

Uh and *Um* as Pragmatic Markers in Dialogues: A Contrastive Perspective on the Functions of Planners in Fiction and Conversation

Ilenia Tonetti Tübben | ORCID: 0000-0002-9006-2808
Department of Languages and Literatures, Englisches Seminar,
University of Basel, Nadelberg 6, 4051 Basel, Switzerland
Corresponding author
ilenia.tuebben@unibas.ch

Daniela Landert | ORCID: 0000-0002-9214-9712
English Department, Heidelberg University, Kettengasse 12,
69117 Heidelberg, Germany
daniela.landert@as.uni-heidelberg.de

Abstract

The planners *uh* and *um* fulfil a range of interpersonal and discursive functions, akin to other pragmatic markers. Although there appear to be differences between the planners' functions across genres, the factors influencing these differences have not been analysed comprehensively so far. In our study, we analyse the frequency and functions of *uh* and *um* across three different genres: spontaneous conversation, scripted dialogues in television series and dialogues in improvised theatre. Our findings suggest that both the fictional nature of a text and its degree of composition have an influence on the planners' functions. Our results confirm that planners are indeed pragmatic markers which serve a range of textual and interpersonal functions, and that the frequency and functions of planners are affected by factors relating to the context of language production, which include spontaneity as well as the content and function of a text.

Keywords

planners – fiction – spontaneous conversation – improvised theatre – telecinematic discourse

1 Introduction

The planners *uh* and *um* – also called fillers, filled pauses and hesitation markers – are typically associated with spontaneously produced spoken language. A growing body of research has shown that they fulfil a range of functions relating to signalling trouble, planning, repair as well as turn management. Sociolinguistic studies have further demonstrated that planners show variation across social groups and, especially, variation across individual speakers. Recently, several studies have pointed out that planners are not restricted to spontaneous spoken interaction. For instance, Tottie (2019) studied the increase of planners in American journalistic writing and Jucker (2015a; 2015b) showed that planners are used in scripted fiction, where they can contribute to characterisation and signal interpersonal relationships between the fictional characters. These studies indicate that there are differences in the functions of planners across different genres. These differences are marked enough for Tottie to suggest that planners fall into different word classes in spoken and written English (2019: 128). However, approaches that compare the use of planners across different genres and registers have been rare so far. As a consequence, we know very little about the factors that influence the functions of planners.

In our study, we treat planners as pragmatic markers that can fulfil a range of interpersonal and textual functions. Our aim is to investigate how these functions depend on the conditions under which texts are produced. More specifically, we are interested in the role of two dimensions that can influence the use of planners: spontaneous/scripted and fiction/non-fiction. So far, linguistic studies have not been able to differentiate between these two dimensions. Comparisons between dialogues in fiction and non-fiction tend to compare spontaneous non-fictional conversation with scripted fictional dialogues (e.g. Bednarek 2012; Quaglio 2009). Our approach sheds new light on how the dimensions spontaneous/scripted and fiction/non-fiction affect the frequency and functions of pragmatic markers.

As previous research has shown, both spontaneous spoken conversation and scripted fictional writing include planners, but it appears that there are differences with respect to the frequency and function of planners. While planners are common in spontaneous conversation, they tend to be used as marked stylistic features in written fiction. We want to investigate these differences in more detail. For this purpose, we analyse the frequency and function of *uh* and *um* across three different sets of data: Spontaneous conversation, scripted dialogues in television series and dialogues in improvised theatre. Spontaneous conversation and scripted fiction have both been used in studies of planners, albeit not in a contrastive way. By including improvised theatre as

a third set of data that consists of spontaneously produced fiction we are able to differentiate between the dimension of fiction/non-fiction and the dimension of spontaneous/scripted. Thus, we can see how spontaneity and fiction affect the frequency and function of planners.

Our analysis is based on three sets of transcriptions. For spontaneous conversation, we rely on transcriptions from the Santa Barbara Corpus of Spoken American English. For the other two data sets, we use transcriptions that we compiled ourselves, based on pilot episodes of TV series and recordings of improvised theatre performances. We analyse all planners in the transcripts with respect to their frequency and function, distinguishing between four different functional categories which were identified by combining a bottom-up approach with categories proposed in previous research, namely cognitive, repair, discourse management and interpersonal. The functional classification is based on a detailed manual analysis of all instances in context, which also makes use of the audio recordings of the Santa Barbara Corpus and video recordings of the other two data sets to take into account audiovisual cues. The categories were tested for their stability by applying intercoder reliability testing.

By comparing the overall frequency of planners and the distribution of the four functional categories across the three data sets, we are able to observe how the use of planners differs in fictional dialogues compared to non-fictional dialogues and how it differs in scripted dialogues compared to spontaneous dialogues. This analysis also helps clarify to what extent the use of planners in scripted texts is related to the use of planners in spontaneous interaction. Are they part of the same linguistic phenomenon? Or are scripted and spontaneous planners, as Tottie (2019: 128) suggests, independent to such an extent that they use warrant separate word classes?

Section 2 provides relevant theoretical background on planners and their status as pragmatic markers. We then introduce our data and overall approach in Section 3. In Section 4, we provide a detailed discussion of all four functions that we identified, including the criteria we used for the classification. Our results concerning the overall frequency and the functions of planners across the three data sets are presented in Section 5. Finally, Section 6 presents our conclusion and outlook towards future research.

2 Planners as Pragmatic Markers

The planners *uh* and *um*, which we use here as including also the British English variants *er* and *erm*, have long been considered signals of speech disfluency. While many style guides and folk opinion continue to advise against

their use (cf. Erard 2007; Fox Tree 2007: 304), linguists have long shown that they fulfil a large range of functions and that “they are most of the time guiding and lubricating elements that facilitate communication” (Kjellmer 2003: 191). Planners have also been studied from the perspective of sociolinguistics (Tottie 2011; Stenström 2012), and have been analysed with respect to their distribution according to socio-demographic factors such as gender, age and socioeconomic parameters (Laserna et al., 2014; Staley and Jucker 2021). In other studies, planners have been looked at from the perspective of language learning and speech fluency (Crible 2017; Crible 2018; Götz 2013; Götz 2019). Finally, previous research has shown that planners fulfil all the criteria to be considered pragmatic markers or discourse markers (Tottie 2014b; Tottie 2016; Tottie 2019).

Formally, planners are inserts that are structurally independent from the sentence or utterance in which they are used (Biber et al. 1999: 1082–3). As Tottie (2014a) shows, they can occur in different positions within a turn, turn-initially, turn-medially, and turn-finally, i.e. they enjoy some freedom in their syntactic position. Norrick (2014) also points out that *uh* and *um* can constitute a fully-formed turn by themselves (Norrick 2014: 256). Like other pragmatic markers, their meaning is highly dependent on the surrounding context; or, as Norrick (2014) puts it, they “tend to have functions rather than meanings” (Norrick 2014: 249).

Functionally, planners are similar to many other pragmatic markers in that they can have both textual and interpersonal functions. With respect to textual functions, planners can be used for turn-management purposes (Clark and Fox Tree 2002; Kjellmer 2003; Tottie 2014a), to signal trouble in online speech production and to gain time to plan the next utterance (Clark and Fox Tree 2002; Kjellmer 2003; Jucker 2015b; Tottie 2016). The interpersonal functions of planners have been discussed in fewer studies. Staley and Jucker (2021) define a functional profile of *uh* and *um*, showing that the planners are not only used in their widely-described functions as signalling hesitation or managing repair, but also with a highlighting or a face-mitigating function. In our study, we include both types of functions, textual and interpersonal. As we will argue based on our results, the emphasis on textual functions in previous research may be a consequence of the type of data that has typically been analysed for the study of planners.

