

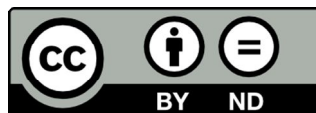


UNIVERSITAT DE
BARCELONA

La variabilidad lingüística nativa y no nativa en escenarios comunicativos

La función de transacción en una situación cotidiana:
corpus y descripción para el español (L1/LE)

Mónica Caballero Benavente



Aquesta tesi doctoral està subjecta a la llicència Reconeixement- SenseObraDerivada 3.0. Espanya de Creative Commons.

Esta tesis doctoral está sujeta a la licencia Reconocimiento - SinObraDerivada 3.0. España de Creative Commons.

This doctoral thesis is licensed under the Creative Commons Attribution-NoDerivatives 3.0. Spain License.

La variabilidad lingüística nativa y no nativa en escenarios comunicativos.

La función de transacción en una situación
cotidiana: corpus y descripción para el español
(L1/LE)

Mónica Caballero Benavente

Tesis presentada para optar
al grado de **Doctor en Lingüística**,
en el programa de doctorado de *Lingüística teórica,*
computacional y aplicada,
Departamento de Lingüística General
UNIVERSITAT DE BARCELONA

Bajo la supervisión de

Dra. Lourdes Díaz Rodríguez
Departamento de Traducción y Ciencias del Lenguaje
Universitat Pompeu Fabra

Dra. Maria Taulé Delor
Departamento de Lingüística General
Universitat de Barcelona



UNIVERSITAT DE
BARCELONA

Diciembre de 2015

Para Antonio y Dolores

ABSTRACT

This dissertation consists of an empirical study of spoken language based on the analysis of a real communicative setting, namely information transaction and ticket purchase in the field of railway transport. Its description of the most significant functional, lexical and socio-pragmatic contents aims to outline a proposal for a core inventory of the language involved in the above mentioned oral communicative situation, to be applied in the learning and teaching of Spanish as a foreign language, levels B1 and B2 in the *Common European Framework of Reference for Languages (CEFR)* (Council of Europe, 2001), although no specific methodology is proposed.

Data were collected from *FerrovieLE*, a corpus created for this thesis. The participants in this corpus include university students learning Spanish as a foreign language at levels B1 and B2 (*CEFR*), as well as native speakers of Spanish of different ages, gender and socio-cultural backgrounds (N=100).

FerrovieLE comprises five subcorpora of dialogues obtained by means of three different methodologies. Every subcorpus was transcribed and annotations on morphology and the pragmatic functions found were made according to the corresponding annotation guide. In this guide, the repertoire of communicative functions in *CEFR* is re-elaborated on the basis of the descriptors in the *Vantage Level* (van Ek y Trim 2001) and the *Core Inventory* by EAQUALS and British Council (North et al. 2010). In addition, the labels of the lexical notions, resulting from the refinement of those in the proposal in the *Plan Curricular del Instituto Cervantes* (B1, B2), are presented. A description of the morphosyntactic annotation is also included.

The inventories resulting from the empirical analysis are pragmatic and strategic-cognitive. They are systematized according to the level of the oral communicative skills. Therefore, they could contribute greatly to the field of Spanish as a foreign language learning and teaching, inspiring future studies and didactic applications aimed at improving learning and teaching materials in this field.

Key words: oral corpus, transactions, communicative settings, transport, interaction prototype, communicative functions, lexical notions, core inventory

RESUMEN

En este trabajo de tesis se realiza un estudio de corte empírico de la lengua hablada, sobre el análisis de un escenario comunicativo real: transacción de información y compra-venta de billetes en el ámbito del transporte en tren. Se trata de describir el contenido funcional, nocional (léxico) y sociopragmático más significativo, con la objetivo final de proponer la descripción de un inventario básico de la lengua oral en la función de transacción de información y compra-venta de billetes de tren, para enseñar y aprender en los niveles B1-B2 del *Marco Común Europeo de Referencia (MCER)* (Consejo de Europa 2001), sin especificar ninguna metodología de enseñanza.

El método de obtención de datos es a partir del corpus *FerrovíELE*, creado específicamente para esta tesis doctoral con la participación de informantes universitarios que estudiaban español como lengua extranjera en los niveles B1 y B2 (*MCER*) y no discentes, y por hablantes nativos de edades, sexos y niveles socioculturales diferentes (N=100).

FerrovíELE está formado por cinco subcorpus que corresponden a tres metodologías distintas de obtención de los diálogos. Cada subcorpus se ha transcrito ortográficamente y se ha anotado morfológica y pragmático-funcional con su correspondiente guía de anotación. En esta guía se reelabora el repertorio de las funciones comunicativas, disponibles en el *MCER* y basadas en los descriptores del *Vantage Level* (van Ek y Trim 2001) y el *Core Inventory* de EAQUALS y el British Council (North et al. 2010). Igualmente, se presenta el conjunto de etiquetas que constituyen las nociones léxicas, que son el resultado del refinamiento de las propuestas para el español del *Plan Curricular del Instituto Cervantes* (B1, B2). Además se incluye la descripción de la anotación morfosintáctica.

De lo analizado empíricamente se obtienen repertorios de carácter pragmático y estratégico-cognitivo, sistematizados por niveles de competencia en una destreza comunicativa, la oral. Estos repertorios constituyen una importante aportación al campo de la enseñanza y aprendizaje en E/LE y servirán para inspirar otros trabajos similares o aplicaciones didácticas con vistas a mejorar los materiales docentes de la enseñanza en el área de E/LE.

Palabras clave: corpus oral, interacción formal, transacciones, escenario comunicativo: transporte, prototipo de interacción, funciones comunicativas, nociones léxicas, inventario básico

AGRADECIMIENTOS

Son muchas las personas que de alguna manera han colaborado en la realización de este trabajo y me han animado a llevar este proyecto hasta el final.

Gracias a la Dra. Mariona Taulé por confiar plenamente en mí, por guiarme en una disciplina nueva, la lingüística de corpus, y por todas sus indicaciones y sus correcciones minuciosas en beneficio de mi aprendizaje.

Gracias a la Dra. Lourdes Díaz por su implicación plena en el proyecto, por sus orientaciones y su motivación, por sus ánimos y por acompañarme desde hace ya unos años en mi formación como investigadora en el campo de la enseñanza del español como lengua extranjera.

Gracias al grupo de investigación Clic: a Montse por ayudarme con los programas informáticos y por sus sugerencias con los datos; a Toni por interesarse por el progreso de esta investigación y a Ricardo por el soporte técnico.

Gracias a Francisco por su paciencia, por entenderme y recibirme siempre con una sonrisa. Gracias, también, por ayudarme con las imágenes y la maquetación de la tesis.

Gracias a mis padres y a María por su apoyo incondicional y creer siempre en mí.

Gracias a Susana, Linda, Meri, Núria, Glòria y Carmen por escucharme, por estar siempre cerca y dispuestas a animarme a acabar este proyecto.

Gracias a Raquel por ayudarme a llegar al final.

Gracias muy especialmente a Yolanda T., Manuel H., M. Luisa B., Marga M., Mari Luz H., Silvia T., Paula del B., Laia E., Ana A., Susana F, Francisco P. y a los alumnos de *Arcadia University*, de la Universidad Pompeu Fabra y de la EOI Barcelona-Drassanes que han colaborado desinteresadamente en este proyecto.

Gracias a los estudiantes que han pasado por mis clases durante estos años: Craig, Olivier, Barbara, Kathy, Hsin, Kayoko, Alina, Francois, Yeun Hee, entre otros. Ellos me han inculcado la curiosidad necesaria sin la que no podría haber llegado hasta el final de este proyecto.

ÍNDICE DE CONTENIDOS

Índice de tablas	XIV
Índice de figuras	XVI
Índice de abreviaturas	XVIII
CAPÍTULO 1. INTRODUCCIÓN	1
1.1. Motivación de la presente investigación	5
1.2. Contexto de la investigación	9
1.2.1. Lingüística aplicada y lingüística de corpus	9
1.2.2. Los corpus en la enseñanza de segundas lenguas: estudiar la lengua en su uso	11
1.2.3. Aplicaciones de los corpus en la enseñanza y el aprendizaje de una segunda lengua	12
1.2.4. Análisis funcional en la enseñanza de lenguas	17
1.2.5. Análisis funcional en los corpus orales	24
1.3. Objetivos e hipótesis	26
1.4. Estructura de la investigación	28
CAPÍTULO 2. ANTECEDENTES Y ESTADO DE LA CUESTIÓN	29
2.1. Modelos nativos: Metodología de corpus orales semiespontáneos y espontáneos para la descripción de variedades de español	30
2.2. Modelos no nativos: Metodología de corpus de aprendices (<i>learner corpus</i>)	47
2.2.1. Corpus orales de aprendices de español como LE y su metodología	51
2.3. Modelos nativo y no nativo: El enfoque psicolingüístico de corpus y su metodología	60
CAPÍTULO 3. METODOLOGÍA	63
3.1. Diseño para la creación del corpus <i>FerrovíELE</i>	64
3.2. Elicitación y obtención de datos orales para la creación del corpus <i>FerrovíELE</i>	66
3.2.1 Modelo nativo: <i>InfoTren-N(ativos)_spontaneous</i>	66
3.2.2. Elicitación de datos: diseño de la tarea por simulación	68
3.2.3. <i>FerrovíELE-L(Learners)_no trained</i>	70
3.2.4. <i>FerrovíELE-L(earners)_trained</i> y <i>FerrovíELE-L(earners)_trained multilingual</i>	72
3.2.5. <i>FerrovíELE-N(ativos)_simulated</i>	74

