

Phrase Frames en un Corpus Oral de Alemán como Lengua Extranjera para el Turismo

María Rosario Bautista Zambrana¹

¹ Universidad de Málaga
mrbautista@uma.es

Abstract. En este estudio partimos de una concepción amplia de la fraseología y nos proponemos analizar las estructuras fijas o *phrase frames* presentes en un corpus de textos orales de dos libros de texto de alemán turístico de nivel A2, con el fin, en primer lugar, de cuantificar su uso y contrastarlo con los datos de un corpus formado por los textos orales de dos libros de texto de alemán general como lengua extranjera, también de nivel A2. En segundo lugar, nos proponemos analizar 20 estructuras fijas, tanto desde un punto de vista cuantitativo como cualitativo, discerniendo, entre otros aspectos, las funciones comunicativas predominantes. Nuestros resultados apuntan a que el discurso oral del ámbito turístico, al menos tal como figura en los libros de texto, está más convencionalizado, y que podría tratarse de un discurso especializado, a pesar de su aparente similitud con el discurso oral general.

Keywords: *Phrase frames*, Corpus, Alemán como lengua extranjera, Turismo.

1 Introducción

Este artículo¹ pretende ahondar en la importancia del aprendizaje de unidades *prefabricadas* en el proceso de aprendizaje de una lengua extranjera, concretamente de la lengua alemana en el ámbito del turismo. Partimos de la idea de que gran parte de la lengua que usamos es de naturaleza fraseológica, tal como establece el *idiom principle* de Sinclair (1991: 110): “the principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.”

Numerosos investigadores, a partir de estudios basados en corpus, han apoyado esta tesis; es el caso, por ejemplo, de Bolinger (1976), Pawley y Syder (1983), Wray (2002), Granger y Meunier (2008) o Morley (2019), entre otros. Esta constatación tiene consecuencias muy importantes para la enseñanza y aprendizaje de lenguas extranjeras, incluso en los niveles iniciales. Según O’Keeffe et al. (2007: 46), la enseñanza de vocabulario en los niveles básicos resulta insuficiente si no se presta

¹ El presente trabajo se enmarca en el seno de los proyectos ‘VIP: sistema integrado Voz-texto para IntérPretes’ (ref. FFI2016-75831-P, MINECO) e INTERPRETA 2.0 (PIE 17-015, UMA).

debida atención a los *chunks* (combinaciones de palabras)² más usuales, ya que muchos de ellos se usan con igual o mayor frecuencia que ciertas palabras individuales comunes. Granger y Meunier (2008: 247) ahondan en esta idea: ya que, según han demostrado estudios psicolingüísticos, la lengua se adquiere, almacena y procesa por medio de *chunks*, es de recibo que la fraseología ocupe un lugar primordial en la enseñanza de segundas lenguas.

Sobre la base de estas premisas, el presente artículo parte de una concepción amplia de la fraseología (véase Granger y Paquot, 2008) y pretende analizar, desde un punto de vista cuantitativo y cualitativo, cuál es el tratamiento que reciben las unidades pluriverbales en los libros de texto de alemán turístico, con especial referencia al componente oral, tanto desde un punto de vista receptivo como productivo. De forma más concreta, nos centraremos en analizar la presencia de estructuras fijas en un corpus formado por los textos orales de dos libros de texto de alemán turístico de nivel A2. Este tipo de unidad pluriverbal viene definido en el texto del Marco Común Europeo de Referencia para las Lenguas (Consejo de Europa, 2002; en adelante MCER), documento que ha servido de base para la unificación de directrices sobre el aprendizaje y la enseñanza de lenguas dentro del contexto europeo.