While most research on planners has focused on spontaneous conversation, some studies have been concerned with the analysis of planners in written texts, showing that planners are used also in written language and fiction, albeit more sparingly and with slightly different functions (Jucker 2015a; Jucker 2015b). For instance, in non-fictional writing planners can be used to

express stance and to show “the writer’s humorous, ironic or euphemistic attitude” (Tottie 2020), a case in point being journalistic prose (cf. Tottie 2019). In fiction, they can be used for characterization purposes, as Jucker (2015b) discusses in his analysis of the Corpus of Historical American English (COHA), as well as in his analysis of *The Hitchhiker’s Guide to the Galaxy* (Jucker 2015a). Jucker’s analysis of the texts in COHA also highlights other functions of *uh* and *um*, namely introducing an awkward topic or a lie (Jucker 2015b: 176). In addition to the aforementioned textual and interpersonal functions, *uh* and *um* in written texts can be used to “conjure up a sense of orality” (Jucker 2021: 16), i.e. they are used “to simulate or represent spoken language” (Staley and Jucker 2021: 23). As a consequence, they have been described as being part of the colloquialisation of written genres like journalistic writing (Rühlemann and Hilpert 2017; Tottie 2019).

So far, studies on planners have either concentrated on spontaneous conversation (Clark and Fox Tree 2002; Kjellmer 2003; Stenström 2012; Tottie 2014a; Tottie 2014b; Tottie 2016), scripted fictional dialogues (Jucker 2015a; Jucker 2015b), or journalistic writing (Tottie 2019). They have shown that there are differences with respect to the frequency and function of planners, but the differences have not been investigated thoroughly across different genres. The aim of our study is to expand the investigation to encompass also a genre which has received little attention, i.e. unscripted fiction in the form of improvised theatrical performances. Given that planners fulfil many functions relating to processing constraints of real-time language production, we can assume planners to be more frequent in spontaneous conversation than in scripted dialogues of television series. If the cognitive constraints of online language production are the main factor affecting the use of planners, then we can assume their frequency in improvised theatre to be similar to that in spontaneous conversation, since both are spontaneously produced. However, other factors are likely to play a role in the production of planners, too. Factors such as the overall topic and function of the interaction, as well as the dramatic content of fiction may affect the use of planners in dialogue. This would result in similarities in the use of planners across improvised theatre and scripted television series, which both present fictional interactions. Thus, by including dialogues from improvised theatre as a form of spontaneously produced fiction, we are able to observe the effects of two dimensions of text production in a comparative fashion: The dimension of fiction/non-fiction, which differentiates between spontaneous non-fictional conversation, on the one hand, and dialogues from improvised theatre and scripted television series, on the other; and the dimension of spontaneous/scripted, which differentiates between scripted television series, on the one hand, and spontaneous conversation and

improvised theatre, on the other. Comparing the use of planners across the three genres will show to what extent the functions of planners are related to either of these two dimensions.

3 Methodology and Data

In our study, we use data from spontaneous conversation, scripted fiction and spontaneous fiction. Table 1 below provides an overview of the data types and the number of files, words, and planners. For a more detailed overview, cf. Tables 3, 4 and 5 in the Appendix.

For spontaneous conversation, we rely on the Santa Barbara Corpus of Spoken American English. The corpus is freely available online and consists of audio recordings of naturally occurring spoken interaction and transcriptions thereof, for a total of approximately 249,000 words. The language is American English and the transcriptions were carried out applying the conventions of the Santa Barbara discourse transcription approach (Du Bois et al. 1992). Most of the spontaneous interactions represented are face-to-face interactions; however, the SBC corpus also includes telephone conversations, task-based interactions, classroom lectures, sermons, story-telling, and other types of spoken interaction. For the purpose of the present analysis and to enhance comparability with the other datasets, we decided to restrict our investigation to selected transcripts of spontaneous face-to-face interaction. We included ten transcripts of such spontaneous face-to-face interaction from the corpus in our study. Since our analysis of the fictional data retrieved many instances of planners in telephone conversations, we decided to also include one telephone conversation from the Santa Barbara Corpus, namely SBC052, a phone conversation between family members. The total amount of data that we used from the Santa Barbara Corpus consists of 11 recordings, amounting to 53,353 words and 275 minutes of conversation.

TABLE 1 Overview of the corpus. Files, number of words and number of planners

	Files	Words	Planners
Spontaneous conversation (SBC)	11	53,353	374
Scripted TV series	14	66,406	305
Improvised theatre	6	47,610	414
TOTAL	31	167,369	1,093

For scripted fiction, we rely on a self-compiled corpus of pilot episodes from TV series (cf. Table 5 in the Appendix), which we transcribed according to the transcription conventions used in the Santa Barbara Corpus, with some minor adjustments. This part of our data consists of 14 pilot episodes of recent TV series (2000 onwards), which amount to 66,406 words and 579 minutes of conversation. The genres represented are drama, comedy, and dramedy (i.e. a genre combining elements of comedy and drama). Excluded from our dataset were series set in past or future time periods, since this may have an influence on language use; additionally, we restricted our data to American English, as we did with the other datasets.

For spontaneous fiction, we use transcriptions of improvised theatre performances, which we transcribed ourselves. Improvised theatre is a theatrical form in which performers improvise scenes spontaneously. Much like traditional theatre, performers perform on stage in front of live audiences. As opposed to traditional theatre, though, there is no script and the performers lack time for planning and revision of the dialogues, as the performances are improvised. As Landert (2021) puts it, “the basics of improvised theatre is the spontaneous response by the performers to what happens in a scene, informed by the performers’ background knowledge and life experience” (Landert 2021: 70). Therefore, improvised theatre dialogues share characteristics with both TV dialogue and spontaneous conversation, in that it is fiction and, at the same time, it is produced spontaneously.¹ For our study, we transcribed a total of 6 performances amounting to 47,610 words and 320 minutes of conversation. The performances consist of shows by two duos of performers, “TJ and Dave” and “Middleditch and Schwartz”. All four performers are professional improvisers with many years of experience in improvisation, and the performances are all in American English. The recordings of the TJ and Dave shows, which were performed in Spring 2015 at the iO Theatre in Chicago, are commercially available on the video sharing platform Vimeo. The shows by Middleditch and Schwartz were released on the streaming platform Netflix. Both duos perform coherent shows that last between forty-five minutes and one hour.

Despite our efforts, the three data sets are not perfectly comparable in all respects. For instance, while we were able to achieve a more or less balanced representation of female and male protagonists for the scripted TV series, we only had access to improvised theatre performances of male performers. Investigating how gender affects planners in fictional dialogues is an aspect that we would like to explore further in the future, once we acquire suitable data from improvised theatre. Another limitation concerns the lack of video recordings of the spontaneous conversation data. Again, studying data from

¹ For a more in-depth discussion of the nature of improvised theatre see Landert (2021).

spontaneous conversation for which video recordings are available would be a welcome addition for future research.

