Colloquialization in journalistic writing
The case of inserts with a focus on well

Christoph Rühlemann and Martin Hilpert
Philipps-University Marburg / University of Neuchâtel

Recent analyses of written text types have discovered significant frequency increases of colloquial or conversational elements, such as contractions, personal pronouns, questions or the progressive. This trend is often referred to as colloquialization. This paper presents a new perspective on colloquialization, with a special focus on the discourse marker well. The paper is divided into two parts. In the first part, we present new evidence of colloquialization on the basis of the TIME Magazine Corpus (Davies 2007), which allows analyses of diachronic change in recent written American English. The focus of our analysis is on highly frequent “inserts” (Biber et al. 1999: 56), which are elements such as discourse markers (e.g., well and oh), backchannels (yeah, uh-huh, etc.), and hesitators (uh and um, etc.). We conclude that inserts significantly increase diachronically in TIME. In the second part of the paper, we focus on the element well in its function as a discourse marker. Through a combination of quantitative and qualitative analytical steps, we analyze its diachronic development in terms of its structural contexts and its pragmatic functions, fleshing out how the process of colloquialization has affected its usage in recent written American English. We argue that the integration of corpus linguistic and pragmatic methods in this case study represents a new step towards the field of corpus pragmatics, that is, “the rapprochement between corpus linguistics and pragmatics and an integration of their key methodologies” (Rühlemann and Aijmer 2014: 23).

Keywords: colloquialization, inserts, discourse markers, well, Variability-based Neighbor Clustering (VNC), corpus pragmatics

1. Introduction

Recent corpus research on short-term diachronic change has been very fruitful. A particularly notable project in this context is Leech et al.’s (2009) work – a large-scale study of grammatical change in the twentieth-century based on the Brown
family, BLOB and BE06. A broad number of areas of change were identified. These include, to name only a few, verbal and negative contractions (e.g., Leech 2003), personal pronouns (e.g., Baker 2009), questions, the progressive and zero relative clauses (e.g., Leech et al. 2009). Also, the findings have been discussed on the basis of a number of explanatory hypotheses. For example, the changes just mentioned have been explained in terms of colloquialization, which is the broad topic of this paper.

Colloquialization has been characterized as a shift of written norms towards spoken norms, so that written language becomes more speech-like (e.g., Leech 2003). Alternatively, colloquialization has been referred to as conversationalization (Fairclough 1995). Interestingly, colloquialization is by no means a new trend but has been observed from the seventeenth century onwards (Leech and Smith 2009: 175).

Despite the fruitfulness of research on colloquialization, work on this topic has been hampered by methodological problems. Many studies have followed a trajectory from previous research to corpus exploration, starting out with a change that has been observed in earlier research, and then moving on to exploration, asking “Does this change occur in my data/corpus too?”. This approach is, for example, the one adopted in Leech et al’s (2009) major study of short-term grammatical change (Nicholas Smith, p.c.). While the approach is not intrinsically problematic, it does not facilitate the discovery of new phenomena: if earlier studies have overlooked a process of change, that process will be overlooked in follow-up studies as well.

We propose, therefore, an alternative methodological route, namely from conversation to exploration. To investigate colloquialization we start out from conversational corpus data, asking “What is typically conversational/colloquial?”. A useful way to answer this question is by doing keyness analysis. Keyness provides two indications: aboutness indication (the topics prevalent in a text) and stylistic indication (how the text is realized) (see Scott and Tribble 2006: Chapter 2). In large and generic corpora, a comparison of conversation against written genres will identify items that are stylistically key in conversation as compared to the written language. With such a catalogue of conversational keywords, researchers can turn to their diachronic corpus and explore the extent to which the key items decrease, increase or remain steady over time.

Further, we wish to analyze the data more deeply than is usually the case. Diachronic studies often show a primary interest in reporting frequencies of a given feature across different time spans and in determining whether the feature exhibits an upward or downward trend (see Hilpert and Gries 2009). While this is a legitimate question (although not without problems if the time spans are wildly discontinuous), it is by no means the only or most rewarding question that can be asked about diachronic data. It is at least equally important to establish whether
there are systematic qualitative differences between earlier and later uses of a linguistic element, which is a question that can be asked regardless of there being a frequency trend or not. After all, a given element can change not only in terms of frequency ups and downs but also in terms of its use and functions in context. In this paper, we will illustrate this type of approach with an investigation of a single element, namely *well*. We will examine its (partly) new functions as a discourse marker, in a specific genre (journalistic writing) thereby taking a semasiological (form-to-function) approach (see, for example, Mosegaard Hansen and Rossari 2005: 179; Brinton 2010: 296).

Finally, our approach also showcases “corpus pragmatics” in action, a “new burgeoning discipline facilitated by the *rapprochement* between corpus linguistics and pragmatics and an integration of their key methodologies” (Rühlemann and Aijmer 2014: 23; see also Taavitsainen et al. 2014). Corpus linguistics typically scans texts vertically (Tognini Bonelli 2010) looking for association patterns within narrow co-textual windows that can be described in quantitative terms, while pragmatic analysis typically processes texts horizontally, thus investigating, in much larger co-textual windows and more varied dimensions of context, functions of language that can be described in qualitative terms. The integrated methodology of corpus pragmatics is shown schematically in Figure 1.

![Figure 1. Integrated-reading methodology in corpus pragmatics (CP) combining the vertical-reading methodology of corpus linguistics (CL) with the horizontal-reading methodology of pragmatics (P)](image)

Our research aims for a tight integration of corpus-linguistic and pragmatic methodologies. We will employ the corpus-linguistic vertical method of analyzing *TIME* magazine texts, for example, by extracting frequency data for the discourse
The marker *well* and by computing its collocates across diachronic stages. The pragmatic method of carefully reading through and interpreting large portions of individual texts is underlying our investigation of how *well* is used in context and what functions it is used for in the interaction of writer and reader. Importantly, the two key methods are not just used side-by-side. Rather, the insights gained from the quantitative corpus-linguistic analysis are used as guidelines towards the most rewarding avenues for the pragmatic qualitative analyses. The study is, hence, a paradigmatic corpus-pragmatic study. Its integrated methodology, deployed for an analysis of the discourse marker *well*, provides insights into how the marker’s use is changing over time. For example, *well* can be shown to enter into new syntactic contexts and develop functions not attested either in conversation, its “home” genre, or any written genres other than news magazine writing.

The study is structured as follows. Section 2 outlines the methods and the corpus data that we used. Section 3 reports the results of our general investigation of inserts and of the case study that focuses on *well*. Section 4 offers a general discussion and a theoretical contextualization of the empirical results. Finally, Section 5 summarizes the main findings and presents our conclusions.

2. Methods and data

Following the methodological compass depicted above, we used BNCweb (Hoffmann et al. 2008) to calculate keywords in the demographically sampled subcorpus of the BNC (BNCdemog) against the whole of the written component. The top twenty key words, ordered by their Log Likelihood values, are given in Table 1. Among them, not surprisingly, we find verbal and negative contractions and personal pronouns, which have, as noted, already been identified as increasing in written discourse.

At the same time, the list prominently features items that have not yet been examined in studies of short-term diachronic change; most notably, it contains what Biber et al. (1999: 56) call “inserts” – that is, items that frequently occur in speech and that can be inserted into discourse with a high degree of syntactic flexibility. Inserts are an extremely broad and heterogeneous word class with a large number of sub-categories. For example, *yeah* (ranked first), *no* (seventh), *mm* (eighth) and *yes* (nineteenth) would be classed as “responses”, *er* and *erm* as “hesitators”, while *oh, well, cos* and *know* (with preceding *you*) would be categorized as “discourse markers”.