Las *estructuras fijas* se abordan en el MCER dentro de la llamada competencia léxica, que es definida como el conocimiento del vocabulario de una lengua — formado por elementos léxicos y elementos gramaticales— y la capacidad para utilizarlo. Dentro de los elementos léxicos encontramos las *expresiones hechas*, que son aquellas compuestas de varias palabras que se utilizan y se aprenden como un todo. Entre ellas se encuentran las *fórmulas fijas* y las *estructuras fijas*. Estas últimas son expresiones “aprendidas y utilizadas como conjuntos no analizados, en los que se insertan palabras o frases para formar oraciones con sentido” (MCER, 2002: 108). Como ejemplos encontramos al. *Könnte ich bitte ... haben?*, o esp. *Por favor, ¿sería tan amable de + infinitivo?* Estas estructuras fijas son equivalentes a lo que Römer (2009: 150) denomina *phrase frame*: “sets of n-grams which are identical except for one word, e.g. *at the end of*, *at the beginning of*, and *at the turn of* would all be part of the p[hrase]-frame *at the * of*.” Consideramos que estas secuencias que presentan algún elemento variable son de extraordinaria relevancia en el aprendizaje de una lengua, no solo por su frecuencia de aparición (como veremos en este artículo), sino también por su productividad. Römer (2010) y Fuster (2014) han realizado estudios

² Hay muchos términos para denominar a las combinaciones frecuentes de palabras: *unidades fraseológicas* (término usado sobre todo por el enfoque más estricto de la fraseología), *secuencias formulaicas* (Wray, 2002), *chunks* (De Cock, 2000), *expresiones hechas* (MCER, 2002), *phrasal patterns* (como tipo de *construcción*; Goldberg, 2006), *expresiones fijas*, *unidades pluriverbales*, *combinaciones usuales de palabras*, etc. En este trabajo hablaremos de *expresiones hechas* o *unidades pluriverbales*, en el sentido que les asigna el MCER (2002): expresiones compuestas de varias palabras que se utilizan y se aprenden como un todo.

muy destacados sobre *phrase frames* que han servido de base para una parte de los análisis presentados en este trabajo.³

En relación con las estructuras fijas, serán de relevancia para nuestro trabajo las funciones comunicativas que el MCER establece dentro de la competencia funcional que debe adquirir el alumno (MCER, 2002: 122-123):⁴ ofrecer y buscar información factual (1.1), expresar y descubrir actitudes (1.2), persuasión (1.3), vida social (1.4), estructuración del discurso (1.5) y corrección de la comunicación (1.6).

En este estudio nos proponemos, pues, analizar las estructuras fijas presentes en dos libros de texto de alemán turístico, con el fin, en primer lugar, de cuantificar su uso y contrastarlo con los datos de un corpus formado por los textos orales de dos libros de texto generales de alemán como lengua extranjera; en segundo lugar, nos proponemos analizar 20 estructuras fijas, tanto desde un punto de vista cuantitativo como cualitativo, discerniendo, entre otros aspectos, las funciones comunicativas predominantes. Nuestros resultados serán de carácter preliminar, debido al tamaño del corpus y el número limitado de estructuras estudiadas. El artículo se encuentra organizado de la siguiente forma: en el Apartado 2 presentamos la metodología seguida para llevar a cabo el trabajo; en el Apartado 3 exponemos los resultados del análisis cuantitativo y cualitativo; finalmente, el Apartado 4 expone una breve discusión de los resultados, así como las conclusiones preliminares de este estudio.

2 Metodología

Hemos seguido una metodología cuantitativa y cualitativa. A fin de realizar los análisis lingüísticos que nos hemos propuesto, hemos compilado en primer lugar un corpus de textos orales a partir de dos libros de texto de enseñanza de alemán turístico de nivel A2: *Herzlich willkommen Neu* (Cohen y Lemcke, 2001) y *Menschen im Beruf Tourismus A2* (Schümann et al., 2015). Hemos seleccionado las transcripciones de los diálogos y monólogos relativos a situaciones profesionales, así como las listas de expresiones usuales (*Redemittel*), orientadas igualmente a la expresión oral de situaciones profesionales.⁵ Hemos descartado aquellos textos orales que no tuvieran relación con situaciones profesionales (diálogos entre turistas, por ejemplo) y los ejercicios de fonética; se han eliminado en los textos seleccionados todos los enunciados explicativos, títulos y referencias a los hablantes al principio de cada intervención. Este corpus, denominado *Corpus 1*, contiene 153 textos, con un total de 15.996 *tokens* y 1.895 *types* (*type-token ratio* 12%).

