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

Abstract

Recent analyses of written text types have discovered significant frequency increases of colloquial or conversational elements, a trend often referred to as colloquialization (e.g., Leech et al. 2009). Among the elements that have made inroads into writing are verbal and negative contractions (e.g., Leech 2002), personal pronouns (e.g., Baker 2009), questions, the progressive, as well as zero relative clauses (e.g., Leech et al. 2009). This paper presents a new perspective on this phenomenon. The paper is structured into two parts. In the first part we present new evidence of colloquialization on the basis of the TIME Magazine Corpus (Davies 2007). Spanning nine decades (1920s-2000s), the TIME Corpus allows analyses of diachronic change in written American English throughout the 20th and early 21st century. 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*, *oh*, etc.), backchannels (*yeah*, *uh-huh*, etc.), and hesitators (*uh*, *um*, etc.), and which are "especially frequent in spoken texts" (ibid.). The main conclusion from the first part of this paper is that inserts significantly increase diachronically in TIME. In the second part of the paper, we study colloquialization more deeply by analyzing morphosyntactic and pragmatic structures in the data. We focus on a single insert, viz. the element *well* in its function of a discourse marker. Through a combination of quantitative and qualitative analytical steps, we analyze the diachronic development of *well* 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 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 & 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 20th 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 2002), 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, 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 going back to the 17th century (Leech & 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, personal communication).While the approach is not intrinsically problematic, it does not lend itself to 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 therefore propose an alternative methodological route, namely from conversation to exploration. To investigate colloquialization we start out from conversation, 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) (cf. Scott & Tribble 2006: Ch 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 (cf. Hilpert & 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. Perhaps an even more revealing approach is to inquire whether there is structure in the data, which is a question that can be asked regardless of there being a frequency trend or not. Structure in diachronic corpus data can offer important insights into how a given feature changes not only in terms of frequency ups and downs but also in terms of the feature's use and functions in context. However, the analysis of diachronic-data structure is statistically and analytically demanding. Statistically, it requires methods of exploratory statistics to generate hypotheses; once a hypothesis is formulated, significance testing is needed. Analytically, it requires close scrutiny of target items in the co-text of their concordances and, if necessary, in the wider context of their individual texts; also, it involves detailed annotating and processing of contextual characteristics. Given the methodological complexity, we will restrict this part of the investigation to a single element, namely *well* in its function as a discourse marker.
Finally, precisely because we will examine structure in diachronic data, this study 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" (cf. Rühlemann & 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* 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 schematically shown in Figure 1:

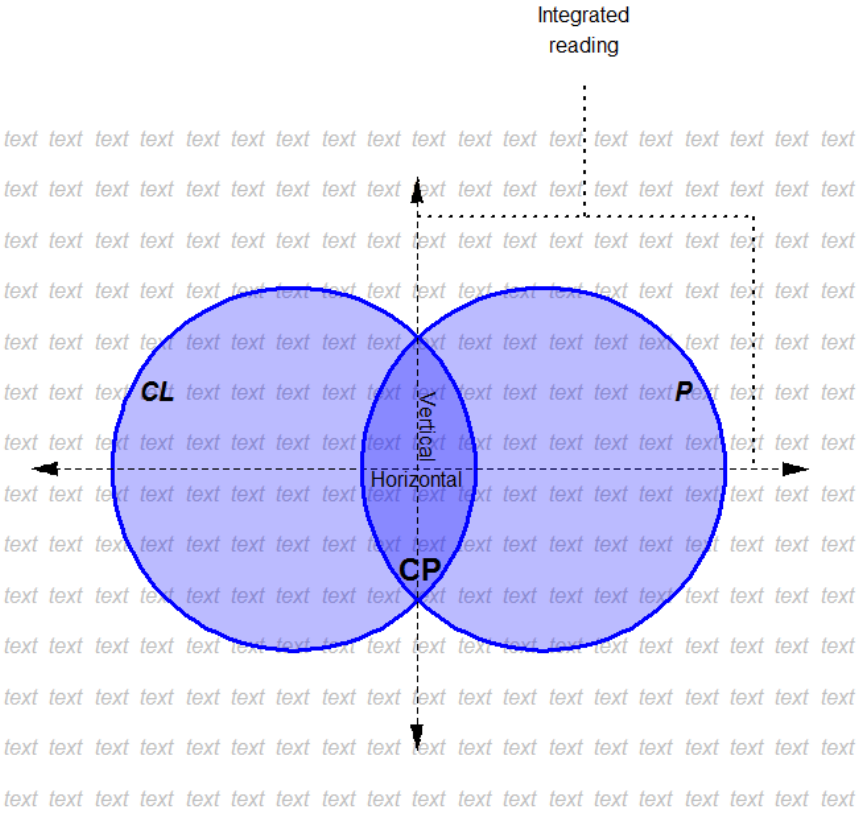

Figure 1: Integrated-reading methodology in corpus pragmatics (CP) combining the vertical-reading methodology of corpus linguistics (*CL*) with the horizontal-reading methodology of pragmatics (*P*)

The present 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 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 valuable 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 a regular research paper, with an outline of the methods and the data used (Section 2), a detailed report of the results (Section 3), a discussion of the results (Section 4), and a final section (Section 5) summarizing the main findings and drawing 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 against the whole of the written component. The top 20 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.