In total, our data consist of 31 files, amounting to 167,369 words and 1,175 minutes of conversation, and they include 1,093 planners. All transcriptions, including those from the Santa Barbara Corpus, were converted to XML to be analysed with the help of the software Oxygen XML Editor. Our analysis proceeded in several steps. In a first step, all planners in the transcripts were automatically searched by looking for the respective forms and tagged as planners. Following this, we manually verified that all tagged instances were planners by listening to the audio and video recordings. In this step, we excluded some instances of affirmative markers and other sounds, which were transcribed as *uh* or *um*, but which were not planners. In a final step, we then classified each planner manually according to its function (cf. section 4). Again, we consulted the audio and video recordings to take into account paraverbal features, gestures and facial expressions. Since we did not have access to video material for the Santa Barbara Corpus, we had less supporting evidence available for this data set than for the other two data sets. This may result in a slight underreporting of functions in spontaneous conversation.

For the functional classification, we distinguish four different functions: cognitive, repair, discourse management and interpersonal (cf. Section 4). These categories were established in an iterative process, combining a bottom-up approach with insights from previous literature. In a first step, we compiled an overview of the planners' range of potential functions based on previous studies. This broad overview formed the basis for conducting a structured observation of the planners in our data. Our aim was to define a small set of functions that account for most instances of planners in our data, and that can be identified reliably, based on clear classification criteria. These classification criteria, which we present in Section 4 below, combine the interpretation of the function of the planner in context with contextual cues, such as other linguistic as well as audiovisual elements which tend to co-occur in the context of the planners' utterances. For instance, contextual cues that would indicate an effort of cognitive processing would be the co-occurrence of the planners with vagueness markers (e.g.: *something, kind of*) and upward gaze.

The functions we identify are not mutually exclusive and our classification allows for the presence of multiple functions for a single planner. The fact that planners, like many other pragmatic markers, can have different functions at once has been noted before (cf. Staley and Jucker 2021, Tottie 2019). For this reason, our analysis allows for multiple classifications, i.e. one and the same element can be classified as fulfilling more than one function at the same time.

We established the stability of our categories by applying intercoder reliability testing. The tests were carried out in several rounds by the two authors, with

adjustments to the classification scheme in-between rounds. Classifications were carried out independently and each round of classification was based on a fresh data sample. Since we allowed for multifunctionality of markers, we classified for each function independently, whether it was present or absent in a marker. Once the agreement per category reached 80%, the categorization was deemed stable. In the final round of coding, we were able to reach a minimum agreement of 80% for each category, although the agreement was much higher for most of the categories.

4 Functions of Planners

In this section, we present the four functions that we applied in our analysis. We include the criteria that we used to identify the functions and we present and discuss selected examples for each function. For space reasons, we are not able to present examples from all three sets of data for each function, but it is important to note that we observed all functions in all three data sets. In addition, as mentioned already, a given planner can have more than one function and, as a consequence, we allowed for multiple classifications in our analysis.

4.1 *Cognitive Function*

The cognitive function is one of the main functions generally associated with planners, especially in spontaneous conversation; it indicates that a speaker needs time to plan the next utterance (Kjellmer 2003; Tottie 2016). Here, we include all uses that occur in a context in which there is clear supporting evidence that the speaker is retrieving or processing information. This evidence can be presented in different forms, including verbally or in the form of gestures, facial expressions and gaze.

There are two different ways in which planners with a cognitive function can be used. They can either be an effect of actual cognitive processing or they can be intentionally used to stage cognitive processing. We can assume that most instances of cognitive functions in spontaneous conversation are instances of the former, whereas most instances in scripted fiction are instances of the latter. However, this distinction is not absolute. Actors in scripted fiction may use planners in a situation in which they are trying to recall information and it is possible that such instances survive post-editing if they are short enough to be unobtrusive or if they are deemed to fit the scene. Perhaps even more common are instances of staged cognitive functions in spontaneous conversation, for instance in connection with politeness functions or to create humour. It is not possible to distinguish genuine and staged cognitive functions based on recordings – one would need to have access to neurological information for

this – and, thus, we do not attempt to make such a distinction. Instead, we include all instances that demonstrate verbal or non-verbal evidence of cognitive processing.

There are a number of recurring patterns that we observed across our data set in which planners with a cognitive function typically occur. For instance, planners with a cognitive function tend to co-occur with other vagueness markers (*sort of, kind of, like*), pauses, explicit comments on the speaker's difficulty in retrieving a lexical item (*what is it called?*), as well as with non-verbal cues such as upward gaze. Syntactically, planners bearing a cognitive function often occur in structures of the type DET + *planner* + Adj/Noun, a pattern that is also observed by Staley and Jucker (2021: 31). Typical local contexts in which planners with a cognitive function occur are before specific information, especially names, specialised vocabulary, idiomatic expressions, and precise amounts; when referring to past memories; deciding among different options; when processing complex or new, surprising information; when reading out loud some type of information (e.g. reading from a document); and when listing items.

Example (1) illustrates the cognitive function with a passage in which the speaker retrieves specific numerical information from memory.

- (1) HAROLD: Have you heard these figures.
 that like=
 um,
 it's something like forty percent of males,
 in ... the Bay Area,
 are supposed to be infected?
 (Spontaneous conversation, SBC002)

In this example, the speaker (HAROLD) is making an effort to recall the precise percentage of people infected in a specific area. While thinking, he utters the planner *um*, which is then followed by a pause. After that, he is able to retrieve the information (*forty percent*). Notice that the percentage is introduced by a vagueness marker (*something like*), which is indicative of the tentative nature of the answer.

Similarly, planners can be uttered while retrieving specialised vocabulary, as in (2). In this example, the speaker (HANK) is trying to recall the precise colour term to define the colour of a building. Pointing to the speaker's cognitive effort are the unfilled pauses as well as the explicit remark *i don't know how to call it*, which refers to his difficulties in retrieving the correct term. In this instance the speaker ends up using a basic colour term, *green*, but following this passage his interlocutor corrects him, providing the colour term *sage*.

- (2) HANK: it's the last house on the right,
 see it?
 (0.6)
 not the two-story one,
 the one next to it.
uh,
 kind of a --
 (0.5)
 i don't know what do you call it *um* --
 (0.9)
 green?
 (Scripted fiction, *Breaking Bad*)

Planners can also be used to process information, as in (3) below. Here the speaker (TJ01) is asked to take a decision; while thinking about what to do, he utters a planner. In this context, the planner serves as a linguistic device to enable TJ01 to gain more time to process what has just been said and to elaborate an appropriate response. The planner is also accompanied by the explicit comment *i don't know*, which can be considered an additional indication that the speaker needs time to come up with a response.

- (3) DA03: do you wanna like a lot of ginger in there then?
 TJ01: what's that?
 DA03: well looking for a lot of ginger in there?
 TJ01: *u=h*,
 i don't know maybe just a regular= --
 DA03: regular amount of ginger?
 TJ01: i think so.
 (Improvised theatre, *Ass In Your Hand*)

For the purposes of this study, we did not distinguish between different subtypes of the cognitive function, such as between retrieving and processing information, but included all of them into the general cognitive function category.

4.2 *Repair*

Another function identified in our corpus is that of repair, a function that was also identified in previous research (e.g. Clark and Fox Tree 2002; Kjellmer 2003; Stenström 2012). A planner functions as a marker of a repair when it indicates that “a more correct or more suitable word or phrase than the one(s) just said will follow” (Kjellmer 2003: 188). For the purposes of our analysis,

planners which occur in the immediate context of a repair, either initiating the repair or within the repair, are always classified as repairs. We include only self-initiated repairs (as opposed to other-initiated repairs), and repairs that were actually carried out, regardless of the type of repair: structural repairs and content-based repairs were both included. For practical reasons, we decided to exclude aborted repairs. Also not included in the repair category were pure hesitations without reformulations, as in (4) below.

