



Target frames in British hotel websites¹

MIGUEL FUSTER-MÁRQUEZ & BARRY PENNOCK-SPECK*
IULMA, Universitat de València

Received: 26/11/2014. Accepted: 06/02/2015.

ABSTRACT

This article centres on four-word phrase frames in British hospitality websites. Our aim is to identify those frames that are specific to this website genre, which we call target frames. Each phrase frame represents an identical sequence of words except for one variable word, that is A*BC or AB*D. The words that fill the slot, marked with an asterisk, are called fillers. We used a corpus-driven approach using KfNgram software to identify the phrase frames in our corpus (COMETVAL). We regard phrase frames as genre-specific when they are significantly more frequent than those found in the written section of the BNC, which represents General British English. We further filtered our selection of phrase frames by eliminating those which were not semantically homogenous with regard to the variable words they contained. Only in this way could the 76 phrase frames we identified be classified according to their primary discourse function. We contend that our study is a valuable addition to the literature on phraseology and can be of use in pedagogical and professional settings.

KEYWORDS: lexical bundle, phrase frame, target frame, hotel websites, corpus linguistics, discourse.

RESUMEN

Este artículo estudia los marcos fraseológicos de cuatro palabras en sitios web de la hostelería británica. Nos proponemos identificar aquellos marcos fraseológicos que son específicos de dicho género, y que denominamos 'marcos meta'. Cada marco fraseológico representa una secuencia idéntica de palabras a excepción de un elemento variable, es decir A*BC o AB*D. Los elementos variables rellenan el lugar indicado por el asterisco. Para identificar de partida estos marcos en nuestro corpus (COMETVAL), se ha utilizado el programa KfNgram. Hemos considerado que estos marcos fraseológicos son representativos de un género solo cuando muestran una mayor frecuencia significativa que en inglés británico general, representado aquí por el subcorpus escrito del BNC. Además hemos filtrado estos marcos mediante la eliminación de aquellos que contenían elementos variables semánticamente heterogéneos. Sólo así podrían los 76 marcos fraseológicos representar una función discursiva distintiva. Sostenemos que nuestro estudio es una estimable contribución a la investigación fraseológica y puede ser de utilidad en contextos pedagógicos y profesionales.

PALABRAS CLAVE: paquete léxico, marco fraseológico, marco meta, sitios web de hoteles, lingüística de corpus, discurso.

**Address for correspondence:* Miguel Fuster Márquez. Departament de Filologia Anglesa i Alemanya, Facultat de Filologia, Traducció y Comunicació, Avda. Blasco Ibáñez, 32, 46010 Valencia, Spain; e-mail: Miguel.Fuster@uv.es

1. INTRODUCTION

There is an increasing interest in the role of phraseology in different types of discourse. A lot of research has been carried out on the importance of collocations and various other phraseological segments. Idioms, perhaps more than any other type of word sequence, have been central in phraseological studies, although it is widely acknowledged that their presence in general and specialised corpora which reflect native speech is minimal compared to the frequency of other phraseological segments (Biber, Conrad & Cortes, 2004). More recently, attention in research has turned towards so-called lexical bundles (LBs), which are “recurring sequence[s] of three or more words” (Biber et al. 1999: 990), such as the four-word bundles *it is important to*, *on the other hand*, *as well as the*, frequently found in academic writing, or *I don't know if*, *I don't think so*, *you want me to*, typical of spoken registers and also found in academic contexts (Biber, Conrad & Cortes, 2004). The methodology used to identify these sequences is frequency-driven and thus most of the sequences identified this way may strike us as unidiomatic and incomplete from a structural point of view, but they are seen to perform relevant specific functions in different registers.

It is undeniable that the phraseological approach has gained momentum ever since Sinclair (1991: 110), within the frame of corpus linguistics, stated the well-known ‘idiom principle’, which holds “[...] that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.” Sinclair’s idiom principle was crucial since it underscored that many of the open choices, as viewed by earlier linguists without the empirical support of corpus analysis, were definitely less relevant in language than the idiomatic (phraseological) choices that the analysis of corpora was able to unveil. Sinclair also highlighted the inadequacy of viewing lexis and grammar as separate components in language analysis. Many other corpus linguists, such as Römer (2010: 95), have claimed that word sequences rather than isolated words are the main meaning-carrier units. Similarly, Baker (2011: 238) has remarked that: “lexis and grammar mingle very closely in the specification of the extended units of meaning which are arguably the main vehicles of our expression and understanding.”

The interest in fixed sequences such as LBs has a starting point in corpus linguistics with Biber et al.’s publication of the *Longman Grammar of Spoken and Written English* in 1999. This, in turn, has generated a plethora of recent publications on ‘discontinuous’ word sequences that ran parallel to studies of LBs (see Biber, 2009; Fuster-Márquez, 2014; Grabowski, 2013; Gray & Biber, 2013; Römer 2009, 2010). However, discontinuous sequences have not gone unnoticed by linguists outside corpus linguistics. Thus, as far back as the 60s, Lyons (1968: 178), for whom phraseology was not a major issue, notes the existence of sequences such as *for *’s sake*, which he classified as phrase-schemas, where different words may fill the slot (*his*, *my mother*, etc.).

For their part, Sinclair and Renouf (1988: 154) became interested in three-word sequences that they called *discontinuous frameworks*. For example, *a ... of* would typically attract items such as *lot, kind, number, sort, couple, matter, bit, series, piece, member*. Cortes (2002: 134) claims that the status of these *discontinuous frameworks* is structurally and semantically contentious. By contrast, longer word sequences considerably narrow down the possibilities of variation and provide more congruous sequences, particularly when one analyses different genres separately.