³ Utilizaremos en este trabajo las denominaciones *phrase frames* y *estructuras fijas* de forma indistinta.

⁴ Se trata, concretamente, de *microfunciones*, definidas como “categorías para el uso funcional de enunciados aislados (generalmente breves), normalmente como turnos de palabra de una interacción” (MCER, 2002: 122).

⁵ Las situaciones profesionales de estos libros de texto abarcan la hostelería, la atención al cliente en recepciones de hotel y la información turística, y tienen lugar generalmente entre un especialista y un consumidor.

A fin de contrastar los datos extraídos del Corpus 1, hemos compilado un corpus general de alemán, con textos orales de manuales de enseñanza de alemán como lengua extranjera de nivel A2; llamaremos a este corpus *Corpus 2*. De la misma forma que en el caso anterior, se han compilado las transcripciones de diálogos y monólogos, descartando aquellos textos orales relacionados con ejercicios de fonética. Los textos proceden de los libros *DaF kompakt A2* (Sander et al., 2011) y *Aussichten A2.1* (Ros-El Hosni et al., 2010). Este corpus contiene 150 textos, que cuentan con un total de 18.665 *tokens* y 3.011 *types* (*type-token ratio* 16%).

El estudio de ambos corpus ha sido en principio de tipo *corpus-driven*, en el sentido de que hemos extraído los *phrase frames* de forma automática sin tener en cuenta ninguna categoría lingüística previa ni ninguna lista predefinida de expresiones.

Hemos llevado a cabo el análisis cuantitativo por medio de *kfNgram* (Fletcher, 2007), con el que hemos extraído todas las estructuras fijas o *phrase frames* de 2 a 6 palabras del Corpus 1 y del Corpus 2, respectivamente. Después realizamos un estudio cualitativo de los *phrase frames* de 4 palabras del Corpus 1, para lo cual extrajimos todos aquellos que aparecían con una frecuencia normalizada superior o igual a 30 veces por 100.000 palabras; esto, en el Corpus 1, supuso que la frecuencia de corte para extraer *phrase frames* fuese de 5 *tokens*. De estos tomamos 20 ejemplos, tratando de seleccionar aquellas estructuras que constituyeran una unidad comunicativa mínima, es decir, que contuvieran un sintagma, aunque incompleto, o que incluyeran al menos sujeto y verbo (aun cuando uno de estos fuera el constituyente variable). Luego asignamos a cada *phrase frame* su función comunicativa y comparamos su frecuencia de aparición con aquella de la misma estructura fija en el corpus general. Para esto extrajimos igualmente todas las estructuras fijas del Corpus 2 con una frecuencia normalizada de 30 *tokens* por 100.000 palabras, lo que corresponde a una frecuencia de corte de 6 *tokens*.

3 Resultados

Ofrecemos en primer lugar los resultados de la extracción de *phrase frames* con *n* 3 a 6, tanto en el Corpus 1 como en el Corpus 2, realizada con *kfNgram*. Hemos reflejado en la siguiente tabla, como explicamos en el apartado anterior, aquellos *phrase frames* que cuentan con al menos una frecuencia de 5 *tokens* en el Corpus 1, y con al menos 6 *tokens* en el Corpus 2. Hemos indicado el número total de *phrase frames* encontrados, así como la cifra total de variantes detectadas. Se ha indicado igualmente cuál es el cociente entre el número de variantes y el número de *phrase frames*.

Tabla 1. Frecuencia absoluta de *phrase frames* en Corpus 1 y 2.