- (4) ALINA: % one of the things that they're doing,
 ... is *um*,
 ... (HX) (TSK) is painting this building on Melrose,
 (Spontaneous conversation, SBC006)

The following excerpt (5) shows a planner used to introduce a self-initiated structural repair.

- (5) DARLENE: ... Oh.
 ... See,
 (H) something was wrong,
 you need to call the telephone company then.
 Because,
 I,
 (H)= either me or ans-
 uh ~Jenn answered it about five times.
 (Spontaneous conversation, SBC052)

Here, the speaker (DARLENE) stops her utterance mid-word (*ans-*), then resumes talking and introduces a new word (*Jenn*). Thus, DARLENE is not merely stuttering, nor repeating the word *answered* after e.g., a false start. She is reformulating her thoughts and structuring her discourse differently on a syntactic-structural level. Additionally, planners can be in the context of self-initiated, content-based repairs as in the following excerpt (6).

- (6) JAKE: i'm doing the best speech from donnie brasco.
 (0.4)
 or *um*,
 actually,
 (1.0)

ten of me are doing the best speech from donnie & brasco.

(Scripted fiction, *Brooklyn Nine-Nine*)

In (6), the speaker (JAKE) is making a claim about his speech (*i'm doing the best speech from donnie brasco*), which he then corrects himself (*ten of me are doing the best speech from donnie brasco*). In this case, the speaker uses the planner *um* to introduce a repair which refers to the content of the proposition, and not to the structure of the sentence.

4.3 *Discourse Management*

Another category that we included in our analysis is that of discourse management. This is, perhaps, the most heterogeneous group of functions in our study. For the purposes of our analysis, we applied the discourse management function to all instances of planners that either signal information structure or that help negotiate the floor. We believe that these functions could be further differentiated in future studies. For the purposes of the present paper, we decided to focus mainly on comparisons between cognitive and interpersonal functions, which is why we did not attempt a more fine-grained differentiation with respect to discourse management.

Typical instances of planners used to manage discourse include planners uttered to direct the attention of the other conversation participants towards oneself and what is about to be said. This could happen when the speaker starts speaking during audience laughter in improvised theatre, or when the interlocutor appears to be distracted, as in excerpt (7) below.

(7) ((Han Lee walks towards the kitchen))

OMAX: wait.

wait *um um um*,

(1.5)

don't hire a new waitress,

okay?

(Scripted fiction, *Two Broke Girls*)

In the above excerpt from scripted fiction, the speaker (OMAX) is trying to get Han Lee, her boss, to pay attention to her and what she is about to ask him (*don't hire a new waitress*). In this case, the other speaker is about to leave. OMAX is introducing her attempt to attract and direct her boss's attention by uttering the imperative *wait* followed by a series of planners (*um um um*).

Planners with a discourse management function can also be used in overlaps, when a speaker wants to keep the other conversation participants from gaining the floor, when the speaker wants to take over the floor, as well as when a speaker wants to cede the floor to another conversation participant. Excerpt (8) below shows a planner used for turn-management to keep the floor (floor holding function).

- (8) SHARON: ... @@@@
 CAROLYN: So they don't know what the hell they're doing.
 [<Q *uh*],
 SHARON: [<Q *pi*=]?
 CAROLYN: (H) [2 Why don't we go out and run some laps Q> 2].
 SHARON: [2 I didn't bring any pie with me today 2].
 (Spontaneous conversation, SBC004)

In excerpt (8), the speaker (CAROLYN) is uttering a planner (*uh*) to signal to the other conversation participant (SHARON) that she intends to continue speaking, i.e. she uses *uh* in a floor-holding function.

Additionally, discourse-management planners can be used to change the topic of the conversation, as in (9) below.

- (9) TJ03: hi,
 how are you.
 DA06: hi.
 i'm okay,
 how are you?
 TJ03: i'm okay,
 i'm okay.
 u=h,
 i need,
 [...]
 a bag of *v=enison*,
 (Improvised theatre, *Toute Suite*)

Here, the speaker (TJ03), who is in a store trying to buy some dog food, is using the planner *u=h* in the context of a topic transition from a greeting sequence (*how are you*) to the reason of his visit to the store (*i need, [...] a bag of venison*).

An additional use that we included into the discourse management function concerns planners that act as quotative markers, i.e. which introduce reported speech, as in (10) below.

- (10) ALICE: ... He said *um*,
 ... ~Mandy had to stay up all by herself and
 & decorate
 the tree.

(Spontaneous conversation, SBC007)

Additionally, planners in a discourse management function are also used at the beginning of turns in phone calls, which we treat as a special case of gaining the attention of the other conversation participant. They are a discourse management device in that they help structure telephone conversations, as the following excerpt from improvised theatre shows (11).

- (11) DA03: *u=h*,
 we'd like to extend our stay by up to two days,
 [...]
 but our flights have been canceled,
 a=nd they're talking about a couple days.

(0.2)

u=m,

also we'd like to know if you= have any= --

(Improvised theatre, *Toute Suite*)

In (11), the speaker (DA03) staying at a hotel. He is calling the reception desk to inquire about dog food. During the telephone conversation, DA03 utters multiple planners. Here the speaker is using *u=h* and *u=m* to structure discourse in the context of a telephone call. More precisely, he uses the planners to introduce the reasons for his call (*we'd like to*). While example (11) is from an improvised theatre performance, in which the phone conversation is staged, we found similar instances in the phone conversation from the *Santa Barbara Corpus*.

Finally, we were able to identify another context of use for planners with a discourse management function, which relates to information structure. The planners *uh* and *um* can be used in storytelling, where they are often used to mark transitions between story parts, introduce elements of narrative sequences, and to open and close asides. The typical syntactic environment of the planners in storytelling is "*And + planner*" / "*Then + planner*". The function of structuring narration in storytelling is presented in (12) below.

- (12) MARILYN: I had a great time.
 (H) And *uh*,
 ... we're pulling up,

... and I see this gir=l.
 Who I'd never seen before,
 sort of d = art out of our driveway.

(Spontaneous conversation, SBC003)

In excerpt (12), the planner *uh* is used to introduce the next event in a series of events related to the telling of a story. The speaker (MARILYN) is telling her friends how she discovered an unauthorized person on her property one day after getting home. After a brief diversion, she resumes the narration of the plot leading up to the incident at the centre of her narrative.

4.4 *Interpersonal Function*

In addition to the functions discussed in the previous sections, planners can also be used with interpersonal functions. For the purposes of our analysis, the interpersonal function applies to all instances of planners which help negotiate the face wants of the interlocutors. Often, this occurs in the context of face-threats, and other face-threat mitigating devices sometimes co-occur with the planner. Some interpersonal functions have already been identified, although primarily in the context of scripted fiction (Jucker 2015a; Jucker 2015b) and task-based interaction (Staley and Jucker 2021). We identified a range of contexts in which planners with an interpersonal function are used. They include: introducing sensitive or potentially offensive or taboo topics; providing dispreferred responses; contradicting someone; announcing unwelcome news; paying compliments and responding to compliments; declining offers; introducing lies; introducing confessions and admissions; and making face-threatening comments about the addressee or a third party.

Examples (13) to (15) illustrate the range of uses that we classified as interpersonal in our analysis. Excerpt (13) shows a planner uttered to mitigate face threats in the context of a request and its refusal.