In this study, we investigated three types of inserts, namely responses, discourse markers and hesitators. The responses include *yeah, yes* and *uh-huh*, the discourse markers *oh, coz, you know* and *well*, and the hesitators *uh, um* and *er*. The
responses *yeah*, *yes* and *uh-huh* have in common that they serve a double function as second part of a question – answer adjacency pair and as a minimal backchannel (“continuer”). Seen as a backchannel, *yeah* is the most frequent backchannel in AmE (Tottie 1991: 264). As regards *uh-huh*, other forms (e.g., *unhuh*, *unhhunh*) were highly infrequent in the corpus (see below). The discourse markers *oh*, *coz*, *you know* and *well* fulfill a broad range of functions in discourse. Aijmer (2013) points out two major functions shared by discourse markers, namely reflexivity (a speaker-centred function) and contextualization (a hearer-based function). As regards hesitators, *uh*, *um* and *er* were the only hesitation forms attested in some numbers in TIME. Generally, hesitators are much more than just indices of hesitation, since they serve a broad range of turntaking functions (e.g., Rühlemann 2007; Kjellmer 2009; Tottie 2014).

Table 1. Top twenty key words in demographically sampled spoken subcorpus (C) against the whole of the written component (W) of the BNC

<table>
<thead>
<tr>
<th>N</th>
<th>Word</th>
<th>Tag</th>
<th>Freq. in C</th>
<th>Freq. in W</th>
<th>Log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yeah</td>
<td>ITJ</td>
<td>58,706</td>
<td>1,386</td>
<td>344216.19</td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td>PNP</td>
<td>167,426</td>
<td>547,004</td>
<td>294395.23</td>
</tr>
<tr>
<td>3</td>
<td>you</td>
<td>PNP</td>
<td>134,910</td>
<td>398,899</td>
<td>256476.85</td>
</tr>
<tr>
<td>4</td>
<td>’s</td>
<td>VBZ</td>
<td>89,938</td>
<td>144,957</td>
<td>248882.05</td>
</tr>
<tr>
<td>5</td>
<td>oh</td>
<td>ITJ</td>
<td>41,555</td>
<td>14,456</td>
<td>190335.54</td>
</tr>
<tr>
<td>6</td>
<td>n’t</td>
<td>XX0</td>
<td>77,480</td>
<td>189,912</td>
<td>168179.21</td>
</tr>
<tr>
<td>7</td>
<td>no</td>
<td>ITJ</td>
<td>32,988</td>
<td>20,352</td>
<td>131822.81</td>
</tr>
<tr>
<td>8</td>
<td>mm</td>
<td>ITJ</td>
<td>21,888</td>
<td>227</td>
<td>130696.3</td>
</tr>
<tr>
<td>9</td>
<td>er</td>
<td>UNC</td>
<td>21,345</td>
<td>952</td>
<td>122125.87</td>
</tr>
<tr>
<td>10</td>
<td>that</td>
<td>DT0</td>
<td>63,324</td>
<td>186,217</td>
<td>120880.07</td>
</tr>
<tr>
<td>11</td>
<td>it</td>
<td>PNP</td>
<td>127,977</td>
<td>799,065</td>
<td>112700.28</td>
</tr>
<tr>
<td>12</td>
<td>erm</td>
<td>UNC</td>
<td>16,605</td>
<td>187</td>
<td>99017.99</td>
</tr>
<tr>
<td>13</td>
<td>do</td>
<td>VDB</td>
<td>42,266</td>
<td>116,247</td>
<td>84742.69</td>
</tr>
<tr>
<td>14</td>
<td>got</td>
<td>VVN</td>
<td>22,545</td>
<td>18,471</td>
<td>82553.54</td>
</tr>
<tr>
<td>15</td>
<td>well</td>
<td>AV0</td>
<td>34,608</td>
<td>75,294</td>
<td>81044.9</td>
</tr>
<tr>
<td>16</td>
<td>know</td>
<td>VVB</td>
<td>21,547</td>
<td>27,658</td>
<td>66381.18</td>
</tr>
<tr>
<td>17</td>
<td>cos</td>
<td>CJS</td>
<td>11,374</td>
<td>353</td>
<td>66085.19</td>
</tr>
<tr>
<td>18</td>
<td>’ve</td>
<td>VHB</td>
<td>22,738</td>
<td>36,628</td>
<td>62939.22</td>
</tr>
<tr>
<td>19</td>
<td>yes</td>
<td>ITJ</td>
<td>17,866</td>
<td>18,813</td>
<td>59743.28</td>
</tr>
<tr>
<td>20</td>
<td>na</td>
<td>TO0</td>
<td>9,966</td>
<td>728</td>
<td>55403.89</td>
</tr>
</tbody>
</table>
The corpus underlying the subsequent analyses is the TIME Corpus (Davies 2007). This corpus contains all texts of TIME magazine from 1923 to 2006. In terms of size, the collection amounts to roughly 110 million words. All of the texts are time-stamped to provide the exact dates of publication. The analyses in this paper will use the periodization into decades that is built into the online corpus interface and that ranges from the 1920s to the first decade of the twenty-first century.

For data retrieval, automatic searches were viable only for some items, such as yeah, which is orthographically stable and virtually always acts as a “reception marker” serving “to acknowledge the receipt of information that is new to the discourse but consistent with currently active information” (Jucker and Smith 1998: 179). Restricted-recall searches were necessary where the item in question is polysemous, as for example, well. To capture, as precisely as possible, all uses of well as a discourse marker, rather than an adverb or a noun etc., we constructed a search string1 that included a series of punctuation signs immediately before and after well (building on the assumption that if writers intend to use well as a discourse marker they will offset it from the neighboring co-text by typographical means). Finally, extensive manual editing of results was necessary for orthographically inconsistent forms such as hesitators2 and the different spellings of COZ (namely, cos, coz and cause).

The subsequent analysis addresses three questions which require three different methods of statistical analysis. Our first question, “Is there a frequency trend?”, can be answered fairly easily by feeding normalized frequencies (tokens per million words) and decades into correlation tests. These yield correlation values such as Kendall’s τ, which range from 0 (meaning no correlation at all) to 1 (indicating a perfect correlation).

The second question, “Are there developmental stages in the data?”, is much less easily answered, but it is one which the temporal continuity of the TIME Corpus allows us to address. One method to detect stages in temporally ordered data is Variability-based Neighbor Clustering (VNC), first proposed by Gries and Stoll (2009) and taken further in a number of other works (e.g., Hilpert and Gries 2009; Hilpert 2013). VNC is a form of Hierarchical Clustering Analysis (HCA). In very general terms, HCAs are useful whenever a researcher inspects a large

---

1. ((),|,|?|!|… well ().|,|:|?|!|…)

2. A search for er in the TIME Corpus returns 488 hits. Inspection of concordance lines reveals, however, only ninety-five hits represent a genuine hesitator rather than, for example, the feminine possessive pronoun her (without “h”) and even the German masculine third-person singular pronoun as in “Hoch soil er leben, Hock soil er leben, Hoch soil er leben. Dreimal Hoch!” (TIME 1929).
dataset with many observations and wants to find out how these observations can be joined together in a set of groups. For instance, a biologist might have a collection of beetle species that are genetically related. On the basis of measuring similarities in the respective genomes of the beetles, cluster analysis can be applied in order to determine which of the species are particularly closely related and whether there are identifiable groups of closely related species (or so-called “clusters”). To arrive at these clusters, the analysis first identifies the two beetle species that are most similar to each another. After that, it determines which third species is most similar to the first two, and so on, until all species are accounted for. We apply a similar logic to the analysis of linguistic datasets: we have a number of observations and successively group together those observations that are most similar to one another. A specific characteristic of the VNC technique is that it is applied to temporally ordered linguistic data in such a way that only temporally adjacent data points are allowed to form clusters (Hilpert and Gries 2009: 390). Like other uses of clustering, VNC is an exploratory method, serving not to test an hypothesis but to generate one. The hypothesis generated by VNC in our study concerns the periodization of a diachronic development into distinct stages, based on frequency values. That is, based on the frequency development of a linguistic form, the VNC algorithm detects stages in that development, states how many such stages there are, and determines which time spans are to be taken together as a single stage. In this study, we apply VNC to the diachronic frequencies of the discourse marker well. This gives us a sequence of diachronic stages for its recent development, thus answering our second research question.