Corpus 1 (frec. absoluta)	Var./PF	Corpus 2 (frec. absoluta)	Var./PF
n=3		n=3	
1.155 PF / 7.380 var.	6,39	774 PF / 6.381 var.	8,2
n=4		n=4	

300 PF / 1.379 var. n=5	4,6	73 PF / 349 var. n=5	4,8
68 PF / 290 var. n=6	4,3	18 PF / 59 var. n=6	3,3
30 PF / 101 var.	3,4	9 PF / 31 var.	3,4

Como observamos, resulta evidente que en el corpus de textos orales turísticos hay un número mayor de estructuras fijas, frente al corpus general, tanto en lo que se refiere a los *phrase frames* de 3 palabras, como a los de 4, 5 y 6 palabras. Resulta llamativo que el número de estructuras fijas baje considerablemente al pasar de 3 a 4 palabras, tanto en el Corpus 1 como en el Corpus 2. Es preciso también resaltar que el número medio de variantes por *phrase frame* va descendiendo a medida que aumenta el número de *n*.

A continuación, hemos extraído 20 ejemplos de entre las estructuras fijas que aparecían 5 veces o más en el Corpus 1 (véase Tabla 2). Para cada uno de ellos, hemos indicado su rango (orden en el que aparece en la lista de *phrase frames* extraídos), frecuencia (número de *tokens*), número de variantes, *variant/p-frame ratio* (VPR) y función comunicativa (según el código asignado por el MCER, véase arriba). El *variant/p-frame ratio* (VPR) es un cociente propuesto por Römer (2010), que expresa la relación entre el número de variantes de cada *phrase frame* con respecto a los *tokens* que existen con ese patrón (cociente de variantes entre *tokens*, multiplicado por cien). Un resultado bajo indica que hay pocas variantes, mientras que un ratio alto significa que una estructura fija cuenta con numerosas variantes.

Tabla 2. Datos de *phrase frames* seleccionados.

Rango	Phrase frame	Frecuencia	Variantes	VPR	Función comunicativa
9	darf ich Ihnen *	18	7	38%	1.2
12	kann ich Ihnen *	16	6	37,5%	1.2
15	haben Sie schon *	14	3	21%	1.1
19	können Sie mir *	11	9	81%	1.2/1.3
21	ich hätte gern *	11	9	81%	1.2
23	dann nehme ich *	11	9	81%	1.2
30	ich bringe Ihnen *	10	5	50%	1.3
37	haben Sie noch *	10	6	60%	1.1
43	wir möchten gern *	9	6	66,66%	1.2
49	wie viel kostet *	9	3	33,33%	1.1
50	das Essen war *	8	6	75%	1.2
57	hier ist * Rechnung	8	2	25%	1.1
64	das hier ist *	8	7	87,5%	1.1
80	ich bringe * gleich	7	2	28,57%	1.3

86	das macht * Euro	7	2	28,57%	1.1
104	hier ist die *	7	6	85,71%	1.1
128	mit dem * fahren	6	2	33,33%	1.1
131	ich hoffe, dass *	6	3	50%	1.2
144	wie kommt man *	6	4	66,66%	1.1
154	ich wünsche Ihnen *	6	3	50%	1.2

Como se puede observar, la mayoría de estructuras fijas siguen el patrón A B C *, mientras que solo cuatro de ellas presentan el patrón A B * D. Según Römer (2010) y Fuster (2014), en estos casos no se trataría estrictamente de *phrase frames*; sin embargo, otros autores como Biber (2009) sí los tienen en cuenta, al tiempo que el propio MCER (2002) considera estructuras fijas todas aquellas que se aprenden y utilizan como conjuntos no analizados, sin tener en cuenta la posición del elemento variable. En el ejemplo del MCER (2002) *Por favor, ¿sería tan amable de + infinitivo?* podemos observar como el último constituyente es el variable, que además puede completarse por medio de más de una palabra.

Otro resultado destacable es que aproximadamente el 50% de los *phrase frames* seleccionados cumplen la función 1.1 (ofrecer y buscar información factual), mientras que el otro 50% presenta sobre todo la función 1.2 (expresar y descubrir actitudes).

En cuanto al cociente VPR, observamos una variación considerable entre los distintos *phrase frames*. La media de todos los resultados es de 53,99%, con una mediana de 50 y una desviación estándar de 22,67.