More recently, Tucker (1996) analysed sequences such as *haven't the * idea*, which admits different insertions, *faintest, foggiest, slightest*, etc., without significant change of meaning. Philip (2008), drawing on earlier research (Francis, 1993; Moon, 1998), re-examines this same sequence and observes that, without losing sight of meaning, other slots allow for variability, e.g. *the slightest **, where the sequence may be completed by the nouns *idea, conception, or notion*. Therefore, in spite of the variability of the sequence, grammatical and meaning affinities remain. As will be shown in later sections, there is greater affinity between our analysis of the discontinuous sequences and the research carried out on the London-Lund Corpus by Eeg-Olofsson and Altenberg (1994). In their study, Eeg-Olofsson and Altenberg (1994: 63) focus on sequences such as *as far as X BE concerned* which host internal fillers that are variable and not fixed.

Our research in this article focuses on a type of non-continuous sequence called a Phrase Frame within the genre of British hospitality websites. It involves a quantitative and qualitative analysis of the 495,891-word subcorpus of 114 British hotel websites in the COMETVAL corpus, collected between the years 2010 and 2013. In our view, both the number of words and the number of different hotel websites in the corpus are sufficient for a meaningful examination of the recurrent phraseology in this genre. In this vein, Flowerdew (2012: 10) underscores that “even smallish corpora of around 300,000 words can throw up useful frequency data, provided that they concentrate on a specific subject area or genre.” As COMETVAL has nearly half a million words and focuses specifically on the language of hotel websites in a single variety, British English, we believe that it can offer relevant insights into this specific genre.

The term ‘Phrase Frame’ (henceforth PF) was introduced by Fletcher in his British National Corpus interface *Phrases In English* (2003-2006). Fletcher (2007) distinguishes these sequences from continuous word sequences, known in the literature as ‘n-grams’ (also ‘clusters’ or ‘chains’), stating that they are “sets of variants of an n-gram identical except for one word [...]” Thus, the 4-word PF *a * place to* in COMETVAL includes an invariant, or fixed, frame made up of an initial “a” and a final two-word sequence “*place to*” plus an intervening empty slot which is filled by different words. In descending order of frequency, the word fillers of this slot in our corpus are: *great, wonderful, lovely, perfect, desirable, tranquil, unique, convivial, pleasant, beautiful, and good*.

We will focus on 4-word PFs, rather than their 3- or 5-word counterparts. There are two main reasons for this; firstly, many 4-word PFs include 3-word PFs and secondly, 4-word PFs are more frequent than 5-word PFs and, as stated by Cortes (2004: 401), 4-word sequences “present a wider variety of structures and functions to analyse.” Our aim in this article is to identify those 4-word PFs which are found more frequently in our corpus than in other genres and thus, in our view, characterise the hotel website genre. We call these PFs *target frames*. To a certain extent, the choice of this group of frames is modelled on Cortes’s (2004) use of the term *target bundles* in her phraseological study of texts written by both students and experts in history and biology. We contend that the characteristic frames that we aim to identify in this article would also be useful in the teaching of English for tourism and the creation of hospitality web pages by highlighting phraseology which is based on the empirical evidence of native usage.

In the following sections, we start with an overview of the discourse features of hospitality web sites. Then we offer a detailed account of our methodology. This is followed by a results section, where we discuss the distribution, composition and discourse-pragmatic function of frames and highlight relevant quantitative and qualitative similarities and differences which we contrast with studies which have similar phraseological goals. Finally, in our conclusion, we also mention the pedagogical implications of our study.

2. THE DISCOURSE OF HOTEL WEBSITES

The relationship between those business specialists who design the texts on hotel websites, and the visitors to the web, who are potential clients, reflects a *Business to Consumer* type of relationship (B2C). The specific features of this commercial relationship are immediately apparent in the choice of vocabulary, phraseology, and grammar as well as other discourse features employed by the creative teams that design these sites. Needless to say, the persuasive overall design of hotel websites, containing abundant audio-visual material, is an essential part of the message and, in that sense, shows its allegiance to other advertising genres. However, our focus in this article will be exclusively on the phraseology of the texts found in our corpus of hotel websites.

Admittedly, the vocabulary of these websites is not as highly specialised as that found in texts where both writers and readers share a high degree of expertise and common interests. A B2C context presumes the use of vocabulary that may be easily understood by a wide readership. However, as has been remarked by other scholars (see Cowie, 1988: 129; Stubbs 1996: 12), we should not overplay the contrast between popular and expert writing or between technical and non-technical vocabulary since the dividing line remains fuzzy.

Apart from the non-specialist nature of communication between hotel owners and customers, the language of these websites is of a hybrid nature. The variety of texts and communication strategies shows that the use of specific linguistics features is dependent on

which text and section of the hotel website we are looking at. Hotel websites are made up of at least three main sections. The first includes texts whose objective is to persuade customers to plan a future stay at the hotel and/or make use of its services. This is a very prominent section of hotel websites that often falls under headings such as *welcome, location, attractions, offers, services, rooms, facilities, restaurant, accessibility*, etc. These promotional texts are perceptually salient and generally include rich multimedia content. The fact that their goal is to persuade potential customers to book a room or service limits technical language to a minimum. Unlike other vehicles of promotional discourse like general print and TV advertising (Byrne, 1992; Pennock-Speck & Fuster-Márquez, 2014; Pennock-Speck & del Saz Rubio, 2013), where the viewer/reader is distracted from his/her main purpose of watching the main TV programming, potential customers who visit hotel websites do so voluntarily. Although one might argue that these customers would really like to get straight to the information about location and price, the lengthy descriptions in many of the most elaborate websites can, in no way, be dismissed as superfluous or irrelevant. Thus hotel websites often give potential customers copious information and invite them to actively seek further information for themselves (see Brown & Levinson, 1987: 179). Their unmistakable goal is clear, to arouse the customers' interest in the specific products or services being offered.