3.3. Grabación, transcripción y anotación de los diálogos	75
3.3.1. La transcripción ortográfica	76
3.3.2. La anotación del corpus <i>Ferroviele</i>	78
3.3.2.1. Codificación morfosintáctica y lematización	79
3.3.2.2. Información sobre los diálogos	83
3.3.2.3. Identificación de los interlocutores	83
3.3.2.4. Codificación pragmático-funcional	85
3.3.2.4.1. Definición de las etiquetas	87
A. Etiquetas de las funciones comunicativas	87
B. Etiquetas de las nociones específicas o léxicas	96
C. Etiquetas para fenómenos suprasegmentales, paralingüísticos y pragmático-discursivos	105
3.3.2.5. Proceso de anotación	107
CAPÍTULO 4. ANÁLISIS DE DATOS: MORFOLOGÍA	109
4.1. Análisis de los datos morfológicos	110
4.1.1. Breve descripción del etiquetario morfológico utilizado	114
4.1.2. Análisis de la muestra sub6	116
4.1.3. Análisis de la muestra sub16	122
4.2. Análisis del vocabulario	126
CAPÍTULO 5. ANÁLISIS DE DATOS: FUNCIONES	141
5.1. Sistematización de modelos de interacción a partir del análisis de los datos (<i>bottom-up</i>)	141
5.2. Los tres modelos prototípicos desde el punto de vista funcional	143
5.3. Los tres modelos prototípicos desde el punto de vista nocional	150
5.4. Análisis del grado de especificidad de las funciones en los distintos subcorpus: funciones de dominio general y de dominio específico (transporte)	162
5.5. Efecto del entrenamiento en los subcorpus	166
5.6. Caracterización sociopragmática de los subcorpus a partir de las funciones comunicativas que implican uso de la cortesía: un ejemplo de aplicación	169
5.6.1. Los actos de habla y la cortesía	170
5.6.2. Los indicadores de cortesía en las peticiones: Hipótesis	174
5.6.3. Indicadores de cortesía: Resultados	177
5.6.3.1. Fórmulas que reflejan el contacto social	177
5.6.3.2. Peticiones de servicio: cortesía a través de la modalidad y de la indirección	178
5.6.3.3. La cooperación como indicador de cortesía	186

CAPÍTULO 6. APLICACIONES E IMPLICACIONES DIDÁCTICAS	191
6.1 Convergencia y divergencia en el <i>MCER</i>	191
6.2. La propuesta del <i>MCER</i> en el <i>PCIC</i> y en los manuales didácticos inspirados en él	194
6.2.1. El <i>PCIC</i>	194
6.2.2. Lo manuales didácticos	198
6.3. Los escenarios como salida a la divergencia	201
6.4. La propuesta de aplicación de esta tesis	209
CAPÍTULO 7. CONCLUSIONES	221
7.1. Contribuciones de esta tesis	224
7.2. Líneas futuras	225
REFERENCIAS BIBLIOGRÁFICAS	227
ANEXOS	I
ANEXO I. Epígrafes del <i>MCER</i>	I
ANEXO II. Diseño de la actividad <i>role-play</i> 1: diálogo tren	V
ANEXO III. Diseño de la actividad <i>role-play</i> 2: para entrenar	IX
ANEXO IV. Cuestionario contacto lingüístico: institución y alumnos	XIII
ANEXO V. Nociones simples	XVIII
ANEXO VI. Nociones combinadas	XVIII
ANEXO VII. Un diálogo en diferentes formatos: diálogo <i>FerrovieLE-L_trained multiling_1</i>	XVIII
VII.a. Formato Word	XVIII
VII.b. Formato html	XIX
VII.c. Formato txt	XXV
VII.d. Formato xml	XXVI
VII.e. Análisis morfológico	XXXIV
ANEXO VIII. Un ejemplo de diálogo de cada subcorpus en formato html	XXXIV
ANEXO IX. Inventario de funciones y nociones del <i>PCIC</i>	XXXV

ÍNDICE DE TABLAS

CAPÍTULO 2. ANTECEDENTES Y ESTADO DE LA CUESTIÓN

Tabla 1. Corpus orales discursivos de hablantes nativos de español	45
Tabla 2. Corpus orales de aprendices de español como L2/LE	59

CAPÍTULO 3. METODOLOGÍA

Tabla 3. FerroviELE corpus: resumen de las características de los subcorpus	65
Tabla 4. Recuento de <i>tokens</i> , <i>types</i> y lemas	82
Tabla 5. Características de los diálogos e interlocutores	85

CAPÍTULO 4. ANÁLISIS DE DATOS: MORFOLÓGICO

Tabla 6. Corpus <i>FerroviELE-sub6</i> : datos morfológicos	112
Tabla 7. Subcorpus <i>FerroviELE-L_trained-multi-sub16</i> e <i>InfoTren-N_spontaneous-sub16</i> : datos morfológicos	113

CAPÍTULO 5. ANÁLISIS DE DATOS: FUNCIONES

Tabla 8. Modelos de transacción: funciones comunicativas	145
Tabla 9. Modelo de transacción 1: funciones comunicativas y naciones léxicas para hacer una reserva	154
Tabla 10. Modelo de transacción 2: funciones comunicativas y naciones léxicas para comprar un billete	157
Tabla 11. Modelo de transacción 3: funciones comunicativas y naciones léxicas para pedir información	160
Tabla 12. Funciones de dominio general y específico para la muestra sub6	164
Tabla 13. Funciones de dominio general y específico para la muestra sub16	164
Tabla 14. Funciones de dominio general y específico por roles para la muestra sub6	165
Tabla 15. Funciones de dominio general y específico por roles para la muestra sub16	165
Tabla 16. Efecto del entrenamiento (grupos simulados)	167
Tabla 17. Diferencias entre nativos espontáneos y nativos simulados	167
Tabla 18. Diferencias entre no nativos, con o sin entrenamiento	167
Tabla 19. Efecto del entrenamiento con la tarea <i>role-play</i>	168
Tabla 20. Efecto del entrenamiento entre las FDG y las FDE en los subcorpus de nativos espontáneos y no nativos B2 para la muestra sub16	169
Tabla 21. Blum-Kulka (1987:133) <i>Table 1</i> : <i>Example of the nine Request Categories used in experiments 1-4</i>	175
Tabla 22. Modalidad léxica para la muestra sub6	180

Tabla 23. Inflexión verbal como modalidad por roles para la muestra sub6 (I)	181
Tabla 24. Inflexión verbal como modalidad para la muestra sub6 (II)	183
Tabla 25: Fórmulas indirectas para preguntar por rol para la muestra sub6	185
Tabla 26. Repeticiones a media distancia y larga distancia para la muestra sub6	187
Tabla 27. Resumir o reformular la información para la muestra sub6	189
Tabla 28. Resumir o reformular la información para la muestra sub16	189

CAPÍTULO 6. APLICACIONES E IMPLICACIONES DIDÁCTICAS

Tabla 29. Cuadro 2. Niveles comunes de referencia: cuadro de autoevaluación. MCER, capítulo 3, epígrafe 3.5	202
Tabla 30. Adecuación sociolingüística. MCER, capítulo 5, epígrafe 5.2.2	203
Tabla 31. Parámetros generales del escenario transacción de información y compra-venta de billetes de tren para E/LE (B1-B2)	210
Tabla 32. Can Do statements (Puedo hacer) en el escenario transacción de información y compra-venta de billetes de tren para E/LE (B1-B2)	211
Tabla 33. Funciones comunicativas en el escenario transacción de información y compra-venta de billetes de tren para E/LE por niveles (B1-B2)	213
Tabla 34. Recursos discursivos en el escenario transacción de información y compra-venta de billetes de tren para E/LE por niveles de lengua por niveles (B1-B2)	215

ÍNDICE DE FIGURAS

CAPÍTULO 3. METODOLOGÍA

Figura 1. Distribución de los (sub)corpus y las metodologías empleadas	63
Figura 2. Desarrollo del corpus y procesos de anotación	78
Figura 3. Repertorio de etiquetas funcionales	88
Figura 4. Repertorio de etiquetas nocionales	96
Figura 5. Etiquetas nocionales especiales	104

CAPÍTULO 4. ANÁLISIS DE DATOS: MORFOLÓGICO

Figura 6. Número total de <i>tokens</i> en sub6	117
Figura 7. Número total de <i>types</i> en sub6	117
Figura 8. TTR en sub6	117
Figura 9. Mrep en sub6	117
Figura 10. Número total de <i>tokens</i> en sub16	123
Figura 11. Número total de <i>types</i> en sub16	123
Figura 12. TTR en sub16	123
Figura 13. Mrep en sub16	123
Figura 14. Número total de lemas en sub6	127
Figura 15. Repertorio de lemas compartidos por los cinco subcorpus: frecuencia total (I)	128
Figura 16. Repertorio de lemas exclusivos de los nativos simulados	130
Figura 17. Repertorio de lemas exclusivos de los nativos espontáneos	132
Figura 18. Repertorio de lemas exclusivos de los aprendices B2 entrenados Multilingües	134
Figura 19. Repertorio de lemas exclusivos de los aprendices B1 entrenados USA	135
Figura 20. Repertorio de lemas exclusivos de los aprendices B1 no entrenados USA	135
Figura 21. Repertorio de lemas compartidos por los cinco subcorpus: frecuencias distribuidas por subcorpus (II)	138

CAPÍTULO 5. ANÁLISIS DE DATOS: FUNCIONES

Figura 22. Distribución funcional	142
Figura 23. Rasgos y secuencias	147
Figura 24. Funciones comunicativas de la situación de transacción	149
Figura 25. Nociones asociadas a las funciones	161
Figura 26. Enfoque del estudio: desde una perspectiva macropragmática a una perspectiva micropragmática	174

CAPÍTULO 6. APLICACIONES E IMPLICACIONES DIDÁCTICAS

Figura 27. Inventario de las nociones específicas para el ámbito de Viajes, alojamiento y transporte (I) del <i>PCIC</i> (2007)	197
Figura 28. <i>Aula 4. Nueva edición</i> (B1.2). Editorial Difusión (2014)	199
Figura 29. <i>PRISMA. Avanza</i> (B2). Editorial Edinumen (2004)	199
Figura 30. <i>Destino Erasmus 2. Niveles intermedio y avanzado</i> (B1-B2). Editorial SGEL (2008)	199
Figura 31. <i>ALTE Can Do statements</i>	202
Figura 32. <i>Core Inventory</i> para el inglés, British Council y EAQUALS (North et al. 2010: 10-11)	205
Figura 33. <i>Appendix B: Scenario: on holiday</i> , British Council y EAQUALS (North et al. 2010: 26)	207
Figura 34. <i>Appendix C: Mapping Text Types</i> , British Council y EAQUALS (North et al. 2010: 36-37)	208
Figura 35. <i>Appendix D: Scenario: Mapping Language Content</i> , British Council y EAQUALS (North, B. et al. 2010: 38-40)	208
Figura 36. Appendix E: Exponents for Language Content (B1), British Council y EAQUALS (North, B. et al. 2010: 52-53)	209

ANEXO I

Figura 37. Descriptores del <i>MCER</i> : Epígrafe 4.4.3.1	I
Figura 38. Esquema general del proceso de la compra de bienes y servicios	III
Figura 39. Descriptores del <i>MCER</i> : Epígrafe 5.1.2. Las destrezas y las habilidades	IV