Si buscamos en el Corpus 2 los *phrase frames* seleccionados, encontramos que solo cinco de ellos se encuentran en las listas de estructuras fijas extraídas por *kfNgram*, teniendo en cuenta incluso aquellos *phrase frames* que solo cuentan con 2 *tokens*. En la Tabla 3 podemos comparar la frecuencia normalizada de aparición (por 100.000 palabras) de estas cinco estructuras:

Tabla 3. Comparación de frecuencias y rango de PF.

Phrase frame	Rango Corpus 1	Frec. norm. Corpus 1	Rango Corpus 2	Frec. norm. Corpus 2
haben Sie schon *	15	87,5	7.107	10,7
können Sie mir *	19	68,8	527	26,8
ich hätte gern *	21	68,8	3.171	10,7
hier ist die *	104	43,8	2.107	16,1
mit dem * fahren	128	37,5	2.839	10,7

4 Discusión y conclusiones

En este trabajo nos hemos propuesto compilar un corpus de textos orales de alemán como lengua extranjera en el ámbito específico del turismo —concretamente de textos

orales de comunicación profesional especialista-consumidor—, a fin de estudiar las estructuras fijas presentes en el mismo. Igualmente, hemos compilado un corpus de textos orales de dos libros de alemán general como lengua extranjera, que nos ha servido para establecer comparaciones.

Aunque el hecho de que ambos corpus estén formados por materiales del nivel A2 se puede considerar una limitación, creemos que de hecho es una ventaja haber contado con textos orientados a un mismo nivel de referencia, ya que esto nos permite realizar contrastes entre textos que (supuestamente) cuentan con estructuras gramaticales parecidas y están diseñados para adquirir las mismas competencias.⁶

Hemos constatado que hay más *phrase frames* (de 3, 4, 5 y 6 palabras) en el Corpus 1 que en el Corpus 2, lo que apunta a que el discurso oral turístico, al menos tal como se presenta en los libros de texto, está más convencionalizado que el discurso oral general. Los constituyentes libres de las estructuras fijas que hemos seleccionado pueden completarse con una palabra o con varias. Hemos comprobado, por otro lado, que el número de variantes no es uniforme, pues algunos *phrase frames* son muy productivos y admiten varias variantes, mientras que otros apenas pueden modificarse con dos constituyentes diferentes; el VPR medio es de 53,99% (mediana: 50; desviación estándar: 22,67). Observamos que ambos corpus presentan un cociente entre variantes y estructuras fijas descendiente, de manera que a medida que aumenta el número de constituyentes de los *phrase frames*, desciende el número de variantes.

Las estructuras fijas seleccionadas (de cuatro palabras) presentan en su mayoría unas funciones comunicativas muy concretas: predominan, por un lado, las del tipo 1.1 (ofrecer y buscar información factual), sobre todo en forma de preguntas, peticiones de información y respuestas; y por otro, del tipo 1.2 (expresar y descubrir actitudes), con expresiones de volición (deseos), modalidad (permiso) y emociones.

Los *phrase frames* seleccionados, por otro lado, suelen presentar la distribución A B C *, con unos pocos del tipo A B * D. Aunque no hemos mostrado el resto de *phrase frames* detectados, estos siguen mayoritariamente el mismo patrón, de manera que predominan los del tipo A B C * o * B C D.

La comparación de las estructuras fijas encontradas en el Corpus 1 con aquellas presentes en el Corpus 2 arroja unos resultados significativos: de las 20 estudiadas, 15 no se encuentran en el Corpus 2; mientras que las cinco que sí encontramos presentan una frecuencia de aparición mucho menor. Esto apunta a que el discurso oral turístico sería un discurso especializado, a pesar de que pueda parecer que se presta a muchas similitudes con el discurso general. Fuster (2014) llega a una conclusión muy parecida al estudiar los *lexical bundles* y *phrase frames* de un corpus en inglés de páginas web hoteleras.