The second type features non-promotional texts, which deal with legal issues concerning the rights and obligations of the service providers and customers, the terms covering the exchange of information on the site, etc. These texts appear under headings such as *terms and conditions, privacy policy, cancellation (policy)*. In sharp contrast to promotional texts, non-promotional discourse is not generally given a prominent place; typically the links to these pages are located at the bottom of the page and smaller type size is employed. These links include long texts that are devoid of any multimedia support. They bear more than a passing resemblance to the small print of ordinary contracts and the same type of text can be found on other websites where business transactions or agreements are found.

The third type of texts is found in the *reservation* section, where hotel owners deal with the details of booking rooms and services. These texts are pragmatically hybrid, since they share some of the features of promotional and non-promotional texts. We have found that together all three text types are essential for any kind of hotel transaction in a B2C context.

3. METHODOLOGICAL CRITERIA

3.1. Identification and selection of Phrase Frames in COMETVAL

In order to obtain a list of target 4-word PFs we had to decide on the methodology we would apply to identify them in our corpus of hospitality websites. Recent research into identifying

PFs has taken two routes. The first has been to find PFs based on the most frequent LBs (see Biber, 2009; Römer, 2010). The second is found in Gray and Biber's study (2013: 121) in which they state that "(...) identifying frames from bundles misses frequent discontinuous patterns which are highly variable." In an earlier publication, Fuster-Márquez (2014) followed this "fully inductive method" (Gray & Biber, 2013: 113); and this has also been the method adopted in this contribution. An obvious advantage is reflected in the results, which show that several relevant PFs in COMETVAL host no LBs at all. PFs like these would not have been detected if the procedure had relied on pre-existing LBs.

The next step was to decide on the minimum observed frequency that determines which discontinuous 4-word sequences are worthy of attention. Frequency thresholds are unavoidable in the analysis of bundles and frames. For Gray and Biber (2013: 115), frequency is an inherent part of phraseology, which is "typically interested in recurring patterns, rather than single occurrences. Thus, most corpus-based studies of phraseology set minimum frequency thresholds for identifying phraseological patterns." Römer (2010) points out the lack of scholarly agreement on statistical measures of phraseology. As a consequence, there is a certain degree of arbitrariness concerning thresholds. Nevertheless, as Byrd and Coxhead (2010: 32) remark, researchers apply cut-off points in the light of "what seems reasonable given the volume of data." As argued in Fuster-Márquez (2014: 92), a relatively high cut-off frequency can guarantee the genre specificity of the results. In this study, we established a cut-off frequency of twenty tokens for a 495,891-word corpus. The extrapolated value, normalised to one per million words, is roughly 40 tokens. This is high considering that Gray and Biber (2013: 112) use a threshold of 40 per million for a written corpus of 5.3 million words. However, COMETVAL makes use of more specialised and homogeneous texts, thus the threshold we have established seems to be reasonable for the phraseological analysis we have carried out. Also, while Gray and Biber (2013: 113) based their analysis on sequences that occurred in at least five texts, we only include sequences that are found in at least fifteen different British hotel websites. In this regard, we follow the more conservative approach taken in Fuster-Márquez (2014). Although the distributional restrictions imposed on this analysis set it apart from other analyses, this is done to ensure that the items analysed recur with sufficient frequency to be taken as truly representative of the domain.

Another methodological consideration affects the position of the variable slots in PFs to be investigated. We have followed Römer (2010: 101), who only takes into account PFs where the empty slot is internal; that is, in the case of 4-word PFs, second or third position: A*CD or third AB*D. The sequences *BCD or ABC* are discarded because "they could either be described as 3-grams (BCD and ABC) with a preceding or following variable slot or they could be part of a larger frame, e.g. a "5-p-frame" A*CDE or a "6-p-frame" ABC*EF" (Römer 2010: 103). Nevertheless, we acknowledge that a comprehensive phraseological profile of hotel websites should include three-word and five-word sequences as long as they

prove to have relevant functions in the discourse of hotel websites, but this falls outside the scope of this investigation.

3.2. Extraction and selection of four-word PFs: methodological overview

In this frequency-driven approach, the automatic generation of PFs was performed by means of *KjNgram* without relying on pre-existing LBs. We filtered out frames which came from the same hotel website and whose co-text showed that they were repeated verbatim at least to the extent of eight to ten words. This occurs, for example, in the description of different rooms or suites where the same sequence was often reused. We also discarded those frames which had punctuation boundaries within them. After the filtering process, where the free concordance software *Antconc 3.2.4.*, designed by Laurence Anthony (2007) was used extensively, the extraction yielded a total of 247 frames, representing 4,689 types, and 11,430 tokens. That is, over 2.3% of all tokens in COMETVAL were members of these sequences.

Whether recurrent collocations in a corpus are representative of specific genre merits attention. This is a central argument regarding the notion of *target frames*, i.e. frames that are genre-specific and hence may be useful for practitioners in a particular field. To detect whether a PF is genre-specific, Vincent (2013: 49) recommends that comparisons be carried out with other genres or with general corpora. We assume that phraseology in COMETVAL will show similarities with other promotional hospitality texts within the tourism macrogenre (see discussion in Calvi 2010), where a B2C relationship is found.