- (13) WALT: hey listen can you do me a favor,
 can you just *uh* --
 (0.8)
 drop me off at a corner (.) somewhere.
 (1.0)
 PARA: *u=h* no,
 (0.4)
 sorry.
 (Scripted fiction, *Breaking Bad*)

In (13), the first speaker (WALT) is being transported to the hospital; he is in an ambulance with a paramedic (PARA). WALT, who wants to avoid being taken to the hospital, tries to convince the paramedic that he is doing well and asks him to let him go (*can you just uh / drop me off at a corner somewhere*). Both speakers are using planners, and the planners are used to mitigate face threats. The face threats are of different nature: while WALT is formulating a request that threatens the addressee's negative face, the second speaker is refusing to comply with the request, i.e. he is using *u=h* to introduce a dispreferred response.

Planners can also be used to introduce a sensitive topic, as in (14).

- (14) SZMA: you guys went out?
 did you --
 i mean sixth grade you didn't --
 MDSP: yeah.
 SZMA: did you kiss or anything?
 MDSP: *u=m*,
 (0.7)
 we made love.
 (Improvised theatre, *Parking Lot Wedding*)

Here the two fictional characters are at a wedding. In this excerpt, the first speaker (SZMA) is inquiring about the past love relationship between the second speaker (MDSP) and the bride-to-be. MDSP and the bride-to-be used to love each other when they were children. Here, the speaker is admitting something embarrassing and potentially taboo, namely having had a sexual relationship with the bride-to-be at a very young age.

Finally, planners can also be used to introduce lies, as in excerpt (15).

- (15) SKYL: how was your day?
 (2.0)
 WALT: oh i don't know.
 (1.9)
 i don't know,
 it was (.) *um* --
 (3.2)
 fine.
 (Scripted fiction, *Breaking Bad*)

In this excerpt from the TV series *Breaking Bad*, SKYL and WALT are a married couple. WALT has just discovered that he is terminally ill. In fact, he has just come home from the hospital when his wife, SKYL, inquires about his day. WALT, who does not want to disclose the diagnosis to SKYL and is visibly upset, lies about how his day was (*it was um / fine*). In this excerpt, the planner is used to introduce the predicate *fine*. The function of introducing a lie was restricted to our fictional data, partly because we usually lack the information to assess whether non-fictional utterances in the Santa Barbara Corpus include lies, unless this is revealed or discussed. In contrast, in fiction we found that planners are used to signal to the audience that what is about to be said is not true while at the same time, on the intradiegetic level of communications, the characters are not aware of that.

5 Frequency and Function of Planners in Conversation and Fiction

Figure 1 presents the overall frequency of planners across the three sets of data. As expected, planners are significantly more frequent in spontaneous conversation (70.1 instances per 10,000 words) than in scripted TV series (45.9 instances per 10,000 words), with a p-value < 0.001.² However, the difference is perhaps smaller than one might have assumed. Given that planners are often interpreted as a consequence of online planning problems, one could have expected them to be far less frequent in scripted TV dialogues. There are two possible interpretations of the relatively high number of planners in scripted fiction. First, it could be due to attempts at representing dialogue in a naturalistic way (cf. Richardson 2010: 78). This would mean that scriptwriters and actors use planners intentionally in order to mimic the communicative behaviour of spontaneous conversation. If this were the main factor contributing to the presence of planners, we would expect the relative distribution of the functions of planners to be similar to those in spontaneous conversation. Alternatively, the rather high frequency of planners in scripted TV series could be related to a higher frequency of planners with functions that are typical of fiction. In this case, scripted fiction should have a functional distribution that is different from that of spontaneous conversation. A more detailed analysis of the frequency of different functions, provided below, will shed more light on this.

² The complete results of significance testing are presented in Table 6 in the Appendix.

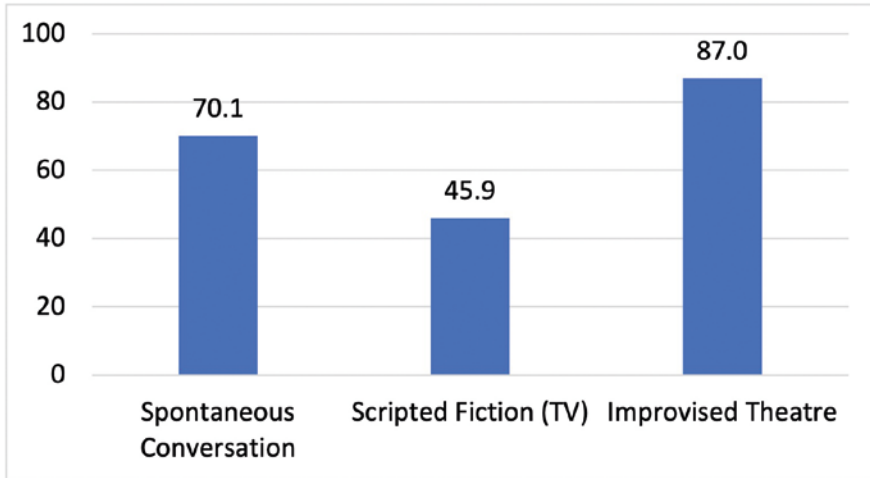


FIGURE 1 Overall frequency of planners across the three data sets, normalised per 10,000 words

As to improvised theatre, our initial assumption was that this kind of data would include a similar number of planners as spontaneous conversation, given that language is produced spontaneously in both sets of data. Our results indicate that planners are slightly more frequent in improvised theatre than in spontaneous conversation, although this difference is statistically significant only with a p -value < 0.01 . Again, a closer look at the distribution of the different functions of planners will provide more information on similarities and differences between the planners in improvised theatre and the other two sets of data. As far as overall frequencies are concerned, our results suggest that planners are more frequent in spontaneous dialogues – both fictional and non-fictional – than in scripted fictional dialogues.

Looking at the distribution of the different functions provides some additional insight into the variation across the three sets of data. Table 2 presents absolute and normalised frequencies of all four functions, and Figure 2 presents a visualisation of these results. Note that because some planners were assigned multiple functions, the sum of all functions is higher than the overall number of planners.

If we compare spontaneous conversation and scripted TV dialogues, we can see that the latter include a lower frequency of planners with cognitive, repair and discourse management functions (significant with p -value < 0.001).³ Thus,

³ For complete results of the significance tests for the function of planners, see Tables 7 to 10 in the Appendix.

TABLE 2 Frequency of functions of planners across the three data sets, normalised per 10,000 words (and absolute)

	Word count	Cognitive		Repair		Discourse management		Interpersonal	
		norm.	(abs.)	norm.	(abs.)	norm.	(abs.)	norm.	(abs.)
Spontaneous Conversation	53,353	35.2	(188)	14.4	(77)	26.6	(142)	4.5	(24)
Scripted Fiction (TV)	66,406	15.8	(105)	2.4	(16)	12.6	(84)	24.7	(164)
Improvised Theatre	47,610	43.5	(207)	15.8	(75)	26.9	(128)	25.2	(120)