ANEXO IX

Figura 40. Inventario de las funciones del <i>PCIC</i> (2007)	XXXV
Figura 41. Inventario de las nociones específicas para el ámbito de Viajes, alojamiento y transporte (II) del <i>PCIC</i> (2007)	XXXVII

ÍNDICE DE ABREVIATURAS

El contenido de esta tesis presenta abreviaturas para hacer más ágil la lectura de ciertos términos que se van repitiendo. Las abreviaturas que se han empleado son las siguientes:

C: Cliente

E/LE: Español como Lengua Extranjera

EL2: Español Lengua Segunda

DDL: *Data-Driven Learning*

FDE: Funciones de Dominio Específico

FDG: Funciones de Dominio General

IL: Interlengua

LA: Lingüística Aplicada

LC: Lingüística de Corpus

LE: Lengua Extranjera

LO: Lengua Objeto

L2: Lengua Segunda

L1/LM: Lengua materna

MCER: Marco Común Europeo de Referencia

Mrep: Media de repetición

M1: Modelo 1 de transacción

M2: Modelo 2 de transacción

M3: Modelo 3 de transacción

O: Operador

PCIC: Plan Curricular del Instituto Cervantes

TTR: Type/Token Ratio

W/P strategies: *Wish/Possibly queries*

CAPÍTULO 1. INTRODUCCIÓN

En la actualidad, la enseñanza de lenguas extranjeras defiende un modelo de lengua y una metodología orientadas a la acción, es decir, a emplear la lengua “para llevar a cabo tareas comunicativas, los usuarios tienen que realizar actividades de la lengua de carácter comunicativo y poner en funcionamiento estrategias de comunicación” (*Marco Común Europeo de Referencia* 2001: 67). En general, los manuales disponibles de español como lengua extranjera (en adelante E/LE¹) recogen este componente situacional en su variedad de prácticas comunicativas, integrando elementos de conversaciones coloquiales. El problema de dichos materiales es que presentan una lengua artificial, ya sea escrita u oral (las audiciones), no basada en datos de producción real. No muestran conversaciones coloquiales reales, sino que recrean situaciones de la vida cotidiana de cuestionable verosimilitud (o relativa artificialidad) que no siempre coinciden con lo que ocurre en la realidad.

Para ayudar a paliar esta situación, una estrategia fundamental es el recurso a la consulta directa o indirecta de corpus reales para obtener un modelo de la competencia nativa básica en un tipo de situación o escenario comunicativo concreto, es decir una muestra representativa de textos (escritos u orales) producidos en situaciones reales. Una consulta directa hace referencia al acceso directo a un corpus existente. Una consulta indirecta, en cambio, hace referencia a la consulta de trabajos basados en corpus o a la consulta o uso de materiales derivados de corpus.

¹ Es conveniente hacer algunas precisiones de conceptos básicos como lengua nativa lengua segunda y lengua extranjera. El término lengua nativa, según el *Diccionario de términos claves de E/LE del Centro Virtual Cervantes* lo identifica como lengua primera (L1), lengua de origen o lengua materna (LM): “lengua materna o L1 se entiende la primera lengua que aprende un ser humano en su infancia y que normalmente deviene su instrumento natural de pensamiento y comunicación (...) Quienes comparten una lengua materna son considerados hablantes nativos de la lengua en cuestión, p. e., hispanohablantes nativos”. De modo que, a los efectos de la siguiente investigación, se siguen las consideraciones del *Centro Virtual Cervantes*.

Por otro lado, aunque los términos lengua segunda (L2) y lengua extranjera (LE) se distinguen, se han usado ambos conceptos para tratar indistintamente el proceso de adquisición y aprendizaje en el país donde se habla la lengua que se adquiere.

Con el objetivo de contribuir a acercar las muestras a la realidad en cualquiera de los dos tipos de consulta, en esta tesis se realiza un estudio de corte empírico de la lengua hablada, basado en el análisis de un escenario comunicativo real para el español: la transacción de información en el ámbito de transporte en tren para pedir información y comprar billetes (interacción cara a cara), de acuerdo con el estándar propuesto en el *Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza y evaluación*² (Consejo de Europa 2001, en adelante *MCER* o *Marco*) y ensayado en el *A Core Inventory for General English*³ (North et al. 2010) para el inglés. Para ello, se extrae el modelo prototípico de la competencia básica en dicho ámbito, obtenido a partir del análisis de corpus orales producidos por nativos de español y corpus de aprendices (no nativos) para describir la variabilidad lingüística nativa y no nativa de este escenario comunicativo.

El *MCER* es un documento que fija las directrices fundamentales para la enseñanza de las lenguas extranjeras. EL *MCER* estandariza, por un lado, los parámetros para definir las etapas de aprendizaje (A1, A2, B1, B2, C1, C2) —dimensión vertical— y, por otro, los parámetros para describir las categorías desde las que se describen los procesos de aprendizaje y uso de las lenguas —dimensión horizontal—. El cruce de estas dos dimensiones proporciona una descripción de lo que los aprendices-usuarios de idiomas deberían ser capaces de hacer y cómo hacerlo, a partir de tareas que incluyen actividades comunicativas de la lengua en todo el proceso de aprendizaje (Llorián 2007).

El *MCER* se refiere a “las tareas” como las actividades concretas de la vida cotidiana, que corresponde a los intereses de los aprendices y a las necesidades de la socialización, de naturaleza diferente, como por ejemplo, escribir cartas de solicitud de trabajo y asistir a entrevistas de trabajo ofreciendo información hablada. La realización de “tareas” se enfoca en la “acción” como el motor que moviliza las estrategias, los

² El *MCER* es un proyecto que se desarrolla dentro de la política lingüística del Consejo de Europa y entre sus objetivos se destaca el siguiente: Desarrollar un enfoque de la enseñanza de las lenguas basado en principios comunes para todas las lenguas de Europa.

³ British Council y EAQUALS (*European Association for Quality Language Services*) —agencia para garantizar la calidad en las instituciones de enseñanza de lenguas extranjeras— han trabajado juntos para crear un *core inventory* o un *inventario básico* de funciones, gramática, vocabulario, etc. para la enseñanza del inglés como lengua extranjera según los niveles del *MCER* (<<http://eaquals.org>>).

procesos, las competencias, las habilidades y los recursos necesarios para aprender y usar una lengua dentro de un contexto.

Una vez establecidos los niveles de referencia, el español fue entre todas las lenguas europeas la primera que concretó y desarrolló las indicaciones del *MCER* con el *Plan Curricular del Instituto Cervantes* (2006) (*PCIC*, en adelante). Posteriormente, *EAQUALS* (*European Association for Quality Languages Services*) y el British Council con su *A Core Inventory for General English* (North et al. 2010) ofrecen pautas a partir de la interrelación de los contenidos (textos y destrezas, textos y lengua, entre otros) del *MCER* a partir de la enseñanza del inglés como lengua extranjera (en adelante LE).

Ambas propuestas, tanto la del Instituto Cervantes como la del British Council, parten de los planteamientos del *MCER*, pero el desarrollo de la propuesta que se hace para el inglés dista mucho de la del español. La diferencia básica es que el *Core Inventory* integra las destrezas y las competencias a partir de la definición de la lengua como “muestras de diferentes escenarios comunicativos” y realiza un inventario de algunos escenarios concretos (sin ser exhaustivo). Se entiende por situación o escenario comunicativo un contexto concreto y cotidiano de la vida, por ejemplo, ir al médico, hacer la compra, quejarse por un mal servicio, etc., en el que se produce un intercambio comunicativo. A partir de un escenario determinado se prevén transacciones y/o interacciones posibles, tanto orales como escritas. Estos escenarios se presentan en forma de *inventarios básicos* que proporcionan un repertorio de los rasgos lingüísticos, pragmáticos y estratégicos más comunes para cada nivel de usuario. Cada escenario se delimita, a su vez, a partir de la descripción del dominio, el contexto, las tareas, las actividades y el texto.

De lo anteriormente expuesto, cabe entender que un *inventario básico* es una recopilación de la información, en términos de competencia comunicativa, que se considera más relevante para el aprendiz en el aula de lenguas extranjeras. “*The Core Inventory does not aim to provide detailed guidance for course book or examination developers, for who detailed specifications already exist in the series Waystage, Threshold Level and Vantage Level. (...) It has been created with teachers and learners in mind. Each language point appears at the level(s) at which it is considered of most relevance to the learner in the classroom” [el subrayado es nuestro] (North et al. 2010: 8).*

En el ámbito del E/LE, el *PCIC* procede de forma distinta: sistematiza objetivos y contenidos para analizar y describir los diferentes componentes —gramatical, pragmático-discursivo, nocional, cultural y de aprendizaje— de la lengua española, ofreciendo repertorios independientes de estos componentes, no interconectados en una situación comunicativa concreta (como EAQUALS). Asimismo, para E/LE todavía no se ha desarrollado un *inventario básico* y creemos que proporcionarlo sería muy útil para los docentes, porque facilitaría su labor en la creación de materiales y en la evaluación de las actuaciones, es decir, de la competencia lingüística de los aprendices. Igualmente, estos inventarios aplicados al español proporcionarían al *PCIC* los repertorios interrelacionados y conectarían sus especificaciones.

En esta investigación se propone contribuir a complementar este panorama, aportando un inventario básico obtenido a partir del análisis de un escenario específico para el español implementado en dos aulas distintas, con anglófonos (nivel B1) y con multilingües (nivel B2) para sustentarlo empíricamente. La tarea es una transacción que tiene lugar durante la compra de billetes de tren, en interacciones cara a cara.

Se parte de escenarios porque permiten contextualizar los intercambios, atendiendo a los elementos *extralingüísticos* (participantes, género, edad, clase social, guion, etc.) de acuerdo con las *necesidades comunicativas* y con la interrelación de destrezas (expresión oral y escrita, comprensión oral y escrita).

Para llevar a cabo la descripción lingüística de este escenario se sigue una metodología basada en corpus, en concreto, se parte de corpus orales producidos por nativos de español y corpus orales de aprendices (no nativos) de español para describir la variabilidad lingüística nativa y no nativa en el escenario comunicativo seleccionado y cubrir los niveles B1, B2 y C2⁴ (un grupo de control nativo).