Our third and most important question is “How has the use of well changed qualitatively over time?” To test what exactly is changing across the developmental stages of well, we carried out a quantitative analysis of its collocates. The choice of this method builds on the assumption that if and when an item changes, this change will not only entail changes in frequency but, more importantly, changes in the way the item is used. These changes in usage manifest themselves in changes in the company the item keeps – in its collocational behavior. To establish whether the observed collocational change can be considered to be statistically significant, we used the Fisher Yates exact test. Specifically, we compared the collocates of well with regard to their frequencies across two time periods, checking for each collocate whether it shows a frequency increase, a decrease or no change at all. The test results are interpreted as indications of language change: if the collocates of well differ significantly in terms of their frequencies from one stage to the next, this is taken as an indication that change has been occurring. To discover exactly how an item has been changing requires concordance analyses where the interplay between node and collocate is investigated in the larger context.
3. Results

In this section, we report the results of our investigations into colloquialization. The section is divided into three main sections. Section 3.1 takes up the research question, “Is there a frequency trend?” detailing the findings on all inserts selected for analysis. Section 3.2 is guided by the question, “Are there developmental stages in the data?” Given the complexity of the methods used to address the question, we narrow the focus to a single marker – namely, the discourse marker well. In Section 3.3, finally, we present the results of close readings and detailed codings of concordances of sentence-medial well.

3.1 Is there a trend?

Let us start with the responses yeah, yes and uh-huh. In Figure 2, the left panel plots frequencies per million words against the decades. We see a modestly strong and significant correlation for yes, a strong and significant one for yeah and no significant correlation for uh-huh. However, uh-huh shares with yes and yeah a noticeable increase over the last two decades. This is also revealed in the right panel, which shows the percentage change on the previous decade: all three responses have their highest increases in the 1990s.

![Figure 2. Frequencies of responses per million words (left panel) and change rates on previous decades (right panel)](image-url)
A similar picture emerges for the discourse markers *oh*, *COZ*, *you know* and *well*, which are shown in Figure 3. The left panel depicts modestly strong correlations for *COZ* and *you know*, and a strong correlation for *well*, but no significant correlation for *oh*. However, even *oh* undergoes a dramatic increase over the last two decades, as do *you know* and *well*. Only *COZ* has its greatest change rate in earlier decades.

![Discourse markers and Change rate](image)

**Figure 3.** Frequencies per million words of discourse markers (left panel) and change rates on previous decades (right panel)

The picture emerging for the three hesitators *uh*, *um* and *er*, shown in Figure 4, looks familiar too: all three forms are modestly strongly correlated with the decades. As regards the rate of change on the earlier decade, only *um* sees its highest relative growth in the 1990s; *uh* and *er* change most dramatically in earlier decades.

To summarize what we have so far: eight of ten features investigated are modestly strongly to strongly correlated with the decades, seven of ten features see their greatest relative increase in the last two decades, specifically in the 1990s. So, overall, we observe positive trends: inserts seem to be on the rise in TIME magazine, accelerating in the 1990s and 2000s. Having established that there are trends, we can further inquire whether an item simply continuously increases over time or whether the item develops across stages – in leaps, as it were – and whether these leaps are correlated with changes in the way the item behaves in discourse. As noted before, this question can be addressed by means of VNC and subsequent collocation and concordance analyses. In order to show how this combination of methods works in practice, we now narrow the focus to a single item, the discourse marker *well*. This marker has been extensively investigated both diachronically (e.g., Finell 1989; Jucker 1997; Crystal 2005; Brinton 2010; Defour 2008;
Culpeper and Kytö (2010) and synchronically (e.g., Svartvik 1980; Levinson 1983; Schiffrin 1985; Fraser 1990; Jucker 1993; Schegloff and Lerner 2009; Aijmer 2013; Rühlemann 2013).

In diachronic research, a number of trajectories have been sketched out for the development of *well* (see Brinton 2010 for a concise overview). Jucker (1997: 95) notes that “the earliest Old English uses of *well* are derived from Indo-European *wel-* ‘to will, wish’”. Finell (1989) traces the origin of the discourse marker *well* back to its use in Old English as a predicative adjective, as in *this/that is (very) well*, where it expresses “a wish to be in good standing with a person” (Finell 1989: 654). Jucker locates the origin of discourse marker *well* in the Old English form *wel la* or *wella*, which performs the function of an attention-getting device at important structural junctures in the context of direct speech (Jucker 1997: 97; see also Crystal 2005: 190). In Middle English, the discourse marker *well* is again primarily used in the context of direct speech, “followed immediately by a parenthesis consisting of a verb of speaking and an indication of the speaker” (Jucker 1997: 99). In Early Modern English, the discourse marker *well* expands into contexts beyond direct speech and “there are the first instances of *well* occurring in fictional texts without any close connection to the spoken language” (Jucker 1997: 105).

In synchronic research, a great many distinct uses and functions have been noted for *well* in a broad variety of contexts “ranging from dispreferred response signal to face-threatening minimiser to qualifier or frame” (Brinton 2010: 297).
An initial approximation to well’s function as a “dispreferred response signal” is Svartvik’s characterization of well as an introducer to “a part of the discourse that has something in common with what went before but also differs from it to some degree” (Svartvik 1980). Subsequent analyses have specified this function in more explicit terms. For example, Levinson (1983: 334) observes that well “standardly prefaces and marks dispreferreds”; in this context, well may function as a “warning particle” (Levinson 2013: 108). In a similar vein, Fraser views well as a marker of “dissonance” (Fraser 1990: 387) and Schiffrin argues that well introduces “a temporary suspension […] for immediate coherence of a response” (Schiffrin 1985: 648; see also Schiffrin 1987: 323). Jucker refers to this function as “insufficiency” (Jucker 1993) and, alternatively, “qualification” (Jucker 1997). Schegloff and Lerner (2009: 91) interpret well-prefaced responses “as general alerts that indicate nonstraightforwardness in responding”, with “nonstraightforwardness” closely corresponding to what Finell (1989: 654) earlier referred to as “an element of indirectness”. Nonstraightforwardness can also be applied to Jucker’s (1993: 450) relevance-theoretic interpretation of well as a signal that “the most immediately accessible context is not the most relevant one for the interpretation of the impending utterance”. Beside well’s function as a dispreferred response signal, a number of scholars have observed its function in the context of face-threatening acts. Jucker, for example, notes that “[a]s a face-threat mitigator, well indicates a problem on the interpersonal level. Either the face of the speaker or the face of the hearer is threatened” (Jucker 1993: 444). Aijmer (2013: 42) observes that well can not only serve to soften a potential face-threat but also to “reinforce it” (e.g., in an argument), especially in contexts where it co-occurs with epistemic adverbs such as certainly, surely, actually, etc. Intriguingly, the face-threat minimizing function of well can be related to the above-mentioned uses of well as a predicative adjective in Old English (Finell 1989: 654). The third commonly acknowledged function of discourse marker well is what, for example, Jucker (1993) calls “frame”. As a frame “well is used to separate discourse units” (Jucker 1993: 446). It performs this separating function, for instance, when speakers seek a topic change and when they use constructed dialogue (e.g., Schiffrin 1985: 658; Svartvik 1980: 170; Jucker 1993, 1997; Aijmer 2013; Rühlemann 2013). In the context of constructed dialogue, well serves as a quote-signal flagging the inception of direct or free direct speech, thus demarcating the boundary between the speaker’s own words and the quoted material (Rühlemann 2013) and thereby “creat[ing] coherence ‘locally’ in the speaker’s turn” (Aijmer 2013: 32). One of the few scholars insisting that the distinct uses of discourse marker well “can all be related to one core meaning” is Jucker (1993: 438). Drawing on relevance theory, he argues that well “can be seen as a signpost which constrains the interpretation process and the concomitant background selection. It signifies that the most immediately accessible context
is not the most relevant one for the interpretation of the impending utterance” (Jucker 1993: 435).