We have followed Vincent's (2013) recommendation to assess the specificity of frames in COMETVAL by establishing a contrast with the written section of the British National Corpus (henceforth BNCWr) which contains 90 million words and represents general written English. As Biber, Conrad and Reppen (1998: 234) claim, "characterizations of *general English* are usually not characterizations of any variety at all, but rather a middle ground that describes no actual text or register." This decision to use the BNCWr was also informed by the results of Fuster-Márquez (2014), which showed that LBs and PFs in British hotel websites had greater affinities with the written English section of the BNC than with the spoken conversational section. The BNC was accessed by means of the CQP-edition of BNCweb (Hoffmann & Evert, 1996-2008) which interacts with the BNC (<http://bncweb.lancs.ac.uk>). This interface provided us with valuable normalised frequencies of all 4-word PFs.

4. RESULTS AND DISCUSSION

In this section we offer a detailed account of the empirical results and discussion in relation to the relevant literature as seen from three different yet complementary perspectives; namely, distributional, structural and semantic, and discourse-pragmatic.

4.1. The distributional features of frames in hotel websites

Our distributional analysis shows that hotel websites host two main kinds of frames, as shown in Table 1 below: one whose fixed elements are all function words, e.g. *as * as the*, and another where at least one fixed element may be a content word, e.g. ‘setting’ in *the * setting for*. For this division between function and content words we have followed Quirk *et al.* (1985: 67). We have also provided the total number of 4-word LBs in COMETVAL for the sake of relevant phraseological contrasts, particularly to find out whether, or to what extent, LBs might be embedded in the relevant PFs. The thresholds for LBs are lower than those established for PFs, a minimum frequency of 14 tokens, and a range of 7 websites, due to the greater frequency of PFs.

	Number of PF types	%
Total PFs	247	
Function-word PFs	166	67.6
Content-word PFs	81	32.4
Total LBs	129	
LBs in all PFs	104	80.6
LBs in content-word PFs	48	45.19
LBs in function-word PFs	56	54.81

Table 1. 4-word PFs and LBs in COMETVAL

In Table 1 we can see that the number of PFs practically doubles that of LBs. Further, the number of function-word PFs (166, i.e., 67.6%) doubles that of content-word PFs (81, i.e., 32.4%). Eeg-Olofsson and Altenberg (1994) also found that content words were less frequently found in the invariant part of the discontinuous recurrent sequences they analysed. Likewise, our quantitative analysis confirms that the vast majority of bundles in COMETVAL are members of a PF (80.6%) of equal length. This evidence proves that hotel website writers very frequently resort to the variability of PFs alongside the fixedness of LBs. Also, it is interesting to note that even though there is a larger number of function-word PFs compared to content-word PFs, a greater proportion of bundles is found in content-word frames than in function-word frames.

In the light of these results, in the present analysis we decided to compare (see Table 2) the frequencies of content-word and function-word PFs in COMETVAL with equivalent frames in the BNC subcorpus of written English (BNCWr). Table 2 shows the frequency contrast between function-word PFs and content-word PFs in the BNCWr and COMETVAL. For the sake of comparison with the BNCWr, figures are normalised to one million words.

Frequencies of PFs per million in BNCWr	Number of content-word PFs (COMETVAL)	%	Number of function-word PFs (COMETVAL)	%
0 to 10	66	80.5	41	24.8
10.1 to 20	8	9.8	13	7.9
20.1 to 30	1	1.2	8	4.8
+30	7	8.5	103	62.5
TOTAL	82	100	165	100

Table 2. Frame frequency contrasts between COMETVAL and BNCWr

While looking at Table 2, it is important to remember that the cut-off frequency of tokens per PF in COMETVAL was set to a *minimum* of twenty tokens, (roughly 40 per million). However, the average number of tokens per PF in COMETVAL is actually much higher –over 46, (approximately 92 per million). When we look at column two in Table 2 we can see that 66 content-word PFs (85%) have counterparts with ten or fewer tokens per million in the BNCWr (see column one). For example, frequent content-word frames such as *the perfect * to, the time * booking, for * information please, we reserve * right*, etc. have a frequency of less than two per million in the BNCWr. The greater frequency of these content-word PFs in our corpus leads us to conclude that this is an important difference between COMETVAL and the BNCWr. One of the reasons may be that many content words in specialised discourses are topic-dependent (Stubbs & Barth 2003: 62).

On the other hand, Table 2 also shows that most of the function-word PFs in COMETVAL (62.5%) have counterparts in the BNCWr with frequencies of over thirty tokens per million. Indeed, the majority of the top PFs in COMETVAL (*the * of the, in the * of, at the * of, to the * of, on the * of, the * of a*) have equivalents in the BNCWr with frequencies of over 400 times per million words. Since the frequency of these same PFs in COMETVAL is lower, it may be argued that they are not genre-specific and thus perhaps might not be of interest to students and professional hotel website writers. In spite of this general tendency among the function-word PFs, a small number are fairly frequent in COMETVAL but show very low frequencies in the BNCWr: *for any * or, your * with us, we will * your, we may * your*. These function-word PFs certainly deserve attention.

In all (see row one in Table 2) we identified 107 PFs, i.e. 66 content-word PFs and 41 function-word PFs that were significantly more frequent in our corpus than in the BNCWr and thus good candidates for genre specificity from a quantitative point of view. These will undergo further qualitative filtering in section 4.3.

4.2. Structure and meaning

To date, studies on PFs have not looked into the grammar and meaning relations among word fillers, even though these have direct bearing on the relevant discourse functions attached to their corresponding frames. As Sinclair and Renouf (1988: 154) remark in their work on *discontinuous frameworks*, it is important to see whether such structures “attract particular lexical words”. For Tucker (1996) and Philip (2008) the phraseological relevance of the

discontinuous sequences they analyse hinges on the semantic closeness of the variable elements. These arguments have led us to analyse the word class/part of speech and meaning congruity of the slot fillers that complied with the genre-specific condition described in section 4.1. We decided that only those PFs in which at least 50% of their fillers were members of a semantically coherent group could be described as phraseologically relevant. Sequences that did not meet this requirement were excluded since the heterogeneity of fillers would point to the absence of stable discourse functions for these sequences.