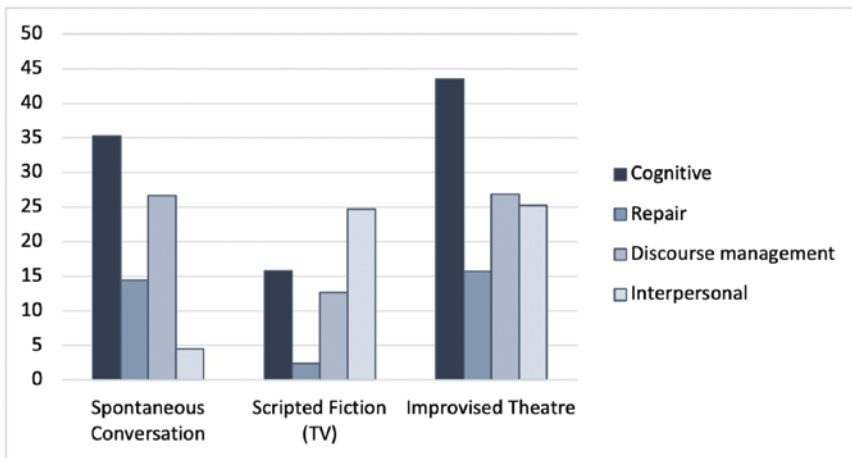


FIGURE 2 Distribution of functions across the three sets of data, normalised per 10,000 words

all functions that have a direct connection to online planning and cognitive constraints of real-time language production are less frequent in scripted fictional dialogues than in spontaneous conversation. This is in line with our initial assumption that planners are more frequent in spontaneous language due to planning problems that occur there. However, the lower frequency of functions related to planning in scripted TV dialogues is in part compensated by a much higher frequency of planners with an interpersonal function (significant with p -value < 0.001). This means that the surprisingly high number of planners in scripted fictional dialogues can, to a large extent, be attributed to the much higher presence of planners with an interpersonal function, compared to spontaneous conversation. This is most likely due to the dramatic

nature of fictional texts, which tend to include many instances of interpersonal conflicts and emotional moments that provide opportunities for using planners with interpersonal functions.

When comparing the results from improvised theatre to those of spontaneous conversation and scripted TV dialogues it appears that, to some extent, the functions of the other two sets of data are combined in improvised theatre. The frequencies of planners with cognitive, repair and discourse management functions are comparable to those of spontaneous conversation (no statistically significant difference at $p < 0.01$) and the frequency of planners with interpersonal function in improvised theatre is comparable to that in scripted TV dialogues (again, no statistically significant difference at $p < 0.01$). This would suggest that the two groups of functions – cognitive, repair and discourse management, on the one hand, and interpersonal, on the other – behave independently from each other. While planners with functions from the first group increase in spontaneous settings, planners with an interpersonal function increase in fictional settings.

As intriguing as these results are, there are certain caveats that should be noted. Most importantly, the results are based on the total frequency of the function of planners across the entire data sets. They do not reflect the fact that the variation within each data set is considerable. For instance, the overall frequency of planners per file varies for spontaneous conversation between 18 and 205 instances per 10,000 words, for TV series between 23 and 103 instances per 10,000 words, and for improvised theatre between 56 and 114 instances per 10,000 words. Thus, the variation within each data set is much higher than the differences between the data sets. This is very much in line with previous findings, which have shown that differences between individual speakers often exceed group differences based on age, gender or social groups (Tottie 2014b; Staley and Jucker 2021). The large variation across individual files means that the sampling of the data sets has a large effect on the results. Since we did not have access to balanced data sets for all types of data, we are not able to investigate social differences in more detail in this study. Expanding the overall amount of data and adding data that is currently missing – such as data from female performers of improvised theatre – will be important in future research to assess the stability of the effects that we observed in our results.

Despite these limitations, our results suggest that there are two factors which, independently, affect the use of planners. On the one hand, spontaneity leads to a higher frequency of planners with cognitive, repair and discourse management functions. These three functions are much more frequent in the two spontaneous data sets than in the scripted TV dialogues. The fact that similar uses still occur in scripted dialogues is probably due to a large extent to attempts at making the scripted dialogues appear naturalistic. In addition,

such instances can contribute to characterisation, for instance by making characters appear hesitant, evasive or lacking appropriate vocabulary. Both characterisation uses and uses which contribute to a naturalistic presentation of dialogue may be intentional and can be included either in the script produced by the screenwriters or added by the actors. If this is the case, we are dealing with staged instances of spontaneity functions. However, it cannot be ruled out that some instances of planners with cognitive, repair and discourse management functions in scripted TV dialogues are produced by actors unintentionally, for instance as a consequence of retrieving lines from memory or in self-repairs. Given that the TV series we included in our sample are expensive productions that must have undergone a very high amount of post-editing, we can assume that most (unintentional) production errors and self-corrections by actors were removed from the final product, but it is still possible that some planners in our data were produced in this way.

The second factor that affects the use of planners in our data is the dimension fiction/non-fiction. Here, fictional texts seem to have a higher frequency of planners with interpersonal function. We explain this difference with the dramatic nature of fictional texts. Compared to the kind of non-fictional spontaneous everyday interactions that are represented in the Santa Barbara Corpus, fictional texts are characterised by a higher density of dramatic conflict and interpersonal problems. Thus, the higher frequency of planners with interpersonal function can be seen as a consequence of the higher presence of interpersonal conflict and tension in fiction. This does not mean that similar planners do not occur in spontaneous conversation. The fact that they are very infrequent in our data from the Santa Barbara Corpus may, at least to some extent, be a consequence of the restrictions of data sampling for linguistic corpora. If we were able to study spontaneous conversations in highly dramatic and emotional situations, comparable to the kinds of situations the protagonists of TV series have to face, it is very possible that we would observe similar uses of interpersonal functions. Given that such data are rarely available for linguistic analysis, the study of fictional data can help identify uses that, so far, have received very little attention.

6 Conclusion

Our study aimed at providing new insight on the functions of the planners *uh* and *um*. Far from being merely signals of disfluency, our research has confirmed that planners can have a broad array of functions. Our results provide further support for the view that planners can be attributed the status of

pragmatic markers, given that they fulfil a range of textual and interpersonal functions typical of pragmatic markers. As pointed out in previous research (Kjellmer 2003; Tottie 2014a; Tottie 2014b; Tottie 2016), planners can be used to plan and repair utterances, and to structure turn-management. Our data also showed uses of planners with text-structuring functions that are related to storytelling. In addition, we found planners with an interpersonal function, especially instances that help mitigate face-threats. These uses are in line with results from previous research on scripted fiction and service encounters (Jucker 2015b; Jucker 2015a; Staley and Jucker 2021).

In our study, we compared planners across spontaneous conversation, scripted TV series and improvised theatre. Including improvised theatre as a form of spontaneous fiction allowed us to observe the effects of the two dimensions of spontaneous/scripted and fictional/non-fictional. Our analysis showed how, overall, planners are more frequent in spontaneous dialogues than in scripted dialogues. However, spontaneity affects not only the overall frequency, but also the distribution of the planners' functions. In spontaneous dialogues, planners are more frequently used with a cognitive, repair and discourse management function than in scripted dialogues. In contrast, planners in fictional texts differ from planners in non-fictional texts with respect to a higher frequency of interpersonal functions. This shows that the two dimensions, spontaneous/scripted and fiction/non-fiction, affect the use of planners in different ways.

As Tottie (2019) pointed out, scripted and spontaneous planners differ at times substantially in their functions. However, our findings suggest that the uses of planners in scripted texts and in spontaneous interaction are related. The fact that all four functions we included in our analysis can be found in all three sets of data indicates that the use of planners in spontaneous conversation and in scripted fiction is not fundamentally different. Rather, we found a difference in the functions' frequency, which is related to the content and the production conditions of these texts.