Table 1. Top 20 key words in demographically-sampled subcorpus (C) against the whole of the written component (W) of the BNC

N	Word	Tag	Freq in C	Freq in W	Log Likelihood
1	yeah	ITJ	58,706	1,386	344216.19
2	i	PNP	167,426	547,004	294395.23
3	you	PNP	134,910	398,899	256476.85
4	's	VBZ	89,938	144,957	248882.05
5	oh	ITJ	41,555	14,456	190335.54
6	n't	XX0	77,480	189,912	168179.21
7	no	ITJ	32,988	20,352	131822.81
8	mm	ITJ	21,888	227	130696.3
9	er	UNC	21,345	952	122125.87
10	that	DT0	63,324	186,217	120880.07
11	it	PNP	127,977	799,065	112700.28
12	erm	UNC	16,605	187	99017.99
13	do	VDB	42,266	116,247	84742.69
14	got	VVN	22,545	18,471	82553.54
15	well	AV0	34,608	75,294	81044.9
16	know	VVB	21,547	27,658	66381.18
17	cos	CJS	11,374	353	66085.19
18	've	VHB	22,738	36,628	62939.22
19	yes	ITJ	17,866	18,813	59743.28
20	na	TO0	9,966	728	55403.89

At the same time, the list prominently features items that have not yet been examined in studies on short-term diachronic change; most notably, it contains what Biber et al. (1999: 56) term '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 1st), *no* (7th), *mm* (8th) and *yes* (19th) 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.

Which inserts were investigated in the present study? Following closely the BNCweb-derived keyword list, 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-pair 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 used (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, viz. 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 2012, Tottie 2014).

The corpus underlying the subsequent analyses is the TIME Corpus (Davies 2007). This corpus contains all texts of TIME magazine ranging from 1923 to 2006. In terms of size, the collection amounts to roughly 110 million words. All the texts are time-stamped offering the exact dates of publication; they are ordered and searchable by nine decades, ranging from the 1920s to the first decade of the 21st century. Thanks to the unique temporal continuity of the data, this corpus allows the analysis of research questions that could not be pursued with corpora that sample language use in a temporally discontinuous fashion. For example, Leech et al.'s (2009) matching one-million word corpora B-LOB, LOB, and F-LOB cover data from the 1930s, 1960s, and the 1990s but the 'gaps' between those decades are unaccounted for.

For data retrieval, automatic searches were viable only for some items, such as *yeah*, which is orthographically stable and virtually always acts as a response token. Restricted-recall searches are necessary where the item in question is polysemous, as for example, *well*. To get as precisely as possible at uses of *well* as a discourse marker, rather than an adverb or a noun etc., we constructed a search string[i] 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 typographic means). Finally, extensive manual editing of results was necessary for orthographically inconsistent forms such as hesitators[ii] and the different spellings of COZ, viz. *cos, coz,* and *'cause*.

The analysis addresses three questions which require three different methods of statistical analysis. The first question, "Is there a frequency trend?", can be answered fairly easily by feeding frequencies per million words (FPMs) and decades into correlation tests. These yield correlation values such as Kendall's tau, which range from 0 (meaning no correlation at all) to 1 (indicating a perfect correlation). The second question, "Is there structure 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 find structure in temporally ordered data is Variability-based Neighbor Clustering (VNC), first proposed by Gries & Stoll (2009) and taken further in a number of other works (e.g., Hilpert & Gries 2009, Hilpert 2013). VNC is a form of Hierarchical Clustering Analysis (HCA) that is applied to temporally ordered data in such a way that only temporally adjacent data points are allowed to merge and form clusters (Hilpert & Gries 2009: 390). Further, it is an exploratory method, serving not to test a hypothesis but to generate one. The hypothesis generated by VNC typically 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. 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 statistically significant, we used the Fisher Yates exact test. The test results are interpreted as indications of language change: if the collocates of *well* differ significantly 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, "Is there structure in the data?". Given the complexity of the methods used to address the question, we narrow the focus to a single marker, viz. 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 1, 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 noticable increase in 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.

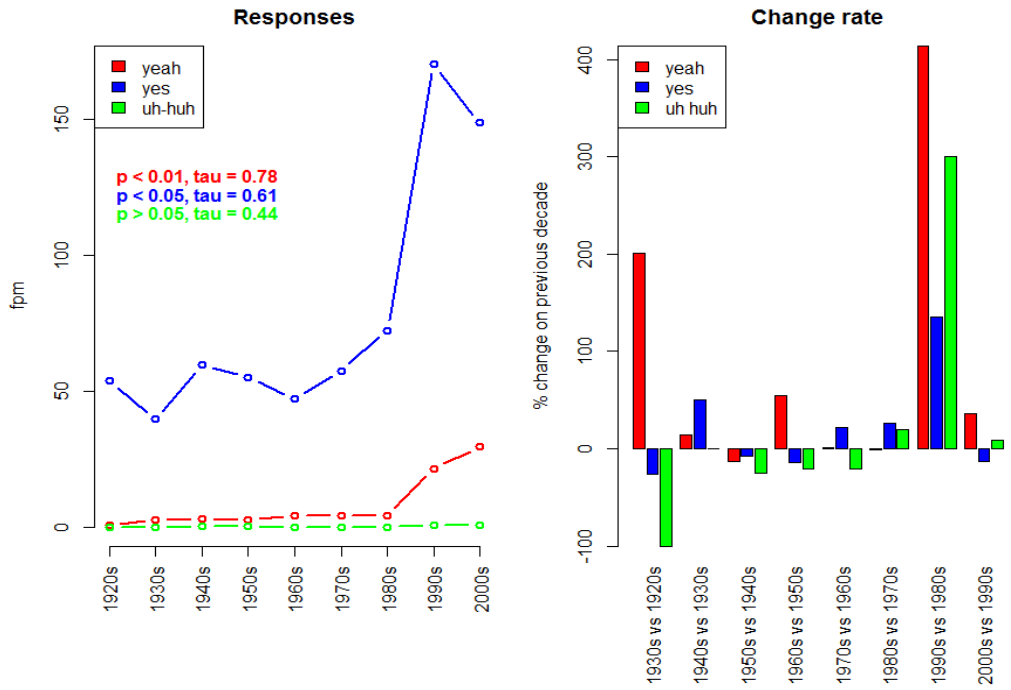

Figure 1: Frequencies per million words of responses (left panel) and change rates on previous decades (right panel)

A similar picture emerges for the discourse markers *oh*, *COZ*, *you know*, and *well*, which are shown in Figure 2 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 in the last two decades, as do *you know* and *well*. Only COZ sees its main change rate in earlier decades.