3.2 Are there developmental stages in the data?

This section first addresses the question of whether the frequency development of the discourse marker *well* can be divided into a sequence of distinct stages. This is done through an application of the VNC algorithm that was described above. On the basis of the normalized text frequencies of *well*, the clustering algorithm returns a visual representation of how those frequencies can be grouped into clusters, a so-called dendrogram, which is shown in Figure 5. The frequency development of *well* is overlaid as a line plot. As can be seen, the algorithm creates clusters of the different time periods, starting with those neighbors that display minimal frequency differences. Hence, the 1990s and the 2000s are merged early on, as are the 1940s and 1950s. A question that the dendrogram does not immediately

![Figure 5. VNC dendrogram with overlaid line plot for frequencies of discourse marker *well* per million words and Period 1 (1920s–1980s) and Period 2 (1990s–2000s)](image)
answer is how many stages should be assumed in the development of *well*. Here, the analyst has to make an informed decision. We used a scree plot (not shown) as a diagnostic, which provided evidence that the development could be reasonably divided into either two or three stages. In order to bring out the differences between earlier and later uses of *well* as clearly as possible, and in order to keep the analysis as simple as possible, we decided to assume only two stages in the development of *well* in the nine TIME decades. The decades that are merged in these stages are shown by the bold horizontal lines in Figure 5: the first stage goes from the 1920s to the 1980s and the second includes the 1990s and 2000s.

Thus, the VNC algorithm and our periodization into two stages generates the hypothesis that discourse marker *well* in TIME undergoes a qualitative change from Period 1 (1920s–1980s) to Period 2 (1990s–2000s). The remainder of this section will discuss that qualitative change. To test the hypothesis that *well* is used differently across the two stages, we conducted a quantitative analysis of the collocates of *well* across the two periods. The collocates had to have a frequency greater than ten occurrences in at least one of the periods to be included in the analysis.

The results of the collocation analysis are given in Table 2. First note that the table includes not only *p* -values and asterisks to denote the significance level but also odds ratios (OR). These provide a helpful indication in which of the two periods a collocate is more frequent: if the OR value is greater than 1, this indicates that the collocate has become less frequent in the second period and, the reverse, if it is smaller than 1, the collocate has become more frequent in the second period. We observe that from Period 1 to Period 2, only the collocate *well* (of the node *well*, as in occurrences of *well well*), which has an OR of 4.57, has decreased in frequency; all other items have increased in frequency. Further, we see that almost all collocates have very highly significantly changed (***) ; only a few items are highly significant (**) or just significant (*). The underlined items in Table 2 will be subjected to closer scrutiny below.

<table>
<thead>
<tr>
<th>Collocate</th>
<th><em>p</em>-value</th>
<th>Odds ratio</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>,</td>
<td>&lt; 0.001</td>
<td>0.236</td>
<td>***</td>
</tr>
<tr>
<td>.</td>
<td>&lt; 0.001</td>
<td>0.403</td>
<td>***</td>
</tr>
<tr>
<td>?</td>
<td>&lt; 0.001</td>
<td>0.616</td>
<td>***</td>
</tr>
<tr>
<td>:</td>
<td>&lt; 0.001</td>
<td>0.366</td>
<td>***</td>
</tr>
<tr>
<td>IT</td>
<td>&lt; 0.001</td>
<td>0.373</td>
<td>***</td>
</tr>
<tr>
<td>THAT</td>
<td>&lt; 0.001</td>
<td>0.441</td>
<td>***</td>
</tr>
</tbody>
</table>
Table 2. (continued)

<table>
<thead>
<tr>
<th>Collocate</th>
<th>$p$-value</th>
<th>Odds ratio</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>&lt; 0.01</td>
<td>0.455</td>
<td>**</td>
</tr>
<tr>
<td>THE</td>
<td>&lt; 0.001</td>
<td>0.177</td>
<td>***</td>
</tr>
<tr>
<td>WELL</td>
<td>0.019</td>
<td>4.567</td>
<td>*</td>
</tr>
<tr>
<td>NOT</td>
<td>&lt; 0.001</td>
<td>0.343</td>
<td>***</td>
</tr>
<tr>
<td>WE</td>
<td>0.021</td>
<td>0.524</td>
<td>*</td>
</tr>
<tr>
<td>A</td>
<td>&lt; 0.001</td>
<td>0.125</td>
<td>***</td>
</tr>
<tr>
<td>THEY</td>
<td>&lt; 0.01</td>
<td>0.420</td>
<td>**</td>
</tr>
<tr>
<td>THERE</td>
<td>&lt; 0.001</td>
<td>0.300</td>
<td>***</td>
</tr>
<tr>
<td>“</td>
<td>0.037</td>
<td>0.464</td>
<td>*</td>
</tr>
<tr>
<td>OH</td>
<td>0.046</td>
<td>0.518</td>
<td>*</td>
</tr>
<tr>
<td>IS</td>
<td>&lt; 0.001</td>
<td>0.160</td>
<td>***</td>
</tr>
<tr>
<td>...</td>
<td>&lt; 0.001</td>
<td>0.068</td>
<td>***</td>
</tr>
<tr>
<td>IN</td>
<td>0.024</td>
<td>0.406</td>
<td>*</td>
</tr>
<tr>
<td>NO</td>
<td>&lt; 0.001</td>
<td>0.180</td>
<td>***</td>
</tr>
<tr>
<td>THEN</td>
<td>&lt; 0.001</td>
<td>0.264</td>
<td>***</td>
</tr>
<tr>
<td>WHAT</td>
<td>&lt; 0.01</td>
<td>0.377</td>
<td>**</td>
</tr>
<tr>
<td>LET</td>
<td>0.035</td>
<td>0.405</td>
<td>*</td>
</tr>
<tr>
<td>THIS</td>
<td>0.045</td>
<td>0.439</td>
<td>*</td>
</tr>
<tr>
<td>LIKE</td>
<td>&lt; 0.001</td>
<td>0.251</td>
<td>***</td>
</tr>
<tr>
<td>FOR</td>
<td>&lt; 0.001</td>
<td>0.174</td>
<td>***</td>
</tr>
<tr>
<td>SO</td>
<td>0.011</td>
<td>0.346</td>
<td>*</td>
</tr>
<tr>
<td>WAS</td>
<td>&lt; 0.001</td>
<td>0.163</td>
<td>***</td>
</tr>
<tr>
<td>BE</td>
<td>&lt; 0.001</td>
<td>0.212</td>
<td>***</td>
</tr>
<tr>
<td>TO</td>
<td>&lt; 0.001</td>
<td>0.192</td>
<td>***</td>
</tr>
<tr>
<td>AND</td>
<td>&lt; 0.001</td>
<td>0.088</td>
<td>***</td>
</tr>
<tr>
<td>ALL</td>
<td>&lt; 0.001</td>
<td>0.194</td>
<td>***</td>
</tr>
<tr>
<td>OF</td>
<td>&lt; 0.001</td>
<td>0.114</td>
<td>***</td>
</tr>
<tr>
<td>HERE</td>
<td>&lt; 0.001</td>
<td>0.243</td>
<td>***</td>
</tr>
<tr>
<td>RIGHT</td>
<td>&lt; 0.001</td>
<td>0.158</td>
<td>***</td>
</tr>
<tr>
<td>ABOUT</td>
<td>&lt; 0.001</td>
<td>0.107</td>
<td>***</td>
</tr>
</tbody>
</table>

The large number of significant collocates provides some support for our hypothesis that the use of *well* has changed across the two periods. Only a fraction of this change can be explored in detail here. We were particularly curious about
the presence of three forms of the verb be, namely is, was and be (underlined in Table 1). Casual inspection of relevant hits pointed to a tendency for these forms to act as copular be and for well to occur within the predicative construction complementing the copula.