In our study, we looked at two dimensions of text production, spontaneous/scripted and fiction/non-fiction. However, there are many other dimensions that are likely to influence the use of planners. In addition to the spoken/written dimension, that has partly been discussed already, for instance in Tottie's study (2019), we encountered a number of promising further aspects, which we have not been able to investigate in detail. One of these aspects relates to mediated communication. When analysing the planners from our fictional data sets, we noticed how prominently phone conversations featured as contexts in which planners were used. The planners in phone conversations of both fictional and non-fictional data showed some indications that planners

may serve functions that are specific to this context. This would certainly deserve more attention, for instance in the form of a study that compares the use of planners across a range of different mediated and non-mediated settings. Another dimension that would deserve further attention is the number of interactants that are engaged in a conversation. Given that some of the functions of planners relate to floor negotiation, we can expect that the number of participants in an interaction can affect how planners are used. We saw some evidence of this in our data, although we did not investigate this aspect in detail. Again, this topic would deserve a separate study.

As we mention in Section 3, there are some limitations to our data sets. We did not have access to a corpus of spoken conversation that includes video recordings and, thus, we had to rely on audio recordings for classifying the functions of planners in this data set. In addition, our data did not allow us to look into gender variation, since we only had data from improvised theatre of male performers. Future research may be able to shed more light on the role of gender variation and, especially, individual variation, which was found to be a strong factor in the frequency of planners in spontaneous conversation (Tottie 2011; Staley and Jucker 2021). Data from improvised theatre has hardly been explored so far from a linguistic perspective. Our study shows that it can add new insight to research on spontaneous conversation and fictional dialogues. Future avenues of research include the variation within improvised theatre, for instance across different performers, different levels of experience and different styles.

Our results point to the need for further research that looks at planners from a contrastive perspective. So far, most research has studied planners in one type of data, typically in spontaneous conversation. This has resulted in a strong focus on functions related to online processing, such as signalling delays and repairs. In contrast, the interpersonal functions of planners have received very little attention so far. While they have been pointed out in a few studies on scripted fiction (Jucker 2015a; Jucker 2015b), service encounters (Staley and Jucker 2021), and journalistic prose (Tottie 2019), we know very little about interpersonal uses of planners in spontaneous conversation. As we discussed above, this may in part be a consequence of the sampling procedures of spontaneous conversation, which lead to few examples of dramatic scenes and interpersonal conflicts in corpora of spontaneous spoken language. However, our sample from the Santa Barbara Corpus included instances of planners with an interpersonal function, so there are certainly aspects to discuss here. Including a contrastive perspective in which data from different contexts are compared may help identify other, previously neglected aspects of planners.

In a broader perspective, our study indicates that different dimensions of text production – spontaneous/scripted and fiction/non-fiction – can affect the use of planners in different ways. For future research, it would be worthwhile to shift the focus from the study of genres and text types towards the study of individual dimensions of text production that cut across genres and text types. Other dimensions that are likely to affect planners but that have received little attention so far include dialogic/monologic production, degree of familiarity between interlocutors, number of interlocutors and degree of personal involvement in topic matter. Not only can focusing on such dimensions provide new insight on the similarities and differences between different text types and genres, but it can also further our understanding of the factors that lead to the use of planners and other discourse markers.

Appendix

Spontaneous conversation

TABLE 3 List of recordings from the Santa Barbara Corpus of Spoken American English (SBC) included for the study (interactions recorded between 2000 and 2005)

Recording number	Title
SBC001	Actual Blacksmithing
SBC002	Lambada
SBC003	Conceptual Pesticides
SBC004	Raging Bureaucracy
SBC005	A Book about Death
SBC006	Cuz
SBC007	A Tree's Life
SBC011	This Retirement Bit
SBC013	Appease the Monster
SBC015	Deadly Diseases
SBC052	Oh You Need a Breadbox

Available at: <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

Improvised theatre

TABLE 4 List of improvised theatre shows included for the study

Duo	Show	Release date	Creators	Director/ Producer	Streaming platform
TJ and Dave	“Ass In Your Hand” “Bonks” “Blurting In Earnest” “Toute Suite”	2016	TJ Jagodowski, Dave Pasquesi	Clark Street Films (prod.)	Vimeo
Middleditch and Schwartz	“Parking Lot Wedding” “Law School Magic”	2020	Thomas Middleditch, Ben Schwartz	Ryan Polito (dir.)	Netflix

Available at: <https://www.netflix.com/title/81122572> (Middleditch and Schwartz); <https://vimeo.com/onde mand/tjanddave> (TJ and Dave)

Scripted fiction

TABLE 5 List of pilot TV episodes included for the study

Series	Release date	Production company	Director(s)
<i>Breaking Bad</i>	2008	High Bridge Productions, Gran Via Productions, Sony Pictures Television, American Movie Classics (AMC)	Vince Gilligan
<i>Brooklyn Nine Nine</i>	2013	Fremulon, Dr. Goor Productions, 3 Arts Entertainment, Universal Television	Phil Lord, Christopher Miller
<i>Desperate Housewives</i>	2004	Cherry Productions, Touchstone Television	Charles McDougall

TABLE 5 List of pilot TV episodes included for the study (*cont.*)

Series	Release date	Production company	Director(s)
<i>Gilmore Girls</i>	2000	Dorothy Parker Drank Here Productions, Hofflund/Polone, Warner Bros. Television	Lesli Linka Glatter
<i>Grey's Anatomy</i>	2005	The Mark Gordon Company, Touchstone Television	Peter Horton
<i>House of Cards</i>	2013	Media Rights Capital (MRC), Panic Pictures (II), Trigger Street Productions, Wade/Thomas Productions	David Fincher
<i>How I Met Your Mother</i>	2005	Bays Thomas Productions, 20th Century Fox Television	Pamela Fryman
<i>How to Get Away with Murder</i>	2014	Shondaland, Nowalk Entertainment, ABC Signature	Michael Offer
<i>Jane The Virgin</i>	2014	Poppy Productions, RCTV, Electus, CBS Television Studios, Warner Bros. Television	Brad Silberling
<i>New Girl</i>	2011	Elizabeth Meriwether Pictures, American Nitwits, Chernin Entertainment, 20th Century Fox Television	Jake Kasdan
<i>Orange Is the New Black</i>	2013	Tilted Productions, Lionsgate Television	Michael Trim
<i>Shameless</i>	2011	Bonanza Productions, John Wells Productions, Warner Bros. Television, Showtime Networks, Sterling Films (v)	Mark Mylod
<i>Suits</i>	2011	Hypnotic, Universal Cable Productions	Kevin Bray
<i>Two Broke Girls</i>	2011	Michael Patrick King Productions, Warner Bros. Television	James Burrows

TABLE 6 Results of significance testing for differences in the frequency of planners across pairs of data sets

	Number of words	Number of planners	Results of Chi-squared test with Yates' continuity correction	
Spontaneous conversation	53,353	374	Scripted TV series p-value = 4.604e-08	Improvised theatre p-value = 0.002893
Scripted TV series	66,406	305	–	p-value < 2.2e-16
Improvised theatre	47,610	414		–

TABLE 7 Results of significance testing for differences in the frequency of planners with cognitive function across pairs of data sets

	Number of words	Number of planners	Results of Chi-squared test with Yates' continuity correction	
Spontaneous conversation	53,353	188	Scripted TV series p-value = 2.277e-11	Improvised theatre p-value = 0.04182
Scripted TV series	66,406	105	–	p-value < 2.2e-16
Improvised Theatre	47,610	207		–

TABLE 8 Results of significance testing for differences in the frequency of planners with repair function across pairs of data sets

	Number of words	Number of planners	Results of Chi-squared test with Yates' continuity correction	
Spontaneous conversation	53,353	77	Scripted TV series p-value = 2.612e-13	Improvised theatre p-value = 0.6468
Scripted TV series	66,406	16	–	p-value = 8.85e-15
Improvised theatre	47,610	75		–