El posible (alto) grado de convencionalización de este tipo de discurso oral y su carácter de especialidad tienen consecuencias relevantes para su enseñanza y aprendizaje. Estos resultados, aunque preliminares, apuntan a que sería conveniente dar más peso a las estructuras fijas en la enseñanza del alemán turístico, y promover

⁶ Véase el descriptor global del nivel A2 del MCER (2002: 26): El alumno “(e)s capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia que le son especialmente relevantes.”

la práctica y repetición de distintos patrones y de sus variantes más frecuentes. Varios estudios han demostrado que las expresiones hechas o unidades pluriverbales (*formulaic language*) presentan ventajas de procesamiento, y que los aprendices son capaces de extrapolar información lingüística a partir de este tipo de unidades (véanse los estudios citados por Vandeweerd y Keijzer, 2018). Esperamos poder ampliar los resultados y discusión de este estudio con un corpus de mayor tamaño y con textos no solo provenientes de materiales didácticos.

Bibliografía

1. Biber, D.: 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 (2009).
2. Bolinger, D.: Meaning and memory. *Forum Linguisticum* 1, 1–14 (1976).
3. Cohen, U., Lemcke, C.: *Herzlich willkommen Neu*. Langenscheidt, Berlín (2001).
4. Consejo de Europa / Ministerio de Educación, Cultura y Deporte: Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación. Secretaría General Técnica del MECD-Subdirección General de Información y Publicaciones, y Grupo Anaya, Madrid (2002).
5. De Cock, S.: Repetitive phrasal chunkiness and advanced EFL speech and writing. In: Mair, C., Hundt, M. (eds.) *Corpus Linguistics and Linguistic Theory*. Papers from ICAME 20, pp. 51–68. Rodopi, Ámsterdam (2000).
6. Fletcher, W. H.: *kfNgram* [software], <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>, último acceso 10/05/2019 (2007).
7. Fuster Márquez, M. Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction* 7(1), 84–121 (2014).
8. Goldberg, A.: *Constructions at Work*. Oxford University Press, Oxford (2006).
9. Granger, S., Meunier, F.: Phraseology in language learning and teaching: where to from here? In: Meunier, F., Granger, S. (eds.) *Phraseology in Foreign Language Learning and Teaching*, pp. 247–252. John Benjamins, Amsterdam & Philadelphia (2008).
10. Granger, S., Paquot, M.: Disentangling the phraseological web. In: Granger, S., Meunier, F. (eds.) *Phraseology: An Interdisciplinary Perspective*, pp. 27–50. John Benjamins, Amsterdam (2008).
11. Morley, J.: About Academic Phrasebank, <http://www.phrasebank.manchester.ac.uk/about-academic-phrasebank/>, último acceso 3/05/2019 (2019).
12. O’Keeffe, A., McCarthy, M., Carter, R.: *From Corpus to Classroom: language use and language teaching*. Cambridge University Press, Cambridge (2007).
13. Pawley, A., Syder, F. H.: Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: Richards, J. C., Schmidt, R. W. (eds.) *Language and Communication*, p. 191–226. Longman, Nueva York (1983).
14. Römer, U.: The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7, 141–163 (2009).
15. Römer, U.: Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction* 3(1), 95–119 (2010).
16. Ros-El Hosni, L., Swerlowa, O., Klötzer, S., Jentges, S., Sokolowski, K., et al.: *Aussichten A2.1*. Klett, Stuttgart (2010).

17. Sander, I., Braun, B., Doubek, M., Frater-Vogel, A., Fügert, N., Köhl-Kuhn, R., Trebesius-Bensch, U., Vitale, R., Behnes, S., Marquardt-Langermann, M.: DaF kompakt A2. Deutsch als Fremdsprache für Erwachsene. Klett, Stuttgart (2011).
18. Sinclair, J.: Corpus, Concordance and Collocation. Oxford University Press, Oxford (1991).
19. Schümann, A., Schurig, C., Schaefer, B., van der Werff, F. Menschen im Beruf Tourismus A2. Hueber, München (2015).
20. Vandeweerd, N., Keijzer, M. J'ai l'impression que: Lexical Bundles in the Dialogues of Beginner French Textbooks. *Canadian Journal of Applied Linguistics* 21(2), 80–101 (2018).
21. Wray, A.: Formulaic Language and the Lexicon. Cambridge University Press, Cambridge (2002).