Given that well used in statements within a predicative construction cannot by definition occur sentence-initially but must occur sentence-medially, we decided to narrow the scope of our attention to occurrences of well preceded and followed by a comma. Searching the TIME corpus for “, well,” we obtained 690 hits and subjected them to concordance analyses. We present the results of these analyses in the following section.

3.3 Concordance analyses of sentence-medial well

Close reading of the 690 hits suggested that three major syntactic patterns were prevalent in the data: (i) well introducing direct speech (hereafter “quote-well”), as in Example 1; (ii) well between clauses and/or constituents (“clause-well”), as in Example 2; and well as an element before the subject or object predicative (“predicative-well”), as in Example 3.

(1) …] people are going to have to say, well, gee, you know, what am I going to do? (TIME 1981)

(2) Cool is an elusive thing. If it weren’t, well, we’d all be cool, wouldn’t we? (TIME 2003)

(3) [The concept of the surround-sound headphone] [is], well, [heady]. (TIME 2004)

This latter syntactic pattern, predicative-well, can schematically be represented thus:

[Subject] [Copula], well, [Subject Predicative]

As shown in Table 3, the three patterns alone account for 62 percent of all 690 sentence-medial hits, with quote-initial use making up 11 percent, clause-well 22 percent and predicative-well taking up 29 percent of the hits total. The 38 percent
labeled “Others” form a wildly heterogeneous group; space considerations prevented their analysis in this paper.³

### Table 3. Frequencies of major syntactic patterns with sentence-medial *well*

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Hits</th>
<th>Percent (out of 690)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quote-<em>well</em></td>
<td>79</td>
<td>11</td>
</tr>
<tr>
<td>Clause-<em>well</em></td>
<td>152</td>
<td>22</td>
</tr>
<tr>
<td>Predicative-<em>well</em></td>
<td>198</td>
<td>29</td>
</tr>
<tr>
<td>Others</td>
<td>261</td>
<td>38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>690</td>
<td>100</td>
</tr>
</tbody>
</table>

All hits were hand-coded for these three categories, as well as a large number of subcategories, for further processing. Wherever necessary, the hits were viewed not only in the context of their concordance line but also inspected in their larger textual contexts to ensure reliable coding. The results of these fine-grained concordance analyses are detailed in what follows.

#### 3.3.1 Quote-*well*

The small sample of seventy-nine hits for quote-*well* nicely reflects the two-period division underlying the collocational analysis discussed in the previous section: while somewhat undecidedly going through ups and downs in the first period (1920s–1980s), quote-*well* takes off in the second period (1990s–2000s), as shown in the left panel of Figure 6.

An intriguing question relates to what types of quotation *well* is used with. Concordance analysis revealed that two such types were predominant,⁴ illustrated in the following two examples. Example 4 shows *well* in the context of interviews,

³. Sub-cATEGORIES of the occurrences labeled “Others” include the following.

- **Restart:**
  - This reason, *well*, where would they get the news as *TIME* gives it to them. *(TIME 1932)*
- **Multiple *well***:
  - S. McCoy, actuary of the Treasury Department. # “*Well*, *well*, *well*,” *said* the Finance Committee.” *(TIME 1924)*
- **Clause-initial** (often in combination with other inserts):
  - That’s another good story gone West; ah, *well*, *see* what you can get on this *(TIME 1930)*
- **Within if-clause**:
  - If they are asked why they think they qualify as, *well*, four Rockmaninoffs, they disarmingly concede (…) *(TIME 1964)*
- **Unclear**:
  - It goes like this (in B-flat):# HELLO, DOLLY, *well*, HELLO, DOLLY (…) *(TIME 1966)*

⁴. Besides occurrences in interviews and prose, there was a single instance of quote-*well* within a haiku poem.
where both the interviewer’s and the interviewee’s discourses are rendered as they may actually have occurred in the interview situation. Conversely, Example 5 illustrates quote-well as an integral part of the author’s prose; moreover, the quote itself is highly unlikely to have occurred in the way in which it is presented (have really all Midwestern football fans shrugged their shoulders and said exactly, “Oh, well, accidents will happen”?). It is much more likely that this quote is an instance of constructed dialogue (i.e., dialogue made-up and used for rhetorical purposes) – for example, to make the writing more lively and, thus, more engaging.

Figure 6. Frequencies-per million words (FPMs) for quote-well per decade (left panel) and FPMs for quote-well in prose and interviews respectively (right panel)

(4) # TIME So is Silicon Valley going to grind to a halt? Do you fear for the future of Intel?
# MOORE There’s still room for creativity. Designers are still going to have to think, Well, how do I use my billion-transistor limit? I don’t anticipate the end of innovation.   (TIME 2000)

(5) COLLEGE FOOTBALL # Illinois lost, 26–7, to Southern Methodist. Minnesota got shut out by Missouri, 24–0. Indiana was embarrassed, 20–10, by little Miami of Ohio. Oh, well, shrugged Midwestern football fans, accidents will happen. (TIME 1966)
To what extent is the overall increase of quote-well owed to changes in frequency of one of the two types? As can be seen from the right panel of Figure 6, both types increase in tandem, albeit only in the second period (1990s–2000s). Quote-well in interviews is not recorded in the data before the 1990s at all.

3.3.2  **Clause-well**

To judge by the size of the sample, the 155 hits coded as clause-well represent a more important use of the marker than the use labeled quote-well. All hits assigned to the category have in common that well is found in syntactic key positions, either between two clauses (either sub- and main-clause or two main clauses) or between a pre-posed adverbial phrase and the main clause. Whatever the type of syntactic relationship, well is invariably positioned in close left-hand proximity to the subject constituent of the main clause.

Consider Examples 6 to 9, where clauses and relevant constituents are annotated:

\[
\begin{align*}
\text{mainCl-well-mainCl:} \\
(6) & \quad [\text{mainCl Then M.J. left}] \text{ and, well, } [\text{mainCl [Subject the team] fell apart faster than a Rodman romance}….] \quad \text{TIME 1999} \\
\text{mainCl-well-AdvCl:} \\
(7) & \quad [\text{mainCl It’s so tempting to give up that distance prematurely}], [\text{AdvCl because, well, [Subject there] is so little distance after you put the Map of the States puzzle (…)}] \quad \text{TIME 1996} \\
\text{AdvPh-well-mainCl:} \\
(8) & \quad [\text{AdvPh As for kissing on the Seattle team bus}, \text{ well, [mainCl [Subject that] was just part of another little game called playing the pansy]. ”}] \quad \text{TIME 1970} \\
\text{AdvCl-well-mainCl:} \\
(9) & \quad [\text{AdvCl once you’ve seen Walking with Dinosaurs (Discovery, April 16, 7 p.m. E.T.)}], \text{ well, [mainCl [Subject you] still won’t have seen real animals do any of that].} \quad \text{TIME 2000}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Hits</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdvCl-well-mainCl</td>
<td>96</td>
<td>63</td>
</tr>
<tr>
<td>AdvPh-well-mainCl</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>main Cl-well-AdvCl</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>mainCl-well-mainCl</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>153</td>
<td>100</td>
</tr>
</tbody>
</table>

**Table 4.** Frequencies of syntactic patterns involving clause-well
As Table 4 shows, by far the largest subset is made up of instances where *well* acts as a relay contact of an adverbial clause and a main clause (the AdvCl-*well*-mainCl pattern illustrated in Example 8). This pattern alone accounts for 63 percent of all clause-*well* instances. It is this pattern that we will focus on.

Adverbial clauses require subordinators “indicating the relationship to the main clause” (Biber et al. 1999: 194). By far the most frequent conjunction in the subset is the subordinator *if*, the “most common and most versatile of the conditional subordinators” (Quirk et al. 1985: 1089) used to introduce condition adverbials, a subtype of contingency circumstance adverbials; condition adverbials “express the conditions which hold on the proposition of the main clause” (Biber et al. 1999: 779). Inspection of the *if*-clauses in the clause-*well* subset showed that their lengths were occasionally remarkable. Two such lengthy instances are shown in Examples 10 and 11.