Se extraen los datos lingüísticos del corpus nativo para la documentación en el proceso de creación del inventario básico del escenario seleccionado, y así, se evitan las muestras de origen introspectivo y, a la vez, se establece una norma —entendido el término como lo hiciera A. Bello (1981), el más normal de todos los usos—. Además,

⁴ En esta investigación el nivel C2 se equipara al nivel de un nativo (véase Ortega 2010).

en esta investigación, se amplía el estudio del escenario seleccionado con el corpus oral de aprendices para ilustrar las dificultades o la distancia respecto a la transacción de información y compra-venta de billetes de tren por los nativos ante el mismo escenario.

Los dos corpus, tanto el nativo como el de aprendices, están acotados a partir de las siguientes características:

a) se incluye un tipo de discurso oral: diálogos o conversaciones;

b) se recoge el registro coloquial y

c) se incluye una situación comunicativa concreta y se delimita a partir de la función de transacción —se conversa para transmitir datos, no hay una motivación fática, caso en el que tan solo se conversa para mantener el vínculo con el interlocutor—.

Tanto el escenario como el corpus creado en este estudio son independientes de cualquier enfoque metodológico en la enseñanza de lenguas extranjeras —enfoque por tareas, enfoque natural, enfoque procesual, entre otros— como indican los investigadores de EAQUALS (cf. documento *A Core Inventory for General English* (North et al., 2010: 20-21)).

Un beneficio posterior previsible es el resultado del análisis de datos y sus implicaciones didácticas. Teniendo los rasgos más comunes del escenario estudiado, se abrirán también vías para la elaboración de propuestas pedagógicas, que no son objeto de este trabajo. Sin embargo, la indicación de las diferencias entre nativos y no nativos —que se pondrán de manifiesto a lo largo del estudio—, y entre los propios nativos, son un *input* importante para todo trabajo aplicado posterior, en particular para la elaboración de repertorios por niveles de competencia lingüística y por especificidad de tratamiento del componente pragmático-sociocultural.

1.1. Motivación en la presente investigación

El docente, usuario de materiales didácticos, se enfrenta en el aula a situaciones comunicativas recogidas en estos desde un punto de vista, por lo general, de impresiones personales. En cierto modo, el profesor se basa en su experiencia y en

sus propios ejemplos —que parten de su propio idiolecto— para completar los materiales o para producirlos, esto es, aborda las cuestiones de lengua en uso desde la introspección. Por otro lado, muchos de los manuales existentes presentan muestras de lengua de cuestionable verosimilitud “per se”. Ambos tipos de datos, los del docente y los de los manuales, para los aprendices desarrollan, pues, modelos de lengua *top-down*. En otras palabras, aportan muestras de lengua introspectivas e individuales (orales/escritas), que distorsionan el modelo comunicativo y satisfacen artificialmente las necesidades comunicativas. Por ello, se considera necesario que tanto profesores como diseñadores de materiales se apoyen en el uso de los corpus para superar esos déficits de realidad y verosimilitud, lo que redundará en que tanto las clases como los manuales ganarán calidad como *input* incorporando evidencia empírica. El modelo se transformará en *bottom-up* o lo que es lo mismo, partirá de la realidad lingüística contrastada cuantitativamente para ilustrar y ofrecer solo las regularidades que aparecen en los corpus (que son datos reales) conformando modelos de uso real, que se puedan aplicar a las descripciones didácticas organizadas por niveles.

El objetivo de este trabajo de investigación es acometer la descripción y el análisis de la situación comunicativa transacción de información y compra-compra de billetes de tren en interacción cara a cara a partir de muestras reales de la lengua para ofrecer una aproximación empírica a los patrones y los exponentes lingüísticos que la conforman. Para ello, se crea el corpus oral *Ferroviele*, formado por diálogos espontáneos y simulados, con participantes nativos y no nativos, que se enmarca en el área de aplicaciones para la enseñanza y aprendizaje de lenguas extranjeras. *Ferroviele* consta de diferentes subcorpus, de nativos y no nativos (aprendices) en lengua española peninsular. Como se ha indicado, *Ferroviele* desarrolla una situación comunicativa concreta para recoger una muestra representativa de la variabilidad lingüística nativa y no nativa desde una aproximación *bottom-up*, esto es, con muestras empíricas. El corpus está anotado, revisado y etiquetado morfológica y pragmático-funcionalmente.

A partir de los datos pragmático-funcionales de los subcorpus nativo y no nativo se describen los modelos de uso real de la lengua (funciones, nociones, etc.) de

ambos subcorpus, de tal modo que el profesor y el creador de materiales dispongan de un repertorio lingüístico-funcional con lemas e información sociopragmática basada en corpus⁵, esto es: con un enfoque sociopragmático para el aula. Por consiguiente, los datos del corpus sirven para realizar la descripción de un inventario básico (*core grammar*) de la situación comunicativa ampliada o escenario que se estudia, es decir, de la transacción de información y compra-venta de billetes en el ámbito del transporte de tren.

En relación con los niveles del *MCER* previstos, este inventario básico de la situación comunicativa analizada detalla los componentes lingüísticos, pragmáticos y estratégicos para el nivel intermedio (B1/B2), contrastados y ampliados a partir del vaciado y análisis de los corpus nativo y no nativo. En los inventarios resultantes se analiza qué elementos son prototípicos de los nativos y cuáles de los no nativos y, a su vez, en qué elementos convergen los dos grupos (nativo y no nativo). De esta forma, es viable realizar una propuesta empírica de descriptores lingüístico-pragmáticos del nivel intermedio, B1-B2, a partir de los elementos analizados.

Esta investigación resulta, sin duda, innovadora, puesto que propone una nueva manera de organizar los contenidos lingüísticos, pragmáticos y estratégicos en E/LE, a la luz de su predecesora no empírica de inventarios por escenarios. Si bien la *core grammar* de EAQUALS ofrece la información organizada o bien por funciones y niveles de lengua o bien por escenarios (con todas las destrezas integradas) de carácter general o guía, la *core grammar* derivada de *FerrovieLE* organiza los contenidos a partir de los modelos de referencia obtenidos de los corpus exclusivamente orales (nativos y no nativos) de una tarea, lo que permite definir y acotar en la destreza oral las (sub)competencias (léxica, sintáctica, pragmática, etc.)

⁵ Tognini-Bonelli (2001) distingue dos formas de aproximarse al uso de los corpus. Distingue, por un lado, la aproximación metodológica *corpus driven* o conducida por el corpus en la que se da prioridad a los aspectos cuantitativos en la recogida de datos y en la que no hay ningún tipo de tratamiento de análisis lingüístico. Los datos del corpus son la única información que se tiene en cuenta.

Por otro lado, la aproximación *corpus-based*, basada en un corpus, en la que se usa el corpus para ejemplificar la teoría que se conoce de antemano y se presupone un modelo previo de representación de los datos. En esta aproximación los corpus se anotan con información lingüística (características morfológicas, sintácticas, semánticas o pragmáticas) y esta información se puede recuperar automáticamente. Estos corpus suelen tener dimensiones más pequeñas, porque el proceso de anotación está condicionado por la intervención humana (Martí y Taulé 2014: 178).

más allá de los objetivos teóricos, afinando en los niveles (B1, B2, C2 —variante nativa—).

Esta manera de organizar (inventarios por escenarios) fue el anclaje al *MCER* que se hizo para la lengua inglesa y como modelo para ilustrar una perspectiva integradora (sociopragmática) en la enseñanza y aprendizaje como lengua extranjera. La diferencia entre nuestra propuesta y las propuestas del British Council/EAQUALS reside en que la nuestra utiliza datos de corpus a partir de un diseño experimental y, así, ofrece una guía para profesores y creadores de materiales en que se proporcionan y se describen los datos a partir de muestras reales. Se consigue un modelo de competencia nativo desde una perspectiva *bottom up*.

Eso sí, no hay que obviar cierta parcialidad del corpus *FerrovieELE*, que se salva con los datos de los dos grupos de hablantes nativos, uno de control y otro simulado.

De todo lo anterior se desprende que el uso de corpus lingüísticos es rentable por ofrecer un banco de datos reales de la lengua. Tanto los profesores como los creadores de materiales se beneficiarán de este recurso, que pueden aplicar para recabar elementos y estructuras que aparecerán en las actividades que diseñen o en los manuales didácticos a partir de criterios parecidos a los que se pueden encontrar en el *English Profile. The CEFR for English*⁶, esto es, por nivel, categoría, destreza, entre otros. Esta forma de proceder encaja con la cada vez más frecuente tesis de especialización, como la de Römer (2004), que defiende la idea de enseñar las estructuras más frecuentes para que las producciones de los aprendices de una L2 estén más próximas a la lengua materna de los hablantes nativos. Sin embargo, esta postura también tiene la restricción de que remite a un modelo de “*core* no contextualizado” en oposición al “*core* contextualizado”, enfoque que defendemos en este trabajo. En otras palabras, el primero, ofrece unidades léxicas, colocaciones, expresiones idiomáticas, etc. con alta frecuencia, recogidos en diferentes textos que tratan diversos ámbitos y situaciones; mientras que el segundo describe los elementos más frecuentes pero acotados a textos que desarrollan una situación comunicativa específica (un dominio determinado).

⁶ <<http://englishprofile.org/english-grammar-profile/egp-online>>

En definitiva, usar una metodología basada en corpus lingüísticos de modelo nativo y de aprendices —*learner corpora*— evita la introspección y proporciona datos nucleares reales de una lengua. Römer (2004: 197) para el inglés señala:

“a lot of corpus-driven work still has to be done to reach the aim of enabling both pupils and teachers to learn and teach an English which is more authentic and closer to that of native speakers”.

Si esta afirmación es adecuada para el inglés, también se puede extrapolar para la lengua española y así mejorar la calidad de los materiales docentes (muestras, criterios de evaluación, *syllabus*, etc.) gracias al rigor empírico.

1.2. Contexto de la investigación

Esta investigación se centra en dos disciplinas distintas: la lingüística aplicada y la lingüística de corpus (1.2.1) con la finalidad de utilizar los corpus como metodología para estudiar la lengua en su uso (1.2.2) y describir las aplicaciones de los corpus en la enseñanza de un LE (1.2.3). Asimismo, en esta sección hablamos del análisis funcional en la enseñanza de LE (1.2.4) y en los corpus orales (1.2.5), puesto que va a ser la perspectiva de análisis que aplicaremos al corpus creado para esta investigación.