Strikingly, when we examined the contents of the target PFs, we discovered that a low percentage, 31 out of 107 (29%), had fillers that did not contribute to homogeneous meanings. Moreover, we found that most of the non-coherent fillers were embedded in function PFs. Therefore, our analysis shows that only 76 PFs deserve to be called target frames (see Table 3). With the exception of the frame *per * per night*, which is made up of two distinct fillers, *room* and *person*, but a high number of tokens, all the others have a single relevant meaning group.

# Tokens	PFs	# Tokens	PFs
142	<i>at the * hotel</i>	36	<i>selection of * and</i>
93	<i>per * per night</i>	36	<i>for * information on</i>
86	<i>the perfect * for</i>	35	<i>to provide * with</i>
76	<i>up to * people</i>	33	<i>on * day of</i>
67	<i>you would * to</i>	33	<i>a * place to</i>
66	<i>the perfect * to</i>	32	<i>can * up to</i>
60	<i>a * selection of</i>	31	<i>use * personal information</i>
58	<i>you * like to</i>	31	<i>the * of arrival</i>
56	<i>you are * for</i>	31	<i>the * room is</i>
52	<i>in * of our</i>	30	<i>perfect * for a</i>
52	<i>the * hotel is</i>	29	<i>credit * debit card</i>
51	<i>if you * any</i>	28	<i>to the * hotel</i>
50	<i>the * of booking</i>	28	<i>of the * hotel</i>
49	<i>to * your stay</i>	28	<i>a great * to</i>
48	<i>up to * guests</i>	28	<i>is available * the</i>
46	<i>the time * booking</i>	27	<i>we may * your</i>
45	<i>find out * about</i>	27	<i>in the * centre</i>
44	<i>of * personal information</i>	26	<i>to make * stay</i>
44	<i>and * making facilities</i>	26	<i>is situated * the</i>
44	<i>find * more about</i>	26	<i>we can * a</i>
43	<i>we * the right</i>	26	<i>to make * reservation</i>
43	<i>your * with us</i>	25	<i>hotel is * in</i>
43	<i>to * you with</i>	24	<i>within * days of</i>
42	<i>we reserve * right</i>	24	<i>not be * for</i>
42	<i>terms and * of</i>	24	<i>for your * day</i>

42	<i>the * hotel in</i>	24	<i>offers a * of</i>
42	<i>in * city centre</i>	23	<i>to * a reservation</i>
42	<i>the * place to</i>	23	<i>is the * place</i>
42	<i>if * have any</i>	23	<i>take a * at</i>
41	<i>for any * or</i>	22	<i>to * your requirements</i>
39	<i>if * do not</i>	22	<i>is located * the</i>
39	<i>offer a * of</i>	21	<i>be * for any</i>
39	<i>you may * to</i>	21	<i>the * setting for</i>
38	<i>for * information please</i>	21	<i>to find * more</i>
37	<i>subject to * and</i>	21	<i>you * to make</i>
37	<i>the ideal * for</i>	20	<i>of our * rooms</i>
36	<i>if * wish to</i>	20	<i>the * venue for</i>
36	<i>at the * we</i>	20	<i>are available * request</i>

Table 3. Target PFs with number of tokens in COMETVAL

Our analysis of these 76 target PFs shows that three PFs contained two LBs each:

- (1) *per * per night: per **room** per night, per **person** per night;*
- (2) *for * information please: for **further** information please, for **more** information please;*
- (3) *for * information on: for **more** information on, for **further** information on.*

PFs that host LBs show structural and semantic coherence since bundles also require a high threshold to be taken into account. We also found 42 PFs that contain one LB, e.g. *a **wide** selection of*, *the **perfect** place to*, or *to provide **you** with*. However, surprisingly, no fewer than 32 PFs contained no canonical or LB sequence at all. This explains why measuring type/token regularity by means of entropy (Eeg-Olofsson & Altenberg, 1994), or VPR (Römer, 2010) was deemed to be non-conclusive for the analysis of PFs in hotel websites. A case in point is the sequence *hotel is * in*, with 25 tokens. The five different fillers are synonyms of location: *located* (12), *situated* (10), *shrouded* (1), *well-placed* (1), and *nestled* (1). If they had been judged by their frequency, none of the types qualifies as a bundle, but the referential function of the frame, thanks to the congruous locative fillers, is preserved. Much more extreme is the case of those frames whose fillers were made up of the names of different hotels. This confirms our initial hypothesis that PFs without very frequent or canonical LBs also deserve attention. The fact that the 76 PFs we have identified as target frames host fillers that belong to well-defined semantic sets contributes enormously to a clearer picture of the highly specific functions of these sequences in the language of hospitality websites.

As regards the structural features of these 76 target frames, while our division into function frames and content frames does not follow the classification adopted by Gray and Biber (2013:123), for the sake of comparability with the results they obtained in their analysis

of conversation and academic prose, we provide a list of our target frames divided into the three groups they mentioned and also quantified them. Thus the grammatical categories of target frames in COMETVAL offer these results, also displayed in Figure 1:

1. **Verb-based frames** (28 types): frame contains one or more modal, auxiliary or main verb (e.g. *can * up to, if * do not, we reserve * right*).
2. **Frames with any content word which is not a verb** (41 types): frame contains one or more nouns, adjectives, or adverbs but no verbs (e.g. *at the * hotel, the perfect * to, a * selection of*).
3. **Function word frames** (7 types): frame consists of only function words such as prepositions, determiners, conjunctions, pronouns, etc. (e.g. *in * of our, at the * we, to * you with*).