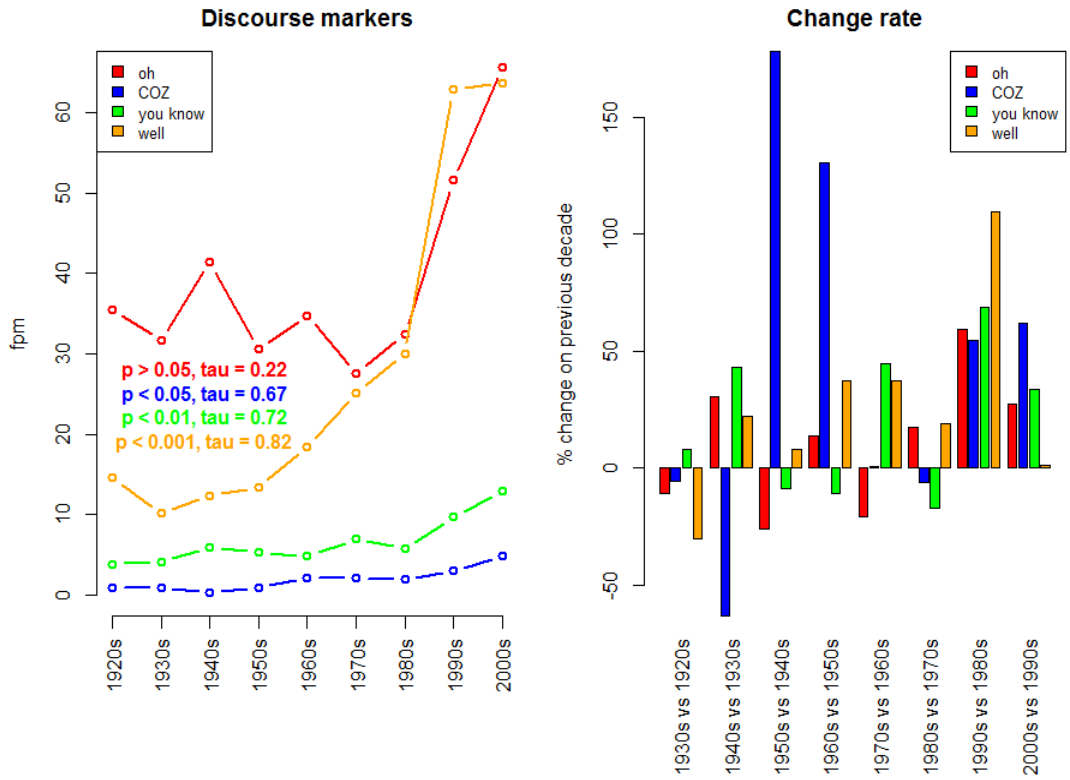

Figure 2: Frequencies per million words of discourse markers (left panel) and change rates on previous decades (right panel)

The picture emerging for the three hesitators *uh*, *um*, and *er*, shown in Figure 3, looks familiar too: all three forms are modestly strongly correlated with the decades. As regards change rate on earlier decade, only *um* sees its highest relative growth in the 1990s; *uh* and *er* change most dramatically in earlier decades:

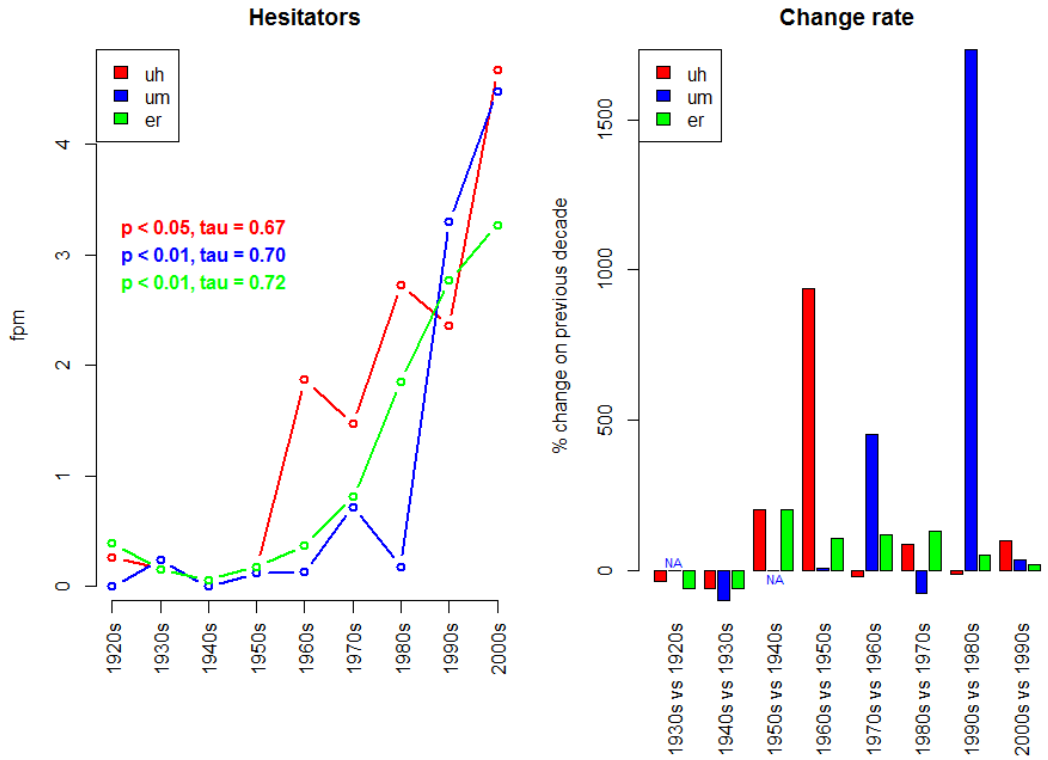

Figure 3: Frequencies per million words of hesitators (left panel) and change rates on previous decades (right panel)

To summarize what we have so far: 8 of 10 features investigated are modestly strongly to strongly correlated with the decades, 7 of 10 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. But how instructive is the observation of trends? Suppose that, for some reason, inserts were to see a sharp drop in the next decades into the 21st century, the overall trend would be seriously diminished or collapse altogether. Spotting a trend is only initially worth your while since the time span underlying it is inevitably random (there is no *principled* reason for examining inserts between the 1920s and the early 2000s in TIME magazine since the magazine is still alive and well in the 2010s and perhaps beyond that too). What is more revealing is to investigate whether, within the random time span, there is structure in the data. We can 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 Variability-based Neighbor Clustering 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*.

3.2 Is there structure in the data?

This section first addresses the question 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 dendrogram, i.e. a

visual representation of how those frequencies can be grouped into clusters. This dendrogram is shown in Figure 4 below. The frequency development of *well* is overlaid as a simple 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 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. For reasons of ease of analysis, we decided to assume only two stages in the development of *well* in the nine TIME decades. Which decades are merged in these stages, is shown by the bold horizontal lines in Figure 4: the first stage goes from the 1920s-1980s and the second includes the 1990s and 2000s.

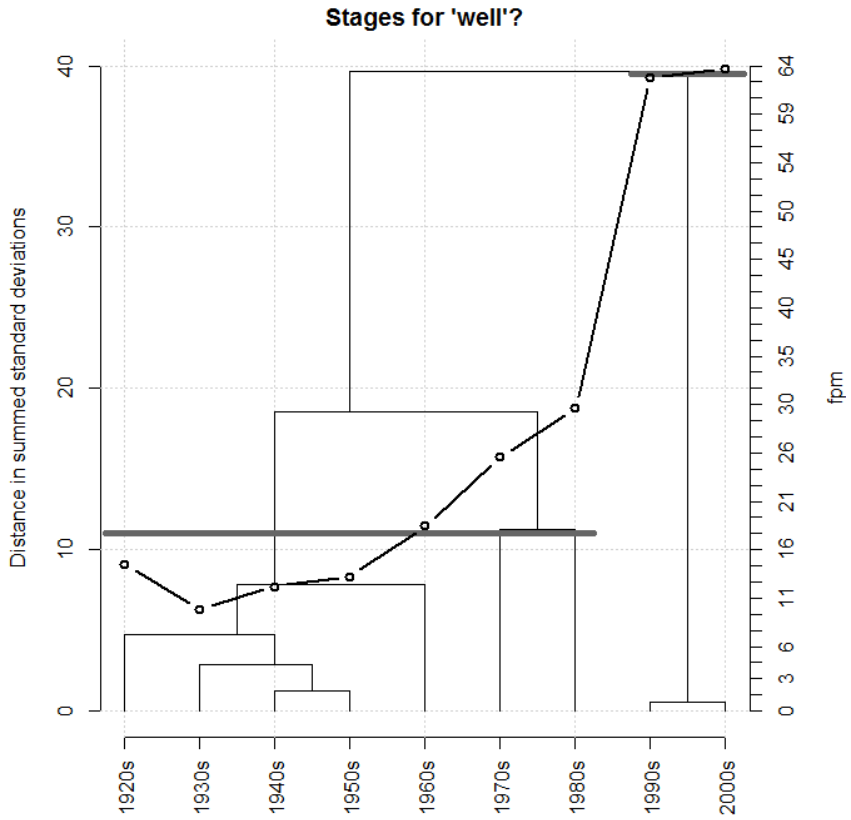

Figure 4: VNC dendrogram with overlaid line plot for frequencies of discourse marker *well* per million words and periods 1 (1920s-1980s) and 2 (1990s-2000s)

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 10 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 are 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 2: Significant collocates (L2-R2) of discourse marker *well* in period 1 (1920s-1980s) vs period 2 (1990s-2000s)

collocate	**p-value**	**odds ratio**	**significance**
,	5.45e-314	0.23559	***
.	3.15E-35	0.40268	***
?	4.82E-10	0.61591	***
:	3.75E-09	0.36621	***
IT	1.29E-09	0.37257	***
THAT	5.00E-05	0.44123	***
HE	0.00154	0.455	**
THE	5.24E-18	0.17676	***
WELL	0.0185	4.56707	*
NOT	0.000202	0.34263	***
WE	0.0208	0.52399	*
A	2.11E-20	0.12542	***
THEY	0.00656	0.42049	**
THERE	3.19E-05	0.30031	***
"	0.0368	0.4636	*
OH	0.0456	0.51757	*
IS	2.10E-12	0.16005	***
...	8.68E-33	0.06785	***
IN	0.0236	0.40564	*
NO	2.53E-08	0.17962	***
THEN	0.000156	0.26381	***
WHAT	0.0033	0.37672	**
LET	0.0346	0.40522	*
THIS	0.0452	0.43913	*
LIKE	2.34E-05	0.25144	***
FOR	1.49E-06	0.17426	***
SO	0.0112	0.34556	*
WAS	1.84E-05	0.16334	***
BE	0.000112	0.21202	***
TO	1.00E-04	0.19156	***
AND	9.91E-14	0.08828	***
ALL	0.000691	0.19449	***
OF	3.13E-07	0.11414	***
HERE	0.000375	0.24317	***
RIGHT	0.000105	0.15756	***
ABOUT	1.18E-05	0.10723	***