1.2.1. Lingüística aplicada y lingüística de corpus

Los estudios sobre lingüística de corpus (en adelante, LC) destacan los beneficios que el corpus lingüístico ha supuesto para otras áreas de la lingüística actual: traducción, lexicografía y, en los últimos años se ha empezado a implementar en el ámbito de enseñanza y aprendizaje de lenguas extranjeras y con fines didácticos de la enseñanza de la lengua en general.

Se entiende por corpus cualquier conjunto textual (escrito u oral) de más de un texto. No obstante, cuando este concepto se aplica al campo de la lingüística su significado está más matizado e incluye rasgos que lo delimitan de forma más concreta.

En efecto, la definición de arriba se puede complementar con la de corpus lingüístico que es “un conjunto de textos informatizados producidos en situaciones reales, que se han seleccionado siguiendo una serie de criterios lingüísticos explícitos que garantizan que dicho corpus pueda ser usado como muestra representativa de la lengua” en Alonso Pérez-Ávila (2007). En esta definición se destacan cuatro aspectos clave que se deberían esperar en los textos que se agrupan en un corpus: i) textos informatizados; ii) textos producidos en situaciones reales; iii) criterios concretos para la selección de los textos del corpus y iv) representatividad de los textos.

Los avances informáticos han permitido tener acceso a los textos escritos y orales (diálogos transcritos). Concretamente, desde la década de los 60 los corpus informatizados han colaborado en el desarrollo de la investigación lingüística, permitiendo el manejo de gran cantidad de datos y el estudio de la lengua, integrada en el contexto discursivo con ejemplos reales. Por lo tanto, la LC permite estudiar el lenguaje-e (*external language* o actuación), es decir, el uso del lenguaje, frente a la introspección y los métodos inductivos, que se centran en el estudio del lenguaje-i (*internal language* o la competencia).

Por otro lado, la LC aplicada a la enseñanza y aprendizaje de lenguas extranjeras, en concreto en el ámbito anglosajón, lleva desarrollándose hace ya años. Su tradición investigadora tiene un largo desarrollo en los ámbitos de la lexicología, la lexicografía, la edición de materiales y la didáctica (Cruz 2012: 99). En cambio, el uso de los corpus en la enseñanza del español como LE se presenta tardíamente, dando resultados positivos.

Es interesante observar cómo la comunidad investigadora de didáctica en E/LE muestra un interés cada vez mayor en el uso de corpus de lengua real, producida de manera natural para mejorar las actividades didácticas. Por ello se puede apreciar “*an important link between the two previously disparate fields of corpus linguistics and foreign/second language research*” (Granger 2002: 1). Emplean el estudio basado en corpus lingüísticos como metodología para obtener muestras que sirvan como fuente para obtener datos empíricos sobre el uso real de la lengua. Por tanto, para esta

investigación, siguiendo a McCarthy&Carter (1995), Sinclair (2004), McEnery et al. (2006)⁷, vamos a considerar los corpus lingüísticos una metodología de estudio.

Asimismo, la didáctica de lenguas extranjeras con corpus ha sido presentada en congresos internacionales como “Teaching and Language Corpora” (TaLC) (Römer 2008) y, en territorio latinoamericano y español, existe una presencia cada vez mayor de comunicaciones en congresos de ámbito nacional, como por ejemplo, de las asociaciones ASELE (Asociación para el Español como Lengua Extranjera) y AESLA (Asociación Española de Lingüística Aplicada) (Cruz 2012).

Desde estas circunstancias, todo esto parece indicar que las investigaciones con corpus serían necesarias en cualquier estudio relacionado con las lenguas, puesto que ofrecerían mejores descripciones lingüísticas tanto de hablantes nativos como de hablantes no nativos. Estos últimos son los llamados corpus de aprendices, que en el siguiente capítulo 2 y en parte del capítulo 3 se examinan con detalle.

1.2.2. Los corpus en la enseñanza de segundas lenguas: estudiar la lengua en su uso

Para el profesor de una LE es importante concienciar a los estudiantes de que aprender una lengua consiste en ser competentes comunicativamente, término propuesto por Gumperz y Hymes (1964), dentro de la antropología social y cultural, en el nuevo marco para el estudio del lenguaje: la etnografía de la comunicación. Para estos autores, la competencia comunicativa da al usuario de la lengua la herramienta para saber si una producción:

- es posible: se ha emitido a partir de unas determinadas reglas, relacionadas tanto con la gramática de la lengua como con la cultura de la comunidad de habla;
- es factible: si las condiciones normales de una persona (memoria, percepción, etc.) permiten emitirla, recibirla y procesarla correctamente;

⁷ Parodi (2008) y Gries (2010) cuestionan si la lingüística corpus se trata de una metodología o una teoría en sí misma, con la revisión de las posturas de diferentes autores. Consideran la LC como una teoría: Leech (1992), Stubbs (1993), Tognini-Bonelli (2001), Teubert (2005); otros consideran la LC como una metodología: McEnery&Wilson (1996), Meyer (2002). No obstante, no nos detendremos en este debate.

-es apropiada: si se adecua a las variables que pueden darse en las distintas situaciones de comunicación;

-se da en la realidad: si resulta posible formalmente, factible y apropiada, es efectivamente usada por los miembros de la comunidad de habla. Según Gumperz y Hymes (1964), “puede que algo resulte posible, factible, apropiado y que no llegue a ocurrir”⁸.

El cuarto y último criterio es un punto de arranque, así como uno de los objetivos de esta investigación: una producción ha de tener carácter real.

Por otro lado, el lingüista Firth (1956) planteaba la necesidad de estudiar el habla en su contexto lingüístico y en su contexto situacional. En otras palabras, estudiar el habla en relación a los elementos no verbales, personas, objetos y sucesos que dan significado a las formas lingüísticas (Zanón 2007: 4). Los dos conceptos, arriba resaltados en negrita, forman los principios básicos de la lingüística de corpus: el significado de las palabras va ligados al uso, de este modo el estudio lingüístico deberá basarse en el análisis del uso lingüístico.

Tras la herencia del contextualismo de Firth, la lingüística sistémico (lengua)-funcional (habla)⁹ de Halliday (1973) parte también del análisis de la lengua en su uso:

“Cuanto más capaces seamos de relacionar las opciones sobre los sistemas gramaticales con los potenciales de significado en los contextos sociales y comportamentales, más cerca nos encontraremos de vislumbrar la naturaleza del sistema lingüístico” (Halliday 1973: 39 en Zanón 2007: 17).

1.2.3. Aplicaciones de los corpus en la enseñanza y el aprendizaje de una segunda lengua

Los corpus lingüísticos se usan en esta investigación así como en otros estudios para la enseñanza y aprendizaje de una lengua extranjera como metodología o recurso

⁸<http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/competenciacomunicativa.htm>

⁹ La unificación de los conceptos de lengua/habla por parte de Halliday dio comienzo a la pragmática.

y, a su vez, la disciplina de enseñanza de una LE utiliza los corpus lingüísticos en función de unos objetivos de estudio.

Estas herramientas metodológicas que son los corpus lingüísticos para el estudio y análisis de la adquisición de una lengua extranjera pueden tener distintas aplicaciones: bien indirectas o bien directas (Leech 1997; Römer 2008). La **aplicación indirecta** significa que los corpus ayudan a tomar decisiones sobre qué enseñar y cómo enseñarlo. Así pues, son una fuente de información para los investigadores y los materiales didácticos, ofreciendo las ocurrencias más frecuentes y los modelos de lengua de diferentes situaciones comunicativas de la vida real. Estos datos deberían influir en el diseño de los programas de enseñanza de lenguas e, igualmente, en los manuales de lenguas extranjeras. Este tipo de enfoque se aproxima a los datos de aprendizaje desde dos perspectivas, desde los *corpus-based* y los *corpus-driven*, si bien es cierto que se aprecian más investigaciones en torno al enfoque *corpus-based*.

En cambio, la **aplicación directa** se refiere a que los aprendices y los profesores tienen acceso directo a los corpus para que sirvan de apoyo en el aprendizaje de la LE/L2. En este caso, los profesores y los alumnos trabajan directamente con los corpus para descubrir ellos mismos modelos de lengua, concordancias, palabras y frases de manera autónoma. Este recurso de aprendizaje en que se aplican los corpus para trabajar la gramática y el vocabulario en clase se conoce con el nombre de *data-driven learning* (DDL) donde profesores y alumnos participan activamente.

El DDL tiene un gran potencial pedagógico (McCullough 2001) y es un complemento en las clases de LE, ya que ofrece información junto a los manuales y a los diccionarios. Hay diversas tareas que se pueden realizar con ellos como, por ejemplo, corrección de errores, observar modelos de lengua, comparar expresiones idiomáticas en la LE estudiada y en su LM (por ejemplo, colocaciones donde los alumnos pueden observar estas en diferentes contextos; al mismo tiempo que pueden observar la variación del significado de una palabra según la variación contextual) o usar corpus, que recogen discursos especializados en un área de conocimiento concreto para utilizarlos en la enseñanza del idioma extranjero con fines específicos (negocios, medicina, entre otros).

Asimismo, los corpus ofrecen *input* de calidad en el ámbito de la enseñanza y aprendizaje de LE. Esto es, los corpus pueden aportar muestras reales (*input*) a las explicaciones o a materiales creados por el profesor. Si, por el contrario, el profesor recurre siempre a sus propios ejemplos que tienen origen en su propio idiolecto, las clases no tendrán la misma calidad en cuanto al *input* que sí se apoya en el uso de los corpus. Con el uso de los corpus en las clases se podría ofrecer un *input* variado y construido desde diferentes fuentes (orales/escritas, lengua formal/informal, etc.). De este modo, se podrían tener ejemplos de situaciones prototípicas de uso —como es el caso de este trabajo—, combinaciones de términos y características pragmáticas.

De lo anteriormente expuesto se deduce la importancia del *input* en el proceso de aprendizaje. De hecho, Krashen (1985) con la “hipótesis del input”¹⁰ afirma que:

“para que haya adquisición de una LE es imprescindible atender a la calidad del input, esto es, a la calidad de los datos de entrada a los que está expuesto el que adquiere una lengua. Las muestras de esa lengua deben ser interesantes, comprensibles y no secuenciadas gramaticalmente. Podemos adquirir más lengua, o más conocimiento implícito —inconsciente— de una LO sólo cuando los datos de entrada son comprensibles, es decir, cuando estamos expuestos a unos datos que contienen unas estructuras que son sólo un poco más difíciles, o van un poco más allá de nuestro nivel de competencia” (en Baralo 1998: 56).