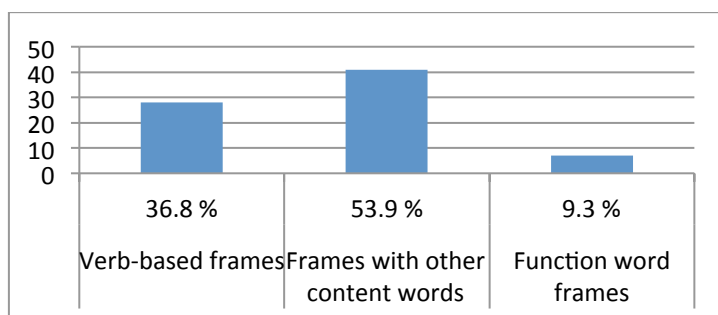


Figure 1. Distribution by form for PFs in COMETVAL

Straightforward comparability with the results obtained by Gray and Biber can only be very partial since a large number of non-target frames have been discarded in our research. Nevertheless, the target frames identified in COMETVAL show quite a sharp contrast with Gray and Biber's findings. Whereas these authors find that the frames with other content words are the least significant of the three structural groups both in conversation and academic prose, in the case of COMETVAL those frames happen to be the most significant. By contrast, the proportion of verb-based frames in COMETVAL appears to be close to that of the academic prose reported in Gray and Biber's study, and quite distant from conversation. Finally, the group of function-word frames is high in academic prose, but low in conversation. In this case, the proportion in COMETVAL is closer to the data from conversation described in Gray and Biber's study.

4.3. The discourse-pragmatic function of word fillers

Research on bundles (see Biber, Conrad & Cortes, 2004; Biber & Barbieri, 2007: 270, etc.) has focused on three discourse functions: (1) stance expressions, (2) discourse organizers, and (3) referential expressions. With regard to the first discourse function, Hyland (2008: 18) adds a finer subdivision of stance expressions to distinguish between (1) stance, as "the ways

writers explicitly intrude into the discourse to convey epistemic and affective judgements, evaluations and degrees of commitment to what they say”; and (2) engagement, which involves writers intervening “to actively address readers as participants in the unfolding discourse.”

We have taken these three discourse functions in the literature on bundles (including Hyland’s 2008 refinement of stance) as our starting point to classify the discourse functions of target frames. A first finding is that discourse organising PFs are simply not found among the target frames in COMETVAL (see also Fuster-Márquez, 2014). Therefore, the two relevant groups in the context of hotel websites are referential sequences, which “make direct reference to physical or abstract entities” (Biber & Barbieri 2007: 270) and stance sequences, which encompass writer-focused and reader-focused expressions. Tables 4 and 5 show the target frames with an indication of the most common filler in bold type to help in the identification of the primary function of each frame in hotel websites. Please note that those fillers which are hotel names are indicated by the word “**name**”, and those which have a random number simply by the word “**number**.”

Indicating location or direction	<i>at the (name) hotel, in one of our, in the city centre, in the city centre, is situated in the, is located on the, to the (name) hotel, of the (name) hotel, is available in the, hotel is situated in, the (name) hotel in</i>
Indication of time	<i>the time of booking, the time of booking, the day of arrival, on the day of, within (number) days of (also quantity specification)</i>
Quality specification of place, often with indication of purpose (by means of the prepositions <i>to</i> and <i>for</i>)	<i>a wide selection of, a great place to, a great place to, the perfect setting for, the perfect setting for, perfect setting for a, is the perfect place, the perfect place to, the perfect place to, the perfect venue for, the ideal venue for, the ideal venue for, selection of vegetables</i>
Identification or focus, with specification of goods or services offered	<i>the (kind of room) room is, and coffee making facilities, the (name) hotel is, our (kind of) rooms, your stay with us (perhaps stance), the dining room is</i>
Quantity specification	<i>up to (number) people, up to (number) guests</i>
Purpose and procedure	<i>per room per night, for any loss or, to make a reservation, to make a reservation, for your big day, credit or debit card, to discuss your requirements (but perhaps stance), to find out more (but perhaps stance)</i>
Others	<i>of your personal information, your stay with us, terms and conditions of</i>

Table 4. Referential Frames

Attitudinal/Modality stance: addressing and interpreting the wishes/needs of hotel guests (the presence of the pronoun <i>you</i> and conditional clause fragments are part of the frames)	<i>you would like to, you would like to, you are looking for, if you have any, if you have any, you may wish to, if you do not, if you wish to, you wish to make, to make your stay, to make your stay, to provide you with, to provide you with, offer a range of, offers a range of</i>
Ability: impersonal. Frames which show the capacity of the hotel in the provision of goods or services	<i>can accommodate up to, are available on request, subject to availability and (perhaps referential of identification or focus, like the following)</i>
Weak or mild obligation by means of frames containing directives (some frames without a verb present)	<i>find out more about, for further information please, for more information on (but might be interpreted as referential of purpose), take a look at</i>
Intention or prediction (often showing negative connotations). The pronoun <i>we</i> representing the hotel as a group of individuals or business team	<i>we reserve the right, we reserve the right, we may share your, not be liable for, be liable for any, we can offer a</i>
Intention by means of impersonal	<i>use of personal information</i>

Table 5. Stance Frames

Note in passing the effect of repetition or duplication of several of these target frames of which we offer a sample: *in * city centre/in the * centre; the * of booking/ the time * booking; a * place to/ a great * to; the * setting for/the perfect * for; etc.*

Referential frames (Table 4) are more numerous (44 frame types, 57.9%), and draw the readers' attention to information about hotels and the services they offer, which Crook (2004: 735) calls the product dimension in advertising. A secondary relevant function of these referential frames is that of self-mention, which is also typical of advertising. The frequent mentioning of the names of hotels in referential sequences is an obvious marketing strategy of shameless self-promotion.