(10) *if* once in a while a letter comes open in the postmistress’ hands, *well*, who can struggle against fate? (TIME 1949)

(11) If he feared that he’s seen as stiff and sardonic, still perceived as a hatchet man by those who recall his slash-and-burn campaign tactics as Gerald Ford’s 1976 vice-presidential running mate, *well*, then maybe he was right to use network TV’s hippest show to lighten his image. (TIME 1994)

Counting cliticized forms as separate words and hyphenated word clusters as single words, the *if*-clauses in the two examples contain thirteen words and thirty-one words, respectively. Are these lengths “unusual”? If they were longer than is usual for *if*-clauses it could be argued that *well* is used as a convenience to the reader, signaling both the termination of the conditional clause and the inception of the main clause. Beside the unusual length of many *if*-clauses we seemed to notice an increase in length over the decades of the TIME corpus. These two intuitions were put to the test. To measure unusualness of lengths we downloaded 100 sentence-initial occurrences of *if*-clauses without *well* preceding the main clause for each of the nine decades in the TIME corpus. From these altogether 900 examples, we randomly selected for each decade exactly the same number of occurrences as were found in the subset for *if*-clauses with *well* – namely eighty-seven occurrences. In both subsets, the number of words in the *if*-clauses was counted manually.

To test whether the *if*-clauses with *well* were longer than the ones without *well* we used the Wilcoxon rank sum test (given that, according to Shapiro – Wilk normality tests, both distributions violated normality). According to the test, the lengths of *if*-clauses are not greater than the lengths of *if*-clauses without *well* ($p = 0.56$, n.s.). So, contradicting our initial impression, *if*-clauses with *well* cannot be seen as unusually long and their assumed role as a convenience to the reader in
marking the boundary between adverbial clause and main clause is not supported statistically.

To test whether *if*-clauses with and without *well* are becoming longer over the nine decades of the TIME corpus (that is, whether there exists a correlation between the lengths and the decades), Kendall’s correlation tests were performed on both subsets (see Hilpert and Gries 2009 for a motivation for the use of rank-order measures like Kendall’s $\tau$ in diachronic corpus analyses). The tests suggest that *if*-clauses without *well* do not show any change in length ($\tau = 0.006, p = 0.94$), whereas *if*-clauses with *well* do show a weak but significant positive trend in which length and time are positively correlated ($\tau = 0.217, p = 0.007$). The trend can be observed in the boxplots shown in Figure 7.

![Figure 7](image)

**Figure 7.** Boxplots of lengths of *if*-clauses with *well* per decade

Despite occasional dips, the median lengths per decade depicted in the bold horizontal lines across the boxes show an overall upward trend as we move from the 1920s to the 2000s. The assumption that the use of *well* between *if*-clause and main clause is strategic may after all not be entirely wrong. We will discuss this possibility in the discussion section later.
3.3.3 Predicative-well

We mentioned earlier that the use of *well* at the onset of a quotation (quote-well) is a widely discussed research topic. The third major category we found in our data, predicative-well, has, by contrast and to the best of our knowledge, not yet been described anywhere. With 198 examples identified, this subset is the largest in our data and may be regarded as the most common and most important use of *well* in sentence-medial position in TIME magazine. As noted, the marquee feature of predicative-well is that the marker is placed before the subject or object predicative constituent of the sentence; in a few rare cases, *well* can act as an element of the adverbial predicative (also referred to as the “adverbial complement”; see Quirk et al. 1985: 1171). The three sub-categories are illustrated in Examples 12 to 14.

Subject predicative:

(12) After all, [Subject Soviets] [Verb are], well, [Subj.Predicative Soviets]. (Time 1987)

Object predicative:

(13) […] reflects the interests of the citizens, [Subject who] [Verb consider] [Object themselves], well, [Obj.Predicative unique]. (TIME 2002)

Adverbial predicative:

(14) […] booming baritone [Subject that], on his five bestselling albums, [Verb sounds] [Adv.Predicative vaguely like, well, a fellow hollering down a drainpipe].

Table 5 shows that, not surprisingly given the distributions in other contexts (see Quirk et al. 1985), the three types are very unequally distributed, with the subject predicative accounting for the lion’s share (89 percent), the object predicative for 9 percent and the adverbial complement for 3 percent.

<table>
<thead>
<tr>
<th>Predicative type</th>
<th>Hits</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>176</td>
<td>89</td>
</tr>
<tr>
<td>Object</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Adverbial</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>198</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

It is, furthermore, not surprising either to see that by far the most frequent copula in all three types of predicative-well is the copula BE: its 144 forms alone account for 73 percent of all copular verb forms. This is post hoc confirmation that in paying close attention to *be*, forms of which emerged as significant collocates of *well* in the second stage (1990s–2000s), we were on the right track. We now know that they were significant collocates because of *well’s* association with the predicative construction, where copular *be* has no close rival.
Table 6. Distribution of copula verbs for predicative-well

<table>
<thead>
<tr>
<th>Copula verb</th>
<th>Hits</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>144</td>
<td>79.5</td>
</tr>
<tr>
<td>SEEM</td>
<td>11</td>
<td>6.1</td>
</tr>
<tr>
<td>BECOME</td>
<td>7</td>
<td>3.9</td>
</tr>
<tr>
<td>SOUND</td>
<td>5</td>
<td>2.8</td>
</tr>
<tr>
<td>Other verbs</td>
<td>14</td>
<td>7.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>181</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

It is instructive to consider predicative-well’s co-text not only in terms of syntax but also in a lexical and stylistic light. Two lexico-stylistic features are noteworthy. First, predicative-well co-occurs with repetition, where a lexical item used before in the same sentence (often but not necessarily in the subject constituent) or in the preceding sentence is simply recycled after well (in the predicative constituent). For example, repetition concerns the lexemes dumb in Example 15, secret in Example 16, tax in Example 17 and healthy in Example 18. Second, many co-texts of predicative-well are characterized by word play. For instance, in Example 19, the author plays with the phonetic similarity of two completely distinct lexemes (decade and decadent); in Example 20, the author juxtaposes opportunity and opportunism thereby playing with the evaluative dissimilarity between them (opportun-ity considered to be ‘good’ and opportunism ‘bad’); in Example 21, the phonetic material contained in the name Cunningham is re-used in the adjective cunning (a pun the poet was probably quite familiar with); in Example 22, a more successful pun, the conceptual association of (female) nanny and household is creatively exploited for the blend formation manny (merging man and nanny).

(15) […] making fun of yourself for being dumb is, well, dumb. (TIME 1999)
(16) Until quite recently, cryptography – the science of making and breaking secret codes – was, well, secret. (TIME 1994)
(17) The worst knock against a gas tax is that it is, well, a tax. (TIME 2004)
(18) […] are well stocked with healthy snack products, but they tend to taste distinctly, well, healthy. (TIME 2003)
(19) After a decade, open tennis is becoming, well, decadent. (TIME 1978)
(20) Why reach for this opportunity? One possible explanation is, well, simple opportunumism. (TIME 1993)
(21) […] and if anything, Cunningham has only got more audacious and more, well, cunning in the past six years. (TIME 2004)
(22) [on male nannies] Trend spotter Faith Popcorn predicts that in the next decade, “manny” (as the guys are called) will become, well, a household word. (TIME 2002)

This results section reported a large number of findings. We will briefly contextualize them in the following section.

