 'exceedingly strong'. Finally, the third set consists of adverbs that are
only used as intensifiers when the following (or preceding, in the last two
cases) item is an adverb, adjective, or pronoun; it contains elements like
amper ('hardly'), redelijk ('reasonably'), and verschrikkelijk ('terribly').

5. Modal verbs [Haeseryn et al. 1997]
The subjectivity lexicon adopted the list of modal verbs presented by Haeseryn
et al. (1997):
behoeven, blijken, dunken, heten, hoeven, lijken, kunnen, moeten, mogen,
schijnen, toeschijnen, voorkomen, willen, zullen
All forms of these verbs (both finite and other verb forms) were annotated as
modal verbs.

A list of frequently used cognitive verbs was compiled. Following Biber
(2004: 111), a liberal approach was adopted in defining the set of cognitive
verbs, including "any form that can be interpreted as expressing author/speaker attitude and assessment of likelihood". The resulting list contains
the following 52 verbs, including non-factive verbs commenting on how
information is communicated, e.g. antwoorden ('answer'), attitudinal verbs,
e.g. hopen ('hope'), factive verbs, e.g. weten ('know') and likelihood verbs,
e.g. betwijfelen ('doubt').
The annotation scripts identified the cognitive verbs on the basis of the
lemma information in the corpus. This ensured that all different forms of
the verbs were annotated as subjective. Additionally, identifying the cognitive verbs by their lemma has the added benefit that versions of the verbs
that are now considered archaic, such as zeide ('said'), are annotated as cognitive verbs.

7. Exclamations
Both exclamations and imperatives are usually marked by the same punctuation mark, the exclamation mark. The scheme for subjectivity annotation used this symbol to identify both features and added a subjectivity code to the XML tags for the exclamation marks.

8. Questions
The identifying patterns identified questions by the punctuation mark that typically marks them, the question mark. The annotation added a subjectivity code to each question mark in the corpus.

The lexicon contains a limited set of deictic elements, such as hier ('here'), nu ('now'), gisteren ('yesterday'), vandaag ('today'), morgen ('today'). These elements were identified through their lemma information in the corpus and were annotated with a subjectivity code.