On the whole, we have identified a lower number of target frames which convey stance discourse features (32 frame types, 42.1%). In the context of hotel websites, most often these stance frames express engagement and are, therefore, participant-oriented. In other words, the addresser overtly interacts with potential customers visiting the website. Note the frequent presence of the pronoun *you*, or the conditional in many of the sequences listed in Table 5. Self-mention is, in the case of stance frames, represented by the use of collective *we* to indicate intention or prediction. The distribution of these PFs in the different texts within COMETVAL is as follows:

1. **Promotional** (58 frames): containing persuasive language, found in hotel website sections such as *welcome, location, attractions, offers, services, rooms, facilities, etc.*: e.g. *in * of our, at the * hotel, to * your stay*.
2. **Non-Promotional** (11 frames): dealing with legal issues, rights and obligations. Found in hotel website sections such as *terms & conditions, privacy policy, cancellation (policy)*: e.g. *we reserve * right, of * personal information*.
3. **Reservation** (7 frames): which provide future hotel guests with specific details about booking: *per * per night, up to * guests, to make * reservation*.

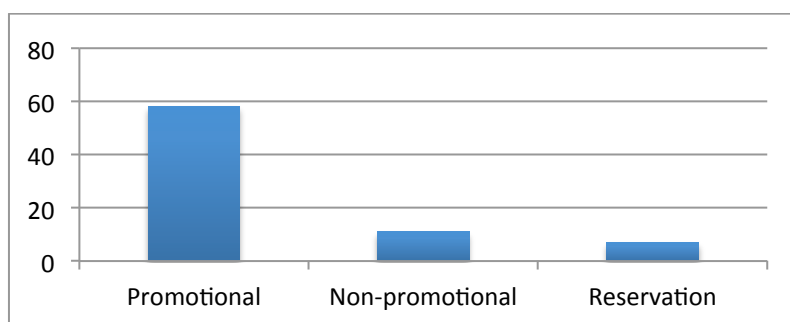


Figure 2. Distribution of PFs per type of text

Figure 2 shows that the vast majority of both referential and stance frames are found in promotional texts. On the contrary, very few are found in those texts which help potential customers to make a reservation, while a slightly higher number of frames are clearly non-promotional, e.g. *we * the right* or *terms and|& * of*. These are used to indicate legal restrictions, responsibility or liability regarding the use of the website or the services offered by the hotel. Finally, pragmatically, most target PFs in COMETVAL embody highly positive connotations, particularly those which are present in the promotional texts of websites.

5. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

Our aim in this article was to identify target 4-word PFs, that is, those 4-word frames which are significant from the point of view of frequency and that contain fillers which are semantically coherent. PFs make up just one example of the growing interest in the phraseological approach which considers that word sequences, both continuous and discontinuous, facilitate production and comprehension of texts by providing pre-fabricated chunks of language.

We have employed a purely inductive-empirical method to identify target PFs. We chose this method rather than selecting a series of common LBs and then finding the PFs that contained them. Thus, we found many frequent and semantically coherent PFs that contained no LBs at all. This would not have been possible using the second method. We then

contrasted our PFs with those from our reference corpus, the BNCWr, to discover which of these were genre-specific. The identification process was based on the relative frequency of our PFs compared to their counterparts in the BNCWr. Finally, we identified the discourse functions of our target PFs and their distribution. Our examination shows that the referential type represents a slightly higher proportion of PFs than the stance type, that is, 54 instances compared to 42. No discourse organising frames as such were found, which differentiates hospitality websites, for example, from the type of phraseology found in academic contexts where discourse organising LBs are found. This comes as no surprise given that academic journals, for example, are argumentative and text-oriented whereas hospitality websites combine the persuasive nature of promotional texts, which involves engaging with the customer, and the more transactional status of the reservation and the non-promotional text types. Pragmatically, the stance frames show how hotels go out of their way to try to win the customers' favour, a clear example of positive politeness strategies. The manner in which promotional texts enumerate the positive attributes of the hotels through purely referential PFs such as their location and the abundance of services they provide, can also be described as embodying positive politeness strategies such as exaggerating and intensifying interest (Brown & Levinson 1987: 104/106). The purely promotional texts in hotel websites are close to advertising genres.

It is not easy to decide on the pedagogical strategies to teach or learn these sequences. A challenging aspect of bundles and frames is that most of them are structurally incomplete. Another is the fact that 4-word sequences may often be better represented by shorter 3-word sequences contained within them, or longer 5-word sequences they are sometimes part of. In addition, not infrequently the frames discussed here contain repeated or overlapping n-grams. For instance, *the * of booking*, *the time * booking* may perhaps more fittingly be presented by a single bundle such as *the time of booking* than by any of these two frames. However, let us recall that not every frame hosts a bundle, e.g. *in the * hotel*. Although most researchers agree that bundles, and we would add frames, should be taught and studied, opinions differ as to the pedagogical approach to be implemented. Cortes (2004: 417-420) makes the point that learners' exposure to texts that contain bundles does not guarantee their acquisition or even that they will be noticed and so helping students to become aware of these sequences in relation to their context of use is a good strategy. On the other hand, Wray (2009: 10) notes that learners "need to operate within a number of different genres." However, she also notes, exposure to one specific genre does not mean that students will achieve proficiency in others. We claim that the target frames identified and selected here may also help students and professionals in the field of tourism to become better familiarised with the type of texts related to their profession. In this respect, these PFs should be taught in the specific context of tourism genres and consequently the materials writers (specialised dictionaries, textbooks, etc.) ought to address this issue. This is not to deny that students need exposure to frames and bundles of different sizes, as well as many other word combinations which are found in the