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 small 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 quickly seemed to point 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 within a predicative construction cannot by definition occur sentence-initially but must occur sentence-medially, we decided to narrow the scope of attention to occurrences of *well* preceded and followed by a comma. Searching the TIME corpus for *, well ,* we obtained 695 hits and subjected them to concordance analyses. We present the results of these analyses below.

3.3 Concordance analyses of sentence-medial *well*

Close reading of the 695 hits suggested that three major syntactic patterns were prevalent in the data: (i) *well* introducing direct speech (hereafter 'quote-*well*'), as in (1), (ii) *well* between clauses and/or constituents ('clause-*well*'), as in (2), and *well* as an element within the subject or object predicative ('predicative-*well*'), as in (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% of all 695 sentence-medial hits, with quote-initial use making up 11%, clause-*well* 22% and predicative-*well* taking up 29% of the hits total.

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

	Hits	**% (out of 695)**
Quote-*well*	79	11
Clause-*well*	152	22
Predicative-*well*	198	29
Others	266	38
Total	695	100

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 79 hits for quote-*well* nicely reflects the 2-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 5:

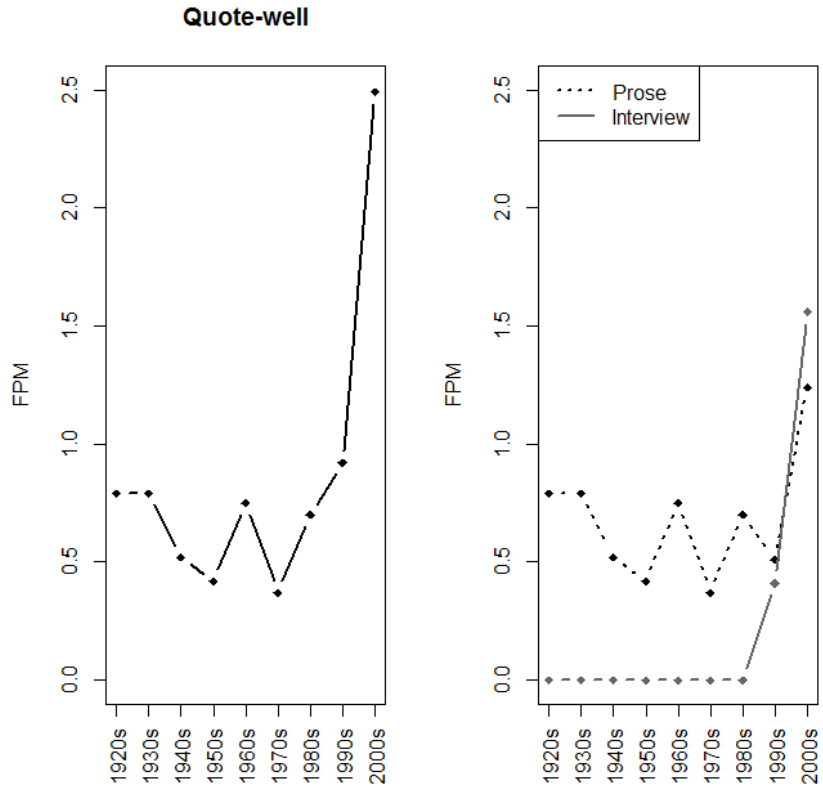

Figure 5: Frequencies-per million words (FPMs) for quote-*well* per decade (left panel) and FPMs for quote-*well* in prose and interviews respectively (right panel)

An intriguing question relates to what *types* of quotation *well* is used with. Concordance analysis revealed that two such types were predominant[iii], illustrated in the following two examples. Example (4) shows *well* in the context of interviews, 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."?). Much more likely, the 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.

(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.

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 5, 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. In a few cases, *well* conjoins a fronted object and a main clause. Consider (6)-(10), where clauses and relevant constituents are annotated:

mainCl-*well*-mainCl:

(6) [^{mainCl} Then M.J. left] and **, well ,** [^{mainCl} [^{Subject}the team] fell apart faster than a Rodman romance....] (TIME 1999)

mainCl-*well*-AdvCl:

(7) [^{mainCl} It's so tempting to give up that distance prematurely], [^{AdvCL} because **, well ,** [^{Subject} there] is so little distance after you put the Map of the States puzzle (…)](TIME 1996)

AdvPh-*well*-mainCl:

(8) [^{AdvPh} Given where we are with the peace process -- we're so close to a real resolution after so many decades of hate -- and given how much we have invested in Peres' being able to continue it on the Israeli side] **, well ,** [^{mainCl} for us [^{Subject} it]'s as if Arafat were running himself]. (TIME 1990)

AdvCl-*well*-mainCl:

(9) [^{AdvCl} once you've seen Walking with Dinosaurs (Discovery, April 16, 7 p.m. E.T.)] **, well ,** [^{mainCl} [^{Subject} you] still won't have seen real animals do any of that]. (TIME 2000)

Obj-*well*-mainCl:

(10) [^{Obj} To the boomers' children **, well ,** [^{mainCl} [^{Subject} it]'s a philosophical conundrum] (TIME 1998)

Table 4: Frequencies of syntactic patterns involving clause-*well*

Pattern	**Hits**	**%**
AdvCl-*well*-mainCl	96	62
AdvPh-*well*-mainCl	25	16
main Cl-*well*-AdvCl	18	12
mainCl-*well*-mainCl	13	8
Obj-*well*-mainCl	3	2

Total	155	100

As can be seen from Table 4, 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 (8) above). This pattern alone accounts for 62% of all clause-*well* instances. It is this pattern that we will focus on in the following.