Esta hipótesis considera que las personas que adquieren (no aprenden) una LE centran su foco de atención en el significado y, como consecuencia, adquieren también las estructuras gramaticales. Por esta razón, se cree en la no manipulación deliberada de los datos de entrada para que sus estructuras sean apropiadas y cumplan la fórmula “i+1”. Es decir, tiene que ser de un nivel ligeramente superior al que se encuentre el alumno. En este sentido, los textos de los corpus disponen de ese *input* interesante y comprensible que Krashen alude y que los aprendices pueden

¹⁰ La hipótesis del input se encuentra dentro del Modelo del Monitor, que contiene una propuesta de cinco hipótesis entre las que se encuentra la del input o “i+1”.

recibir. Con este *input* que proporcionan los corpus, se pueden crear hipótesis sobre las estructuras y las reglas gramaticales de la nueva lengua que se estudia.

En general, el uso de los corpus con un enfoque didáctico tiene implicaciones en el rol del alumno y del profesor. Por un lado, el alumno participa de manera autónoma. Se arriesga ante las dudas que resolverá interactuando con un corpus. Por otro lado, el profesor se presenta como experto del aprendizaje, no del lenguaje. En este sentido, facilita estrategias de aprendizaje, guía a sus estudiantes y favorece la interacción (Bernardini en Sinclair 2004: 27-28).

Igualmente, los corpus ayudan a aplicar en el aula los fundamentos del enfoque comunicativo para la enseñanza y el aprendizaje de lenguas extranjeras porque, i) usa el lenguaje significativamente; ii) favorece la autonomía del estudiante frente al profesor y, en consecuencia, la motivación; iii) trabaja tanto la forma como el significado e iv) introduce variación en el proceso de aprendizaje.

Por otro lado, como se ha señalado más arriba, los corpus pueden tener una aplicación indirecta, que son aquellos que se emplean para la investigación en el contexto de la enseñanza y aprendizaje de lenguas extranjeras. Aquí aparecen los estudios con corpus de aprendices donde se estudia la interlengua de los aprendices.

El término de interlengua (en adelante, IL), introducido por Selinker (1972) hace referencia al sistema lingüístico de los hablantes no nativos en el proceso de aprendizaje y adquisición de una lengua extranjera, en contraste con el sistema de lengua materna del aprendizaje.

La evolución de los fundamentos en torno a la IL ha evolucionado según las teorías lingüísticas y de aprendizaje vigentes. Al principio, la investigación se centraba en los aspectos más gramaticales de la IL y, después, se pasó a estudiar la IL desde distintos paradigmas: el análisis del discurso, la sociolingüística y la pragmática, que se concreta en: i) el Análisis Contrastivo; ii) el Análisis de Errores; iii) el Análisis de la Actuación y la IL o modelos de frecuencia y iv) el Análisis del Discurso.

Entre estos enfoques, en la actualidad, los estudios de corpus de aprendices pueden seguir dos líneas: por un lado, el análisis contrastivo de la IL y, por otro, el análisis de errores asistido por ordenador. El primero centra el análisis en la comparación entre las producciones de los no nativos con las de los nativos, es decir,

con un corpus de control, que está formado por textos de características similares. Si se define el estudio en el ámbito de una L1 los textos del corpus de control serán producidos por adultos; en cambio, si el área de estudio es una L2/LE los textos serán producidos por nativos, que es el campo de investigación de este proyecto. Se comparan los dos corpus, el de aprendices y el de control para observar la frecuencia de una estructura concreta o para obtener un uso estándar de cada nivel de lengua de los aprendices; o se realiza una comparación entre las producciones de no nativos con diferentes L1 para ver cómo funciona la transferencia lingüística. Este tipo de análisis no necesita el etiquetado de errores, pero en algunos estudios los corpus se anota morfológicamente para hacer comparaciones del uso de categorías entre nativos y no nativos.

Por su parte, los recientes estudios, concretamente en E/LE, se han centrado principalmente en la frecuencia, en el uso de las colocaciones Orol y Alonso Ramos (2013), distribución de pronombres en la interfaz sintaxis-discurso Lozano (2009), combinaciones de palabras Pino (2012), entre otros.

El segundo modelo de estudio, el análisis de errores asistido por ordenador etiqueta los errores (*error-tagged*) con las técnicas de la lingüística de corpus no para ver las carencias lingüísticas que presentan sino para recuperar las incorrecciones a partir de criterios de búsqueda y sacar frecuencias de errores. Algunos proyectos, creados desde la metodología del análisis de errores con distintas etapas (Corder 1971), son *el Corpus para el Análisis de errores de aprendices de E/LE (CORANE)*¹¹ (Cestero et al. 2002); el trabajo de tesis y los artículos derivados de esta (Campillos 2014a, 2014b); el mal uso de las colocaciones (Alonso et al. 2010), entre otros. De este modo, con la metodología de corpus se evitan afirmaciones intuitivas de las realizaciones de los aprendices y se busca la objetividad en el análisis de los datos.

Otro de los enfoques mencionados para el estudio de corpus de aprendices es el modelo de Análisis del Discurso que relaciona los enunciados con el contexto y entiende el lenguaje como uso. Desde estos paradigmas, se basa el análisis del corpus *FerrovíELE*, que centra el estudio en la conversación y en la interacción. Para ello se

¹¹ Se trata de un corpus de redacciones escritas por alumnos de los cursos de lengua del año 2000 de la Universidad de Alcalá.

anotan las funciones comunicativas que se han empleado en los diferentes diálogos, producidos en un escenario comunicativo específico.

1.2.4. Análisis funcional en la enseñanza de lenguas

Las funciones comunicativas habitualmente suelen estar presentes en la mente de los profesionales en el ámbito de LE. Si se revisan diversos índices de manuales para la enseñanza de E/LE, vemos que se refieren a ellas o bien con el término explícito funciones, o bien con la expresión “recursos comunicativos” (en el manual *Aula 4. Nueva edición*, figura 28). Sea como fuere, al detallar las funciones pueden ser recogidas desde macro-ámbitos (*expresar opiniones, pedir y dar informaciones, expresar opinión...*) o desde actos más puntuales (*explicar cómo se llega a un lugar, preguntar cuánto dura algo...*).

En cualquier caso, se entiende por funciones en didáctica de segundas lenguas aquellas categorías de clasificación conectadas al uso comunicativo de la lengua, de modo que estas sirven para describir la lengua de estudio desde las intenciones y las acciones comunicativas. Son de naturaleza pragmática y de acuerdo con Matte Bon (2004: 811) los contenidos funcionales en el sector de enseñanza de idiomas tratan sobre “los diferentes actos sociales que un enunciador puede llevar a cabo en la lengua extranjera (también llamados funciones): presentarse, preguntar la hora, pedir permiso, ofrecer, aceptar, rechazar, etc.”. No obstante, si nos referimos al término función desde el punto de vista de la Lingüística, se observa que las acepciones son variadas. Si bien todas convergen en entender la lengua desde una perspectiva pragmática. Cuando nos referimos al término funcional se puede pensar en: i) las funciones del lenguaje que proponía Jakobson (1963), que describe los elementos que intervienen en la comunicación —el emisor (función expresiva), el receptor (función conativa), el mensaje (función poética), el referente (función referencial), el código (función metalingüística) y el canal (función fática)—; ii) el funcionalismo de Martinet (1970) o iii) la propuesta de funciones de Halliday (1973) que concibe las funciones como herramientas de comunicación en sociedad.

En general, estas acepciones en estos términos no se emplean en la enseñanza de lenguas extranjeras. Si bien es cierto que los parámetros funcionales en LE, que deben tenerse en cuenta en la programación de cursos de lenguas hoy en día, derivan de las teorías de las funciones de Halliday (1973) y de los actos de habla de Searle (1969). Las listas de funciones que conocemos actualmente se detallaron en los años 70 en *The Threshold Level* (van Ek y Trim, 1990) para el inglés. Este documento incluye listados de funciones, basados en las intenciones de los interlocutores en diferentes situaciones sociales entre otros componentes, para definir el nivel mínimo de dominio de un estudiante de LE. El documento para el español, *Nivel Umbral*, se publicó posteriormente (Slagter 1979). También el *PCIC* (2006), el documento de referencia que concreta el *MCER*, incluyó el componente funcional para detallar el dominio en los diferentes niveles de lengua.

Como se ha podido observar el término función tiene ya una cierta trayectoria. Veamos, a continuación, cómo apareció en la enseñanza de idiomas esta dimensión nocio-funcional con los programas comunicativos que se desarrolló en los años 70, en torno al proyecto del Consejo de Europa con la finalidad de fomentar el aprendizaje de lenguas extranjeras. En este enfoque se encuentra el primer intento de diseñar un modelo comunicativo.

En la década de los 60 y los 70 se destaca el estudio del lenguaje como comunicación, y se presentan trabajos innovadores desde la psicología, la antropología y la filosofía del lenguaje. Ello influyó en la enseñanza de idiomas, que se centró más en la competencia comunicativa que en el conocimiento de las estructuras.

Desde el campo de la filosofía del lenguaje, Searle publica *Speech Acts* (1969) e introduce el concepto de acto de habla, que no es del todo nuevo. Ya en 1962 lo propuso Austin en *Cómo hacer cosas con palabras*. De hecho, Searle lo retoma para defender el doble componente del lenguaje como conocimiento y acción con efecto en el mundo. Un enunciado emitido por un hablante, como mínimo, desarrolla cuatro actos diferentes:

1. acto de emisión: emitir palabras;
2. acto proposicional: referir y predicar sobre elementos de la realidad;
3. acto elocutivo: ordenar, prometer, preguntar, negar... y
4. acto perlocutivo: persuadir, convencer, alegrar, etc.

El lenguaje no sólo es un conjunto de reglas sintácticas, semánticas y fonológicas, también necesita el contexto para crear sus producciones. Aprender a hablar es una actividad para la que se han de conocer las reglas y se han de aprender a utilizarlas en un contexto de comunicación (Zanón, 2007).

En el campo de la sociolingüística (etnografía del habla), Hymes (1972) propone sustituir el concepto de competencia lingüística¹² de Chomsky (1965) por el de competencia comunicativa. Chomsky diferenciaba *la competencia* —conocimiento lingüístico implícito que tiene el hablante nativo de su lengua, dejando de lado los aspectos individuales, sociales y culturales— y *la actuación* —producción e interpretación lingüística que cada hablante hace de las reglas. Hymes integra en la propuesta de Chomsky el aspecto social del uso de la lengua y su adecuación al contexto.