TABLE 9 Results of significance testing for differences in the frequency of planners with discourse management function across pairs of data sets

	Number of words	Number of planners	Results of Chi-squared test with Yates' continuity correction	
Spontaneous conversation	53,353	142	Scripted TV Series p-value = 4.843e-08	Improvised Theatre p-value = 0.9827
Scripted TV series	66,406	84	–	p-value = 5.893e-08
Improvised theatre	47,610	128		–

TABLE 10 Results of significance testing for differences in the frequency of planners with interpersonal function across pairs of data sets

	Number of words	Number of planners	Results of Chi-squared test with Yates' continuity correction	
Spontaneous conversation	53,353	24	Scripted TV series p-value < 2.2e-16	Improvised theatre p-value < 2.2e-16
Scripted TV series	66,406	164	–	p-value = 0.9131
Improvised theatre	47,610	120		–

Acknowledgements

We would like to thank Sophie Rosset and Arianna Hettenbach for their work on the transcriptions of improvised theatre performances and pilot episodes of TV series on which our analysis is based. We would also like to thank the anonymous reviewers, the editors of this special issue, the participants of the research colloquium at the English Department of the University of Basel and the audience at the “Contrastive corpus approaches to pragmatic markers: variation in use, translation issues and pedagogical applications” panel of the IPrA Conference in 2021 for their helpful comments and feedback on earlier versions of this paper. This research was supported by the Swiss National Science Foundation in the form of an SNSF PRIMA grant (PR00P1_185737/1).

References

- Bednarek, Monika. 2012. "Get us the hell out of here": Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics* 17(1): 35–63.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan (eds.). 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Clark, Herbert H. and Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1): 73–111.
- Crible, Ludivine. 2017. Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics* 22(2): 242–269.
- Crible, Ludivine. 2018. *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. Amsterdam: John Benjamins.
- Du Bois, John W, Susanna Cumming, Stephan Schuetze-Coburn, and Danae Paolino (eds.). 1992. *Discourse Transcription*.
- Erard, Michael. 2007. *Um ... slips, stumbles, and verbal blunders, and what they mean*. New York: Pantheon Books.
- Fox Tree, Jean E. 2007. Folk notions of *um* and *uh*, *you know*, and *like*. *Text and Talk* 27(3): 297–314.
- Götz, Sandra. 2013. *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins.
- Götz, Sandra. 2019. Filled pauses across proficiency levels, L1s and learning context variables: A multivariate exploration of the *Trinity Lancaster Corpus Sample*. *International Journal of Learner Corpus Research* 5(2): 159–180.
- Jucker, Andreas H. 2015a. Pragmatics of fiction: Literary uses of *uh* and *um*. *Journal of Pragmatics* (86): 63–67.
- Jucker, Andreas H. 2015b. *Uh* and *um* as planners in the *Corpus of Historical American English*. In: Irma Taavitsainen, Merja Kytö, Claudia Claridge, and Jeremy Smith (eds.), *Developments in English: Expanding electronic evidence*. Cambridge: Cambridge University Press, 162–177.
- Jucker, Andreas H. 2021. Features of orality in the language of fiction: A corpus-based investigation. *Language and Literature: International Journal of Stylistics* 30(4): 341–360.
- Kjellmer, Göran. 2003. Hesitation. In Defence of ER and ERM. *English Studies* 84(2): 170–198.
- Landert, Daniela. 2021. The spontaneous co-creation of comedy: Humour in improvised theatrical fiction. *Journal of Pragmatics* 173: 68–87.
- Laserna, Charlyn M., Yi-Tai Seih, and James W. Pennebaker. 2014. *Um ... Who Like Says You Know: Filler Word Use as a Function of Age, Gender, and Personality*. *Journal of Language and Social Psychology* 33(3): 328–338.

- Norrick, Neal R. 2014. Interjections. In: Christoph Rühlemann and Karin Aijmer (eds.), *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 249–276.
- Quaglio, Paulo. 2009. *Television Dialogue: The sitcom Friends vs. natural conversation*. Amsterdam/Philadelphia: John Benjamins.
- Richardson, Kay. 2010. *Television Dramatic Dialogue. A sociolinguistic study*. Oxford: Oxford University Press.
- Rühlemann, Christoph and Martin Hilpert. 2017. Colloquialization in journalistic writing. The case of inserts with a focus on *well*. *Journal of Historical Pragmatics* 18(1): 104–135.
- Staley, Larssyn and Andreas H. Jucker. 2021. “The uh deconstructed pumpkin pie”: The use of *uh* and *um* in Los Angeles restaurant server talk. *Journal of Pragmatics* 172: 21–34.
- Stenström, Anna-Brita. 2012. Pauses and hesitations. In: Gisle Andersen and Karin Aijmer (eds.), *Pragmatics of Society*. Berlin/Boston: Walter de Gruyter, 537–567.
- Tottie, Gunnel. 2011. *Uh* and *Um* as sociolinguistic markers in British English. *International Journal of Corpus Linguistics* 16(2): 173–197.
- Tottie, Gunnel. 2014a. Turn management and the fillers *uh* and *um*. In: Karin Aijmer and Christoph Rühlemann (eds.), *Corpus Pragmatics. A Handbook*. Cambridge: Cambridge University Press, 381–407.
- Tottie, Gunnel. 2014b. On the use of *uh* and *um* in American English. *Functions of Language* 21(1): 6–29.
- Tottie, Gunnel. 2016. Planning what to say: *Uh* and *um* among the pragmatic markers. In: Gunther Kaltenböck, Evelien Keizer, and Arne Lohmann (eds.), *Studies in Language Companion Series*. Amsterdam: John Benjamins Publishing, 97–122.
- Tottie, Gunnel. 2019. From pause to word: *uh*, *um* and *er* in written American English. *English Language and Linguistics* 23(1): 105–130.
- Tottie, Gunnel. 2020. Word-search as Word-formation? The Case of *uh* and *um*. In: Paloma Núñez-Pertejo, María José López-Couso, Belén Méndez-Naya, and Javier Pérez-Guerra (eds.), *Crossing Linguistic Boundaries: Systemic, Synchronic and Diachronic Variation in English*. London: Bloomsbury Academic, 29–42.

Biographical Notes

Ilenia Tonetti Tübben is a research assistant at the University of Basel. Her research interests include the pragmatics of fiction, discourse analysis, linguistic documentation and typology. She holds a BA in General and Comparative Linguistics and English Linguistics from the University of Regensburg (Germany) and an MA in Applied English Linguistics from the University of Dalarna (Sweden). In the past, she has published on the morphosyntactic

coding of proper names, and her MA thesis received Dalarna University's award for best thesis on sustainable development. She is currently working on her PhD project, in which she studies emotional expressions in fictional dialogues.

Daniela Landert is Professor of English Linguistics at Heidelberg University. Her research interests include the study of (spontaneous) spoken language, the pragmatics of fiction, mass media communication, historical pragmatics, modality and corpus pragmatics. Before joining Heidelberg University in 2022, she was the Principal Investigator of a research project on the pragmatics of improvised theatrical fiction, located at the University of Basel and financed through a PRIMA grant by the Swiss National Science Foundation. She is the author of the research monograph *Personalisation in Mass Media Communication* (2014) and she has published internationally in edited volumes and linguistic journals, including *Journal of Pragmatics and Discourse, Context and Media*.