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 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). Upon inspection of the *if*-clauses in the clause-*well* subset, their lengths were occasionally remarkable. Two such lengthy examples are (11) and (12):

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

(12) **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 13 words and 31 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 corpus's decades. 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 amount of occurrences as were found in the subset for *if*-clauses with *well*, viz. 87 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. The tests suggest that *if*-clauses without *well* are not correlated ($p = 0.94$, $tau = 0.006$), whereas *if*-clauses with *well* do show a weak but highly significant positive trend ($p = 0.007$, $tau = 0.217$). The trend can be observed in the boxplots shown in Figure 6:

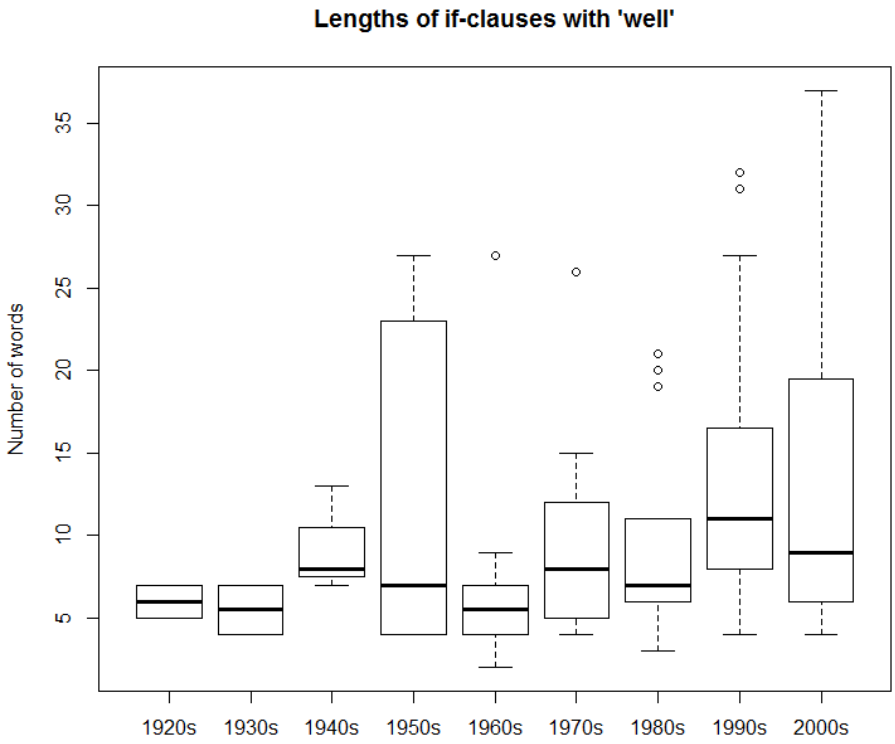

Figure 6: Boxplots of lengths of *if*-clauses with *well* per decade

Despite occasional dips, the median lengths per decade depicted in the bold horizontal lines across the boxes show an overall upward trend as you 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 further down.

3.3.3 Predicative-*well*

We mentioned earlier that the use of *well* at the onset of quotation (quote-*well*) is a widely discussed research topic. The third major category we found in our data – predicative-*well* –, by contrast, has, to our knowledge, not yet been described anywhere. With 198 examples identified, this subset is the largest in our data and may be considered 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 involved in 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') (cf. Quirk et al. 1985: 1171). The three sub-categories are illustrated in (13)-(15):

Subject predicative:
(13) After all, [$^{\text{Subject}}$ Soviets] [$^{\text{Copula}}$ are] **, well ,** [$^{\text{Subj.Predicative}}$ Soviets]. (Time 1987)

Object predicative:
(14) (…) reflects the interests of the citizens, [$^{\text{Subject}}$ who] [$^{\text{Copula}}$ consider] [$^{\text{Object}}$ themselves] **, well ,** [$^{\text{Obj.Predicative}}$ unique]. (TIME 2002)

Adverbial predicative:

(15) (…) booming baritone [$^{\text{Subject}}$ that], on his five bestselling albums, [$^{\text{Copula}}$ sounds] vaguely [$^{\text{Adv.Complement}}$ like **, well ,** a fellow hollering down a drainpipe].

Table 4 shows that, not surprisingly given the distributions in other contexts (cf. Quirk et al. 1985), the three types are very unequally distributed, with the subject predicative accounting for the lion share, viz. 89%, the object predicative for 9% and the adverbial complement for 3%.

Table 4: Frequencies of types of predicative-*well*

Predicative type	**Hits**	**%**
Subject	176	89
Object	17	9
Adverbial	5	3
Total	198	100

It is further 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% 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 involvement in the predicative construction, where copular BE is without any close rival.