4. Discussion

4.1 Inserts

In Section 3.1 we examined the possibility that inserts show a positive trend in the TIME Corpus. We found that, indeed, the large majority of the features investigated (eight of ten) showed moderate-to-strong frequency increases. Since inserts are key components of colloquial speech, we certainly have a strong case of colloquialization here: inserts contribute to TIME magazine writing becoming more speech-like. A number of explanatory hypotheses have been proposed in recent research to account for the phenomenon of colloquialization, which is “presumably attributable to external, social factors, rather than purely internal, linguistic change” (Leech et al. 2009: 49). The factors proposed include editorial changes of style guidelines (e.g., Millar 2009: 212), an increase in direct speech (Leech et al. 2009, see below), the reduction of male bias (Baker 2009), as is the case in the increasing tendency to avoid generic he, popularization (Biber 2003), a trend intended to appeal to wider audiences by making written text “more engaging, accessible and easy to process” (Leech et al. 2009: 245), and, most notably, democratization (Leech 2003; Millar 2009). This latter hypothesis describes a “tendency to phase out markers of distance, respect, superiority or inferiority, and to aim at the expression of greater equality and familiarity” (Leech at al. 2009: 259). Democratization can plausibly be observed in the well-attested fall in British and American English of “modals of authority” such as shall, ought and must (Leech 2003; Millar 2009). To the extent that conversation “as the most commonplace, everyday variety” (Biber et al. 1999: 1038) represents the most common linguistic resource shared by everyone, the increased use of inserts may be seen as an “expression of greater equality and familiarity” and thus of democratization. At any rate, inserts certainly contribute to “a kind of spontaneous directness which (though often contrived) is clearly supposed to inject into journalistic discourse some of the immediacy of oral communication” (Leech et al. 2009: 239).

We further witnessed seven of ten features make their greatest relative leaps in frequency in the two most recent decades, specifically in the 1990s. Acceleration in the use of informal or colloquial features has also been noted in other news
varieties: Westin (2002) reports accelerated use of informal style features towards the end of the twentieth century for English newspaper editorials. Investigating what motivates this accelerated colloquialization in the 1990s and the 2000s is far beyond our aims here, although it represents an exciting research question for future research. All we would like to do here is point to an intriguing temporal coincidence. Much of the language of the new online media – such as e-mail, tweets on Twitter and chats on Facebook – undoubtedly “looks like speech” (Baron 2003: 85). While Twitter, What’s App and Facebook are more recent inventions, the “proliferation of email within the broader public arena (…) seriously began in the 1990s” (Baron 2003: 85, emphasis added) – precisely the decade in which the majority of the most dramatic hikes in frequency were observed for inserts. Many people spend long hours using e-mail and even longer hours on the more recently developed online media, where the language is even more informal and more colloquial, and users are becoming more and more familiar with seeing “written speech” (Crystal 2001) on their electronic devices. It is, therefore, tempting to assume that extended exposure to written speech somewhat blurs the strict separation of speech and writing, and that TIME magazine writers take more liberties in printed news reportage because they know that in their readers’ language experience the two modes have come to overlap. To the extent that a correlation can be posited between speech and writing on the one hand and the private and the public on the other, the increased familiarity of written speech makes an interesting parallel to the increased blending of the private and the public spheres that has been observed for the twentieth and the early-twenty first century (see Landert and Jucker 2011). We note that the possibility that the colloquial character of online media language has any effect on written news magazine prose is at present merely a speculation, but one, we feel, worth investigating in more detail in future.

4.2 Quote-well

We saw that one important pattern of use for well was quote-well (well used at the onset of direct speech). It was also noted that quote-well sees its major increases in the 1990s and 2000s. These findings then support earlier research suggesting that colloquialization is in part due to the fact that writers include more and more quotation in their writing. For example, Leech et al. (2009: 249) observe that “direct speech, an environment traditionally appropriate for colloquial forms, has increased”. So, the findings on sentence-medial well indicate that a potential language-internal factor contributing to colloquialization may lie in “dramatization”, that is, an increase in (direct speech) quotation in news magazine discourse (“writing containing more speech”). A two-way distinction between quote-well in interviews and in prose was made, with quote-well in interviews introducing the
interviewee’s actual speech and quote-well in prose presenting hypothetical speech made up by the writer for rhetorical purposes. Functionally, the two forms occupy clearly distinct territory. Quote-well in interviews is a fair reflection of quote-well in conversation, where the discourse marker is commonly used at the onset of direct speech, thus separating the speaker’s own discourse from the quoted discourse. The quote-initial use of well in casual conversation has been amply documented in the literature (for overviews see, for example, Aijmer 2013; Rühlemann 2013). Its function has been described as an “utterance opener” signaling that the speaker “is embarking on direct speech quotation” (Biber et al. 1999: 1118). This function is particularly convenient to the hearer in that well used at the onset of direct speech attends to the “boundary issue” (Rühlemann 2013: 119) demarcating the left-hand boundary of discourse that requires processing “in another context”: not as the present speaker’s words but as a remote speaker’s words with all referential expressions pertaining to that remote speaker (for a discussion see Rühlemann 2013: Chapter 4). It then serves as a “contextualization cue”, “marking off segments in the discourse thus helping the hearer to understand how the stream of talk is organized” (Aijmer 2013: 6) and its function is that of providing coherence. The need to facilitate coherence in written discourse is greatly reduced because of the availability of typographical means such as colons, commas and quotation marks; these sufficiently set off the quote as a separate discourse unit and as to be processed in another light, as shown in Example 23.

(23) The inquiring police take a cheeky tone with him: “Yes, well, your… friend has gone a bit missing, to tell you the truth. (TIME 1990)

Quote-well in prose then works less as a contextualization cue and its coherence function is not primary (a role largely taken over by typography). Its central role must lie in creating a conversation-like “rapport” with the reader where events and ideas are expressed in a narrative style using (hypothetical) direct speech and thought.

4.3 Clause-well

The most important finding in the subsection on clause-well was the intimate association of well with syntactic structure. We found that well was used in complex sentences, where it showed a strong attraction to positions right before the inception of the main clause and, thus, in close proximity of the subject constituent of that main clause. Given the complexity of the sentences well was involved in, and its marked preference for occurrence right before the main subject, the role of well probably lies in signaling to the reader that the structural “preliminaries” of the sentence expressed in the fronted adverbial clause or phrase are over and that the
sentence is embarking on its main business, as expressed in the main clause. An interpretation of *well* occurring in these contexts seems particularly plausible in cases where *well* bridges over to the main clause from a left-branching adverbial clause of some length. Here, it may be argued, the intrusion of *well* is a convenient service for the reader to whom *well* announces the eventual arrival, after a long-winded adverbial prelude, at the sentence subject and its predication. Clause-*well*, then, serves a coherence function elucidating sentence and syntactic structure.

We tested whether *if*-clauses that contained *well* in left-of-subject position were of greater length than a random subset of *if*-clauses without such *well*. The result was negative: no significant differences in length could be found. However, this cannot be taken as evidence that *well* has no structure-marking function. TIME magazine authors can still deploy *well* as a structural marker although the preceding *if*-clause is not longer than other *if*-clauses.

We also tested whether *if*-clauses with *well*, although on average not of greater length than *if*-clauses without *well*, could be seen as increasing in length over the decades represented in the corpus. This test yielded a significant result: while *if*-clauses without *well* did not become longer over time, *if*-clauses with *well* did exhibit a modestly strong but significant correlation with the TIME decades. There is a possibility, then, that we are witnessing the emergence of a new function for *well*: as an analytical marker sign-posting, for the benefit of the reader, clausal structure in complex sentences whose complexity is exacerbated by ever growing pre-subject material. Its macro-function, if this interpretation is correct, is as a processing aid contributing to text coherence for the reader by “providing the ‘grease’ between parts of discourse” (Aijmer 2013: 31).

### 4.4 Predicative-*well*

The largest subset in the data showed *well* in close association with the predicative constituent. It was noted that the contexts of this type of *well*, called predicative-*well*, were marked lexico-stylistically in that repetition and word play were frequent in the predicative.