Otra teoría lingüística emparentada que defiende el lenguaje como instrumento de interrelación social es la descripción funcional de Halliday (1973) que sustenta de forma evidente las descripciones de Wilkins (1976). Halliday describe siete funciones: instrumental, reguladora, interactiva, personal, heurística, imaginativa y representativa.

El enfoque nocio-funcional planteaba el análisis de las intenciones comunicativas que un estudiante de LE podía producir o entender. Estructuraba la enseñanza en unidades y para determinar dichas unidades se apoyaba en el documento de Wilkins (1972) que daría lugar al libro *Notional Syllabuses* (Wilkins 1976), que tuvo una gran influencia en la didáctica de idiomas de corte comunicativo en Europa.

Este enfoque organizaba el aprendizaje en unidades o formas lingüísticas, formadas por funciones y nociones, que a su vez daban lugar a los exponentes o realizaciones (*¿Quieres que te+pres. subj.?* es el exponente de la función *ofrecerse para hacer algo*). Su programa proponía listados de funciones y nociones. Este listado, en principio, es igual en todas las lenguas. Sin embargo, las realizaciones lingüísticas son únicas de cada lengua.

¹² Dicho concepto aparece en la publicación *Syntactic Structures* (Chomsky, 1965).

Respecto a los exponentes, claramente, su forma interna es compleja y va más allá de las estructuras gramaticales. Están integrados por diversos elementos de tipo morfológico, sintáctico o léxico. A grandes rasgos los exponentes se asemejan a las oraciones aunque en función de la situación e intención comunicativa pueden variar. Otros factores que hacen cambiar los exponentes consisten en si la producción es oral o escrita, el estilo de los interlocutores, la interpretación de cada exponente, entre otros. Así pues, una realización puede aplicarse a diversas funciones o una función puede aparecer bajo diferentes exponentes. El grado de variabilidad es una característica de los exponentes. Por ello, se pueden encontrar fórmulas habitualmente fijas (como por ejemplo, *buenos días*) y, también, existen otras realizaciones que pueden variar (como por ejemplo, añadir otros elementos o los elementos de una parte pueden usarse en otros exponentes y, así cambiar de función).

En la enseñanza del E/LE el enfoque nocio-funcional tuvo un gran peso, en los años 80, a la hora de organizar los contenidos (*syllabus*) de los cursos de lengua. Si bien es cierto que el programa comunicativo evolucionó y dio pie a otros enfoques en la enseñanza del E/LE, por ejemplo el enfoque por tareas, que se sitúa más en el proceso, es uno de ellos (hoy en día en la enseñanza de lenguas coexisten diferentes enfoques). De este modo, actualmente, resulta difícil encontrar un enfoque claramente funcional. No obstante, parece que los planteamientos del enfoque nocio-funcional se han implantado en documentos de referencia y manuales. Pues, en principio, un buen número de libros para enseñar E/LE organizan sus unidades en base a funciones de la lengua o competencias pragmáticas. Asimismo, el *MCER* contiene parámetros de referencia nocio-funcionales. De hecho, el antecedente del *MCER* fue el *Threshold Level* (van Ek 1990).

Sin lugar a dudas, la dimensión comunicativa de la lengua en relación con la enseñanza comunicativa de esta tiene una base teórica funcional de gran solidez en la que se fundamenta el *MCER*. Por eso, se describe la dimensión del uso social de la lengua y se da una constante tensión entre forma y significado en el *MCER*.

En la dimensión horizontal¹³ —una de las organizaciones propuesta por el *MCER*— se articulan los parámetros y las categorías que posibilitan la comunicación a partir de los procesos de aprendizaje y el uso de las lenguas. Es decir, presenta un enfoque orientado a la acción. En ella se tratan objetivos comunicativos a partir de tareas lingüísticas comunicativas (acciones centradas en la consecución de un objetivo con efecto ilocutivo y/o perlocutivo). Estas tareas exigen del usuario un análisis de los factores del contexto (situación, participantes, condiciones, etc.) y a su vez relacionan diferentes planos de la vida social (personal, público, educativo, etc.) (Goffman, 1959). Mientras se realizan las tareas, el usuario emplea sus competencias generales y comunicativas de la lengua, a través de líneas de actuación conscientes y reguladas, en otras palabras, hace uso de las estrategias. De modo que la actuación lingüística comunicativa se concreta en actividades comunicativas que generan textos, entendidos en sentido amplio como producciones orales y escritas, por ejemplo, escribir cartas, mantener conversaciones, etc. (Llorián, 2007).

La competencia descrita tiene que medirse en el proceso de aprendizaje del individuo: la evaluación. Dado que en dicha competencia se integran diversos saberes serán evaluados todos: tanto los componentes lingüísticos como los componentes cognitivos (saber), afectivos (saber ser), procedimentales (saber hacer) y metacognitivos (saber aprender).

Los componentes saber, saber ser, saber hacer y saber aprender forman las competencias generales y están relacionados indirectamente con la lengua, pero se activan cuando se interactúa con miembros de la cultura en la que se habla la lengua. El currículo de lenguas puede considerar estos saberes dentro de los objetivos, contenidos, metodología y evaluación.

Entre todos los saberes nos interesa el saber hacer, que es la capacidad potencial del usuario de manifestar conductas. El *MCER* distingue dos tipos de destrezas o habilidades: las prácticas (sociales, de la vida, profesionales y de ocio) y las interculturales. Es decir, por un lado, la destreza práctica en el ámbito social, por ejemplo, capacitará al usuario para mantener la distancia física adecuada durante los

¹³ El *MCER*, además, cuenta con la dimensión vertical donde se estructuran los niveles y cursos. Son los dos ejes que equilibran el documento.

intercambios conversacionales con los hablantes nativos. Por otro lado, la destreza intercultural enseña al usuario a actuar con adecuación en las convenciones sociales de la comunidad en la que se habla la lengua, y le ayudará a superar prejuicios que lleven a malentendidos culturales (ver en el anexo 1, figura 39, los descriptores para ambos componentes). Para cada nivel de competencia (inicial, intermedio, avanzado) esos saberes se reflejan en comportamientos o en producción descrita con detalle en los saber hacer, que suponen la interacción de los dos niveles o planos mencionados al principio del párrafo.

Esta investigación se enmarcaría en la intersección, por una lado, del área de la dimensión horizontal que trata la competencia pragmática dentro de la expresión oral, para observar la lengua desde una perspectiva funcional y las capacidades de los usuarios de una lengua para asociar las frases a contextos adecuados, y por el otro lado, del área de la dimensión vertical correspondiente a los niveles de competencia intermedios B1-B2.

La competencia pragmática en el *MCER* se plasma en dos competencias: discursiva y funcional, ambas regidas por la competencia organizativa. Es cierto que la delimitación entre dichas competencias no es sencilla, como muestra la dificultad de establecer a qué competencia responden los parámetros que determinan el uso de la cortesía¹⁴ en las interacciones, sobre todo, entre desiguales.

La competencia funcional¹⁵ la integran tres componentes:

- 1) las Microfunciones: el *MCER* las entiende como “categorías para el uso funcional de enunciados aislados (generalmente breves), normalmente como turnos de palabra de una interacción” (en el *MCER*, 5.2.3.2., dentro del punto las Microfunciones). Ejemplos de microfunciones son: *expresar y describir actitudes, persuasión, vida social*, etc. De modo que se observa una relación entre los conceptos de función comunicativa y de exponente lingüístico, que ya aparecían en el documento *Threshold Level* (van Ek, 1990). El *MCER* también

¹⁴ Trabajos de Grice (1981) y de Levinson (1983).

¹⁵ En *MCER*, apartado 5.2.3.2.

La inclusión de las escalas microfunciones y macrofunciones en el *Marco* es muy novedosa, porque cumple la función de medir o valorar la fluidez oral y la precisión (Guitérrez Quintana 2007).

hace uso de las funciones y las macrofunciones como herramienta para el aprendizaje y enseñanza de una LE.

- 2) las Macrofunciones: este concepto se relaciona al de secuencia discursiva en el texto escrito o en el discurso hablado (descripción, narración, instrucción, etc.), en el que hace hincapié el análisis del discurso (Stubbs, 1987).
- 3) los Esquemas de interacción: son unidades superiores donde se incluyen funciones y exponentes, por ejemplo *compras*, cuyo esquema de interacción general reproduce el *MCER* en la sección de actividades de interacción oral¹⁶ (ver anexo 1, figura 37). Dichos exponentes del cuadro 1 describen la interacción oral en una transacción. Es decir, trata los modelos de intercambio verbal y la capacidad de utilizarlos. En cuanto a la descripción del procedimiento, el *MCER* parte del par adyacente (pregunta-respuesta; afirmación-acuerdo/desacuerdo, etc.) para describir el desarrollo de los intercambios y la articulación de turnos. Sin embargo, las características de la conversación espontánea no se recogen para ningún nivel. Lo que sí ofrece es la descripción global de la conducta en el saber hacer (ver anexo 1, figura 39). En el *MCER*, el ámbito de los intercambios dialógicos transaccionales no aparece. El documento recoge únicamente un esquema general (etología o comportamiento) que se observa en el proceso de la compra de bienes y servicios del capítulo 8 del *Threshold Level 1990*¹⁷ (ver anexo 1, figura 38, segunda mitad).

El descriptor¹⁸ (bandas de criterios y comportamientos observables por nivel) del *MCER* para la tarea *comprar* es deductiva, y por supuesto, puede estar sujeta también a variación entre nativos. De ahí que una parte de la propuesta de esta investigación consista en hacer un estudio empírico de las transacciones a partir del

¹⁶ En *MCER*, apartado 4.4.3.1.

¹⁷ En el apartado 5.2.3.2. La competencia funcional, en el punto 3, Esquemas de interacción.
<http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cap_05.htm#p523>

¹⁸ Un descriptor es la explicación, formulada en términos de conducta observable, de la actuación satisfactoria (lingüística y no lingüística) del individuo en una determinada situación de comunicación. Se formula normalmente en pequeñas descripciones que son bandas criterios establecidas para cada uno de los niveles de competencia previstos en el *MCER*.

análisis real, basado en un corpus de nativos en un escenario concreto, para poder sistematizar un ejemplo de la función de transacción más próximo a la expresión oral en situaciones reales. El estudio se centrará solo en las tareas orales del escenario seleccionado, para acotar la variabilidad de los exponentes únicamente en el ámbito oral.