Table 5: Distribution of copula verbs for predicative-*well*

Copula verb	**Hits**	**%**
BE	144	73
SEEM	11	9
BECOME	7	4
SOUND	5	3
Other verbs	31	16
Total	198	100

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 (16), *secret* in (17), *tax* in (18), and *healthy* in (19). Second, many co-texts of predicative-*well* are characterized by word play. For example, in (20), the author plays with the phonetic similarity of two completely distinct lexemes (*decade* and *decadent*); in (21), the author juxtaposes *opportunity* and *opportunism* thereby playing with the evaluative dissimilarity between them (opportunity considered 'good, opportunism 'bad'); in (22), 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 (23), a more successful pun, the conceptual association of (female) 'nanny' and 'household' is creatively exploited for the blend formation *manny* merging 'man' and 'nanny'.

(16) (…) making fun of yourself for being dumb is **, well ,** dumb . (TIME 1999)

(17) Until quite recently, cryptography -- the science of making and breaking secret codes -- was **, well ,** secret. (TIME 1994)

(18) The worst knock against a gas tax is that it is **, well ,** a tax. (TIME 2004)

(19) (…) are well stocked with healthy snack products, but they tend to taste distinctly **, well ,** healthy . (TIME 2003)

(20) After a decade, open tennis is becoming **, well ,** decadent . (TIME 1978)

(21) Why reach for this opportunity? One possible explanation is **, well ,** simple opportunism. (TIME 1993)

(22) (…) and if anything, Cunningham has only got more audacious and more **, well ,** cunning in the past six years. (TIME 2004)

(23) [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 of inserts showing a positive trend in the TIME Corpus. It turned out that indeed the large majority of the features investigated (8 of 10) were modestly strongly to strongly correlated with the decades thus increasing in frequency. Since inserts are key components of colloquial speech we certainly have a strong case of colloquialization here: inserts do make a contribution to writing in TIME magazine becoming more speech-like. A number of explanatory hypotheses have been proposed in recent research to account for the phenomenon of colloquialization. For colloquialization itself 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 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 7 of 10 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 20th century for English newspaper editorials. Investigating what motivates this accelerated colloquialization in the 1990s and, as our data suggest, the 2000s is far beyond our present aims, 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 email, tweets on Twitter and chats on Facebook doubtlessly "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; added emphasis), precisely the decade in which most of the most dramatic hikes in frequency were observed for inserts. Many people spend long hours using email and even longer hours on the more recent 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. 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, Norrick 2014). 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 typographic 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 (24):

(24) 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* most likely 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 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 corpus decades. 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*, termed 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' 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 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 (cf. Aijmer 2013). Obviously, magazine writers do not use *well* because they were at a loss for 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 sign-post 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.[iv]

Bluntly re-using lexical material in close vicinity to its first use is generally considered bad style and therefore avoided. Seen through this lense, *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 (cf. Crystal 1998) thus involving 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 counterevidence, to assume that predicative-*well* so far exists only in writing (specifically TIME magazine writing) but not in conversation. If this is the case, predicative-*well* represents 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-*well* is indeed not conversational and hence not colloquial, it becomes questionable whether it can be seen as an instance of colloquialization. Strictly speaking, predicative-*well* is colloquialized only inasmuch as it involves the colloquial marker *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-*well*, the adaptation is to a context which is syntactically highly complex and generically quite constrained, viz. TIME magazine writing.

5. Concluding remarks

To conclude this paper, this section spells out a number of implications of our empirical findings. Some of these are related to the over-arching theme of colloquialization, others relate to questions of methodology.

As regards colloquialization, our study suggests that there is far more colloquialization going on than has hitherto be observed. Not only contractions, personal pronouns, questions, progressives and zero relative clauses are on the rise in writing; inserts are making inroads too gaining accelerated momentum in the last two decades of the corpus (the 1990s and the 2000s). The close analysis of *well* indicated a number of subtle factors co-driving this rise. Analysis of quote-*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 *well* acts as a rhetorical device to inject direct-speech-like immediacy into the text. Moreover, examination of clause-*well* and predicative-*well* suggested that *well* is gaining in frequency because it is entering into new syntactic contexts, viz. the left-of-subject position in complex sentences and the predicative construction, and that, in these new contexts, it is taking on new functions, viz. 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 email and the massive increase in 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' (cf. Crystal 2001).

Methodologically, the study has implications on three levels. First, the study suggests that, in investigating colloquialized writing, it pays dividends to start with an examination of colloquial speech (instead of starting with writing itself or previous research). Building on the keyness procedure a large bulk of (lexical) colloquial key items can readily be identified, catalogued, and examined as to their development over time. We are aware that using, as

we did, key items pulled from the British National Corpus and examining them (or their equivalents) in a corpus of American English is not without problems. It is doubtlessly 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 actually on-going changes in American English.

Second, the case study of *well* is a very strong indication that mere trend-spotting is at best a preliminary exercise and that much more, and more rewarding, insights can be gained from analysis of diachronic-data structure. As noted, however, such structure analyses require temporally continuous data, a condition perfectly met by the TIME Corpus. A desideratum for future diachronic research is therefore not only an update of the TIME Corpus to also include texts from the second decade of the 21st century but more importantly the creation of more such continuous diachronic corpora to cover a wider range of written genres.