What pragmatic functions does *well* fulfill in these contexts? It appears that the common functional denominator to both lexico-stylistic variants is as a marker of word choice. As a word-choice marker, predicative-*well* prepares the reader for upcoming wording which is, in one way or another, peculiar (or “marked”), either as repetition or as word play. Repetition is peculiar in that it violates the principle of “elegant variation” that authors of news magazines, as other written text types, are normally held to observe. Word play is peculiar in that, in order to be enjoyed by the reader, it needs to be recognized as word play. Subtle word play may go unnoticed, so *well* may help to make the reader become aware of it.
The function as word-choice marker stands in an interesting relation with the function as a marker of word search *well* commonly fulfills in conversation (see Aijmer 2013). Obviously, magazine writers do not use *well* because they were at a loss to supply the right word. The recycled lexical item (in repetition) or the manipulated item (in word play) in the predicative is deliberately chosen and so is the use of *well* at the onset of the predicative: it serves to signpost the choice as such and bring it to the reader’s attention. Nonetheless it seems possible to argue that the word-search function is the model after which the word-choice function is crafted. That is, *well* is used as if the writer were searching for the appropriate wording. The effect is carefully calculated: just as conversational word-search *well* indexes the speaker’s planning difficulties drawing the interlocutor’s attention to the searched-for wording so too does predicative-*well* focus the reader’s attention on the expression to follow.\(^5\)

Bluntly re-using lexical material in close vicinity to its first use is generally considered bad style and therefore avoided. Seen through this lens, *well* in the context of repetition is reminiscent of *well*’s function in conversation, where it “standardly prefaces and marks dispreferreds” (Levinson 1983: 334) – that is, adjacency second-pair parts which are in some way contrary to the expectations raised by the preceding adjacency first-pair part (e.g., refusing an invitation). Just as *well* indexes the conversationalist’s awareness that the (negative) response is generally avoided and dispreferred over the positive one, so TIME authors use *well* to index their awareness that re-use of a lexical item is stylistically dispreferred. Unlike conversationalists, however, who may have no alternative to producing dispreferreds due to constraints beyond their control (e.g., being busy at the time an invitation takes place), TIME authors are certainly linguistically versatile enough to have alternatives at their disposal. Still, they do not use them, thus violating the principle of elegant variation. The violation is purposeful: the aim is to create a bond with the reader by means of language play (see Crystal 1998) and so involve them in the text (Aijmer 2013: 37). While the playfulness may not be immediately obvious to the reader in repetition, it moves center-stage in word play. The ludic manipulation of lexical material is an act of camaraderie on the part of the author intended to strengthen the bondage with the reader and make them an accomplice in the discourse. While, then, the context predicative-*well* is found in serves an involvement function, the marker itself serves to draw attention to this context.

As noted, we are not aware of any discussion of predicative-*well* in the literature. If indeed this use has not yet been discovered elsewhere, either in written or in spoken text types, we feel justified, in the absence of counter-evidence,

\(^5\). The use of *well* as if it were for planning marks a switch from indexical use in conversation (as a symptom of planning ahead) to iconic use in writing (as a likeness of planning ahead).
to hypothesize that predicative-\textit{well} may exist only in writing (specifically TIME magazine writing) but not in conversation. If this were the case, predicative-\textit{well} would represent an intriguing case of conversational language entering the written language, emancipating itself from its conversational origin, as it were, and developing its own life there. Further, if predicative-\textit{well} were indeed not conversational and hence not colloquial, it would become questionable whether it can be seen as an instance of colloquialization. Strictly speaking, predicative-\textit{well} is colloquialized only to the extent that it involves the colloquial marker \textit{well}; its precise use in writing, however, as an element within the predicative construction, is anything but colloquial: rather, we have support for Aijmer’s (2013: 12) notion that part of the “meaning potential” of pragmatic markers is their adaptability to new contexts – in the case of predicative-\textit{well}, the adaptation is to the context of TIME magazine writing, which is syntactically highly complex and generically quite constrained.

5. Conclusion

To conclude this paper, this section spells out the implications of our empirical findings for the over-arching theme of colloquialization and for questions of methodology.

As regards colloquialization, our study suggests that the repertoire of colloquialized features is larger than has hitherto been observed. Next to the attested features of colloquialization, inserts are gaining ground, especially in the 1990s and the 2000s. The close analysis of \textit{well} indicated a number of factors co-driving this increase. Analysis of quote-\textit{well} suggested the possibility that the rise may be fueled in part by increases in the use of interviews (a genre in which the construction of direct speech is inevitable) and by increased use in prose (where \textit{well} acts as a rhetorical device to inject direct-speech-like immediacy into the text). Moreover, examination of clause-\textit{well} and predicative-\textit{well} suggested that \textit{well} is gaining in frequency because it is entering into new syntactic contexts, such as the left-of-subject position in complex sentences and the predicative construction, and that, in these new contexts, it is taking on new functions, for instance as an analytical marker flagging clausal structure and, respectively, as a word choice marker indicating playful language use.

Finally, we noted the possibility that colloquialization may be driven not only by the factors identified in previous research such as democratization, reduction of male bias, and changes in editorial guidelines. Given the temporal coincidence in the 1990s of the mass adoption of e-mail and the massive increase in the use of inserts in TIME, we speculated that another factor may be the influence from new media language which is often highly colloquial in style and vocabulary and
where daily exposure may favor processes of familiarization with “written speech” (see Crystal 2001).

Methodologically, our study has implications on three levels. First, the study suggests that, in investigating colloquialized writing, it is useful to take colloquial speech as the starting point. Through keyness analyses, colloquial items can be readily identified and examined. We are aware that using key items from the British National Corpus and examining them (or their equivalents) in a corpus of American English is not without problems. It is undoubtedly preferable in future research to base analyses of colloquialization in a given national variety on lists of key items derived from the same national variety. However, we maintain that the hypotheses generated by the data from British English did lead us towards the discovery of on-going changes in American English.

Second, the case study of well indicates that the identification of frequency trends needs to be followed up with investigations of qualitative changes. Temporally continuous data, as provided by the TIME Corpus, are ideal for that purpose.

Third, the analysis of well illustrates how corpus pragmatic research can integrate key methodologies of corpus linguistics and pragmatics. The quantitative corpus-linguistic analyses have indicated what to look for, highlighting novel uses of well that a solely qualitative analysis might have overlooked. On the other hand, the qualitative pragmatic analyses have unearthed the marker’s new forms and functions. Without this contribution, a merely quantitative analysis of well might have failed to find and appreciate what the marker contributes to the context and the writer – reader interaction. To conclude, we hope that the results reported in this paper will inspire further analyses that take full advantage of the corpus pragmatic approach.

Sources


References


doi: 10.1017/CBO9781139164771


doi: 10.1016/0378-2166(89)90054-4


doi: 10.1080/09296170902975692


doi: 10.1093/llc/fqn012


doi: 10.1017/CBO9781139004206


doi: 10.1075/pbns.57.10juc


doi: 10.1017/CBO9780511642210


doi: 10.1017/CBO9780511611841

Scott, Mike and Christopher Tribble. 2006. Textual Patterns: Key Words and Corpus Analysis in Language Education. Amsterdam: John Benjamins. doi: 10.1075/slcl.22


Authors’ addresses

Christoph Rühlemann
Fremdsprachliche Philologien
Institut für Anglistik und Amerikanistik
Philipps-University Marburg
Wilhelm-Röpke-Straße 6
35032 Marburg
chrisruhelmann@googlemail.com

Martin Hilpert
Université de Neuchâtel
Institut de langue et littérature anglaises
Espace Louis-Agassiz 1
CH-2000 Neuchâtel
mhilpert@gmail.com

Biographical notes


**Martin Hilpert** works as Professor of English Linguistics at the University of Neuchâtel. He holds a PhD from Rice University. His interests include cognitive linguistics, language change, construction grammar and corpus linguistics. Martin Hilpert is the author of *Germanic Future Constructions* (John Benjamins, 2008), *Constructional Change in English* (Cambridge University Press, 2013), and *Construction Grammar and its Application to English* (Edinburgh University Press, 2014). He is editor of the journal *Functions of Language* and associate editor of *Cognitive Linguistics.*