language of tourism. However, exposure to less specific phraseology, not exclusive of hotel websites, can be dealt with in more general courses with a focus on writing. Although the actual pedagogical implementation of PFs in the tourism classroom is beyond the scope of this article, it is easy to see that the PF *the perfect *for* could be used as part of a gap fill exercise in what Vincent (2013: 54) calls “guided discovery” which fits roughly within the data-driven learning tradition. Vincent (2013: 53) warns against a selection of sequences based on frequency counts alone as they would not “warrant pedagogical attention.” In this sense, our selection of PFs, identified by a quantitative and qualitative approach, avoids this problem.

NOTES

1. This contribution has been produced within the framework of the Research Project awarded by the Ministry of Economy and Competitiveness (earlier Ministry of Science and Innovation), reference FFI2011-24712, *Lexical and discursive analysis of parallel and comparable corpora (Spanish, English and French) of tourism promotional websites, 2011–2014*.

REFERENCES

- Anthony, L. (2007). *AntConc 3.2.4w (Windows)*. Tokyo: Waseda University.
- Baker, P. (2011). Times may change, but we will always have money: Diachronic variation in recent British English. *Journal of English Linguistics*, 39(1), 65–88.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311.
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263–286.
- Biber, D. & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181–190). Amsterdam: Rodopi.
- Biber, D., Conrad, C. & Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at ...: Lexical bundles in university teaching and textbooks*. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Brown, P. & Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. (2nd ed.). Cambridge: Cambridge University Press.
- Byrd, P. & Coxhead, A. (2010). *On the other hand: Lexical bundles in academic writing and in the teaching of EAP*. *University of Sydney Papers in TESOL*, 5, 31–64.
- Byrne, B. (1992). *Relevance Theory and the Language of Advertising (Dublin, CLCS Occasional Paper 31)*. Dublin: Trinity College.
- Calvi, M. M. (2010) Los géneros discursivos en la lengua del turismo: Una propuesta de clasificación. *Ibérica*, 19, 9–32.
- Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds), *Using Corpora to Explore Linguistic Variation* (pp. 131–145). Philadelphia: John Benjamins.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397–423.

- Cowie, A.P. (1988). Stable and creative aspects of vocabulary use. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 126–139). Longman: London and New York.
- Crook, J. (2004). On overt communication in advertising. *Journal of Pragmatics*, 36, 715–738.
- Eeg-Olofsson, M. & Altenberg, B. (1994). Discontinuous recurrent word combinations in the London-Lund Corpus. In U. Fries, G. Tottie, & P. Schneider (Eds.), *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora. Zürich 1993* (pp. 64–77). Amsterdam / Atlanta: Rodopi.
- Fletcher, W. H. (2007). Implementing a BNC-Comparable web corpus. In C. Fairon, H. Naets, A. Kilgarriff, & G.M. de Schryver (Eds.), *Building and Exploring web Corpora. Proceedings of the WAC3 Conference* (pp. 43–56). Louvain: Presses Universitaires de Louvain.
- Flowerdew, L. (2012). *Corpora and Language Education*. Basingstoke: Palgrave MacMillan.
- Francis, G. (1993). A corpus-driven approach to grammar. Principles, methods, and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In honour of John Sinclair* (pp. 137–156). Amsterdam: John Benjamins.
- Fuster-Márquez, M. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84–121.
- Grabowski, Ł. (2013). Register variation across English pharmaceutical texts: A corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics. *Procedia - Social and Behavioral Sciences*, 95, 391–401.
- Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), 109–135.
- Hoffmann, S. & Evert, S. (1996–2008). *BNCweb (CQP-edition)*. URL: <http://bnc-web.lancs.ac.uk> (Last accessed on November 20, 2014).
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21.
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Moon, R. E. (1998). *Fixed Expressions and Idioms In English: A Corpus-Based Approach*. Oxford: Clarendon.
- Pennock-Speck & del Saz Rubio, M. M. (2013). A multimodal analysis of facework strategies in a corpus of charity ads on British television. *Journal of Pragmatics*, 49(1), 38–56.
- Pennock-Speck, B. & Fuster-Márquez, M. (2014). Imperatives in voice-overs in British TV commercials: ‘Get this, buy that, taste the other.’ *Discourse and Communication*, 1–16. doi:10.1177/1750481314537578.
- Philip, G. (2008). Reassessing the canon: ‘Fixed’ phrases in general reference corpora. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 95–108). Amsterdam: John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London / New York: Longman.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7, 141–163.
- Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, 3(1), 95–119.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140–160). London / New York: Longman.
- Stubbs, M. (1996). *Text and Corpus Analysis: Corpus-assisted Studies of Language and Culture*. Oxford: Blackwell Publishers.
- Stubbs, M. & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language*, 10(1), 61–104.
- Tucker, G. H. (1996). So grammarians haven’t the faintest idea?: Reconciling grammar and lexis in a systemic functional model of language. In R. Hasan, C. Cloran & D. G. Butt (Eds.), *Functional Descriptions: Theory in Practice* (pp. 145–178). Amsterdam: John Benjamins.
- Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent words: A methodological exploration. *Journal of English for Academic Purposes*, 12, 44–56.

Wray, A. (2009) Future directions in formulaic language research. *Journal of Foreign Languages*, 32(6), 1–17.