Third, the analysis of *well* is a prime example of corpus pragmatic research integrating (rather than using in a side-by-side fashion) the key methodologies of corpus linguistics and pragmatics. Exemplifying the vertical corpus linguistic methodology, the investigation of the marker started with three successive quantitative analyses. First, a frequency count revealed a significant trend towards increased use of the marker across the nine decades. Further, to establish whether the trend is continuous or whether it develops in leaps, Variability-based Neighbor Clustering (VNC) was performed, an exploratory statistical method for hypothesis-generation. This analysis suggested the hypothesis that *well* develops in two stages, with a break between the 1980s and the 1990s. To test that hypothesis, the marker's collocational profiles in the two time spans are compared statistically. The comparison identified various forms of the verb BE as significant collocates of *well*. At that point, to not simply note but understand those collocation patterns, we changed the mode of our analysis from quantitative to qualitative, i.e. to reading and classifying all instances of *well* bracketed by commas in the corpus. This qualitative analysis discovered striking, and strikingly novel, patterns: *well* owes its rise in TIME magazine not only to an increase in quote-initial position, nor is it attributable only to increased use between clauses but it is also largely due to occurrence within predicative constructions with BE as the copula. In sum, the analysis of *well* in TIME demonstrates how corpus pragmatics can reap the benefits of both pragmatics and corpus linguistics. The quantitative corpus-linguistic analyses have indicated what to look out for; without that indication a merely qualitative analysis of *well* might have overlooked the marker's novel uses in the vast masses of data. On the other hand, the qualitative pragmatic analyses have unearthed the marker's new forms and functions; without them a merely quantitative analysis of *well* might have failed to find and appreciate what it contributes to the context and the writer/reader interaction. Much more of this fruitful marriage can be expected in future corpus pragmatic research.

References

Aijmer, K. 2013. *Understanding pragmatic markers. A variational pragmatic approach*. Edinburgh: Edinburgh University Press

Baker, P. 2009. 'The BE06 Corpus of British English and recent language change', *IJCL* 14/3: 312-337.

Biber, D. 2003. 'Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy', in J. Aitchison and D. M. Lewis (eds.) *New media language*, London: Routledge, pp.169-181.

Crystal, D. 1998. *Language play*. London: Penguin

Crystal, D. 2001. *Language and the internet*. Cambridge: Cambridge University Press

Davies, Mark. 2007-. *TIME Magazine Corpus: 100 million words, 1920s-2000s*. Available online at http://corpus.byu.edu/time/

Fairclough, N. 1995. *Media discourse.* London: Arnold.

Gries, S. Th. and S. Stoll. 2009. 'Finding developmental groups in acquisition data: Variability-based Neighbor Clustering', *Journal of Quantitative Linguistics* 16(3): 217-242

Hilpert, M. and S. Th. Gries. 2009. 'Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition', *Literary and Linguistic Computing* 24(4): 385-401

Hilpert, M. 2013*. Constructional change in English. Developments in allomorphy, word formation, and syntax*. CUP.

Hoffmann, S., S. Evert, N. Smith, D. Lee, and Y. Berglund Prytz, 2008. *Corpus linguistics with BNCweb – A practical guide.* Frankfurt am Main: Peter Lang

Kjellmer, Göran. 2009. 'Where do we backchannel? On the use of mm, mhm, uh huh and such like.' *International Journal of Corpus Linguistics* 14(1): 81–112.

Leech, G. 2003. 'Modals on the move: The English modal auxiliaries 1961-1991', in R. Fachinetti, F. R. Palmer and M. G. Krug (eds.) *Modality in contemporary English*, Berlin/NY: Mouton de Gruyter, pp. 223-240.

Leech, G., M. Hundt, C. Mair, and N. Smith. (2009). *Change in contemporary English. A grammatical study*. Cambridge: Cambridge University Press.

Levinson, S. C. 1983. *Pragmatics.* Cambridge: Cambridge University Press.

Millar, N. 2009. 'Modal verbs in TIME. Frequency changes 1923-2006', *IJCL* 14/2:191-220

Mindt, D. 2000. *An empirical grammar of the English verb system.* Berlin: Cornelsen

Rühlemann, C. 2007. *Conversation in context: A corpus-driven approach*. London: Continuum

Rühlemann, C. and K. Aijmer. 2014. 'Corpus pragmatics: Laying the foundations', in Aijmer, K. and C. Rühlemann (eds.) *Corpus pragmatics*. *A handbook.* Cambridge: Cambridge University Press, pp. 1-26

Scott, M. and C. Tribble. 2006. *Textual patterns. Key words and corpus analysis in language education*. Amsterdam/New York: John Benjamins

Taavitsainen, I., A. H. Jucker and J. Tuominen (eds) 2014. *Diachronic Corpus Pragmatics*, vol. 243 of Pragmatics & Beyond New Series, Amsterdam: John Benjamins.

Tognini Bonelli, E. 2010. 'The evolution of corpus linguistics,' in A. O'Keeffe and M. McCarthy (eds.) *The Routledge handbook of corpus linguistics*, London/New York: Routledge, pp. 14–27

Tottie, G. 1991. Conversational style in British and American English: The case of backchannels. In K. Aijmer & B. Altenberg (eds) *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 254-271). London: Longman

Tottie, G. 2014.. 'Turn management and the fillers uh and um', in Aijmer, K. and C. Rühlemann (eds.) *Corpus pragmatics*. *A handbook.* Cambridge: Cambridge University Press, pp. 381-407

Westin, I. 2002. *Language change in English newspaper editorials.* Amsterdam/New York: Rodopi

Westin, I. and C. Geisler. 2002. 'A multi-dimensional study of diachronic variation in British newspaper editorials', *ICAME Journal* 26: 115-134

[i] (*.|,|:|?|!|... well .|,|:|?|!|...*)

[ii] A search for *er* in the TIME Corpus returns 488 hits. Upon inspection of concordance lines, however, only 95 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)

[iii] Besides occurrences in interviews and prose, there was a single instance of quote-*well* within a haiku poem.

[iv] 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).