1.2.5. Análisis funcional en los corpus orales

El interés por la competencia pragmático-discursiva en el campo de la adquisición de LE se ha estudiado desde el enfoque del Análisis del discurso o desde la Pragmática interlingüística. La metodología basada en un análisis de las funciones ofrece resultados cualitativos que sirven para observar el proceso de adquisición de una LE.

En esta investigación nos centramos en las funciones comunicativas del lenguaje, que se basan en principios pragmáticos (*quejarse, expresar la opinión, dar consejo*), si bien es cierto se pueden aplicar otros enfoques de estudio desde el análisis funcional del significado (referencia específica/no específica), del significado y de la gramática (sujeto/agente y objeto/paciente) y del discurso (tópico/tema).

El método de análisis funcional en el ámbito de aprendizaje y enseñanza de LE se emplea para estudiar la variabilidad de la lengua. Dentro de este tipo de análisis, primero, cabría recoger un análisis funcional de la interlengua española de los aprendices y, segundo, se haría necesario comparar los resultados ya obtenidos con la producción nativa como punto de referencia lingüístico.

La variabilidad de los nativos, principalmente, se ha estudiado desde el análisis del discurso, por un lado, y el análisis de la conversación, por otro. El análisis del discurso analiza aspectos comunes y diferentes de la producción en un diálogo específico; mientras que el análisis de la conversación estudia la variación que se produce en un conjunto de diálogos. Ambos enfoques estudian la conversación nativa en función de una metodología basada en corpus lingüísticos orales (véase apartado 2.1). Por otro lado, es habitual que las investigaciones de variabilidad no nativa sigan una metodología basada en el análisis de errores.

En vista de lo antes expuesto, no deja de llamar la atención que los planteamientos de los estudios para describir la variabilidad nativa son bien diferentes de los establecidos para analizar la variabilidad no nativa. Las características que surgen del análisis de la producción nativa son diferentes de los habituales estudios de la interlengua para los aprendices. Por ello, si se acepta el principio de interrelación entre la producción nativa y no nativa, hemos de analizar cómo se desarrolla esa conexión: se hace necesario, entonces, plantear un tipo de análisis donde ambos grupos compartan un mismo principio de estudio.

De este modo, se ofrece como planteamiento el análisis funcional comunicativo, puesto que en el área de aprendizaje de LE sirve de eje organizador de las clases y de los *syllabus*. Esta metodología de estudio se toma desde dos perspectivas diferentes, que atiende a cuál sea el objetivo de análisis: a) de la función a la forma y/o b) de la forma a la función (Ellis-Baruiken 2005; Ellis 1992).

En el primer tipo de análisis, de la función a la forma, se identifican las funciones comunicativas (por ejemplo, *expresar planes e intenciones*) y el estudio de estas; después, nos llevará a examinar las formas lingüísticas que se emplean en la función seleccionada para obtener el significado de la función (por ejemplo, *pienso+inf.*; *mi propósito es...*; *tengo la idea de+inf.*). El segundo tipo, de la forma a la función, el análisis parte a partir de la dirección opuesta: se observa una forma gramatical específica (por ejemplo, verbos modales como *quiero, puedo, prefiero*), a continuación, se pasa a verificar el significado específico que esta forma desarrolla en la muestra lingüística, es decir, en la función comunicativa. Estos dos tipos de análisis son complementarios. Es más: este planteamiento de análisis funcional de la interlengua de los aprendices proporciona una herramienta muy útil para estudiar cómo: i) los aprendices utilizan sus recursos lingüísticos para resolver sus necesidades comunicativas y ii) las necesidades comunicativas que dirigen al proceso de la adquisición de una LE.

Estas dos perspectivas de análisis de la interlengua, de la función a la forma y de la forma a la función, son las que se aplicarán para el análisis de los datos obtenidos del corpus *FerrovieLE* (véase capítulos 4 y 5).

1.3. Objetivos e hipótesis

El objetivo general de esta tesis es, pues, la implementación de un escenario comunicativo concreto —transacción de información y compra-venta de billetes (interacción cara a cara)— para la obtención de muestras empíricas que permitan la descripción del inventario lingüístico propio de la función de transacción en la lengua hablada, destinado al aprendizaje del español como lengua extranjera o como segunda lengua para un nivel intermedio (B1-B2 del *MCER*). El trabajo resultante consistirá en un documento de referencia que liste de forma organizada, a partir del concepto de escenario, la competencia pragmlingüística materializada por niveles que van del B1 al B2 circunscritos a la destreza oral que permita llevar a cabo las diferentes etapas de la transacción de información y compra-venta de billetes. El documento resultante, por tanto, recogerá exponentes, nociones y funciones como el modelo del EAQUALS y el British Council (North et al. 2010), de forma que ofrezca una perspectiva general del contenido lingüístico más relevante de la situación comunicativa analizada, con vistas a su aplicación posterior en la didáctica del español como lengua extranjera. Servirá para acotar, más allá del *MCER* y de las concreciones del *PCIC*, cómo enseñar y aprender la lengua y los recursos sociopragmáticos necesarios en un escenario y en un nivel concreto, sin constreñirse a ninguna metodología de enseñanza concreta.

El punto de partida de toda la investigación es el modelo de inventario básico del proyecto EAQUALS-British Council para el inglés como L2. Hasta la fecha, no existe nada parecido para el español como L2 en el ámbito de las *core-grammars*. Tampoco existe antecedente de trabajo basado en corpus para los inventarios básicos del inglés (Thornbury 2011).

Para conseguir este objetivo de base empírica se parte del estudio de corpus lingüísticos producidos por hablantes nativos y no nativos en el que se apoye la descripción del escenario seleccionado para la enseñanza de E/LE.

Dicho objetivo se concreta en los siguientes objetivos más específicos:

- (1) Explotación de un corpus nativo basado en un escenario de transacción para modelizar los estadios en que consiste y llevar a cabo su descripción pormenorizada.

- (2) Análisis lingüístico de ese escenario concreto a partir de la información extraída del corpus oral de nativos con el objetivo de proponer una caracterización minuciosa de las transacciones.
- (3) Sistematización y etiquetado funcional de las estructuras y las características de dichas transacciones para aplicarlas en la propuesta del inventario básico y de las recomendaciones didácticas (como en EAQUALS).
- (4) Obtención y análisis del corpus oral de aprendices para observar y describir cómo ejecutan ellos las transacciones en el mismo escenario trabajado por los nativos.
- (5) Detección de las carencias y de los aspectos de dificultad de los aprendices (variabilidad respecto a los nativos) para considerarlos por niveles de competencia (B1, B2) en la propuesta del inventario básico y de las recomendaciones didácticas (como en EAQUALS).

Para acometer estos objetivos específicos se usa una metodología del análisis del discurso de corte pragmático y se fundamenta en corpus reales de producción oral nativa y no nativa de español peninsular. Asimismo, estos objetivos se sustentan en diferentes **hipótesis**:

- (1) Trabajar con corpus proporciona modelos de lengua más verosímiles: su mayor valor cualitativo viene avalado por los datos cuantitativos. Por consiguiente, basar empíricamente la descripción lingüística de los manuales didácticos y las explotaciones didácticas contribuye a mejorar la calidad de ambos.
- (2) El análisis funcional basado en corpus (*bottom-up*) permitirá refinar las categorías funcionales, además de las sintácticas y las léxicas a lo largo de los niveles B1, B2 y C2 (variante nativa). Este rasgo no es evidente en el *PCIC*.
- (3) El entrenamiento en la tarea (en este caso en interacción) tiene un efecto cualitativamente distinto en la mejora de la actuación lingüística de los aprendices de nivel B2 en comparación con los aprendices de nivel B1.
- (4) Las diferentes percepciones sociopragmáticas ligadas a la L1 del alumno en relación con diferentes rasgos (por ejemplo, el uso de las funciones en el escenario de estudio, las peticiones con el empleo de diversos modales, entre otros) hace que haya variación sociopragmática, asociada al uso distinto de los

elementos lingüísticos que se emplean para esta situación comunicativa en las diferentes culturas.

- (5) Ciertos rasgos habrán de ser discriminatorios y permitirán distinguir a los nativos de los no nativos y a los de nivel B1 de los del B2, C2 (grupo nativo).

1.4. Estructura de la investigación

La presente tesis se divide en 7 capítulos. En primer lugar, se desarrolla la parte teórica donde se incluyen **antecedentes y estado de la cuestión** (capítulo 2) de los corpus lingüísticos orales. En esta sección se realiza una revisión detallada y actualizada de los corpus disponibles para español LE y su metodología. En segundo lugar, se encuentra el núcleo de la tesis: se describe la **propuesta metodológica** (capítulo 3) que se ha desarrollado en este trabajo de tesis doctoral para la obtención del corpus *FerrovíELE*, que se fundamenta en los principios teóricos tratados en la primera parte de la tesis.

A continuación, se presenta el **análisis de datos: morfológicos y funcionales** (capítulos 4 y 5). Se describen los principales resultados sistematizados (morfosintácticos, funcionales y pragmático-discursivos) y se discuten.

A raíz del análisis de los resultados obtenidos, a partir del estudio detallado del corpus *FerrovíELE*, se propone una **aplicación e implicación didáctica** (capítulo 6). Aquí se establecen qué elementos se recomiendan trabajar en la situación comunicativa de estudio para los niveles B1, B2 (en algunas secciones se incluye también la variante nativa (C2)). Para finalizar se recogen las **conclusiones** (capítulo 7) que se derivan de toda la tesis doctoral, así como qué aspectos se deberían seguir investigando.

El volumen concluye con la **bibliografía** empleada en este estudio y unos **anexos** donde se ofrecen descriptores de algunas escalas del *MCER*; el diseño de las actividades *role-play*; el cuestionario de contacto lingüístico; ejemplos de diálogos del corpus *FerrovíELE* en cada uno de los niveles de análisis y tablas de lemas y de búsquedas de nociones.

CAPÍTULO 2. ANTECEDENTES Y ESTADO DE LA CUESTIÓN

El propósito de este capítulo es examinar los corpus orales del español — y usados para aplicaciones de E/LE en algún momento—en tanto que método empírico para aproximarse a la lengua oral como fuente de datos fiables para describir y modelizar el uso lingüístico real de la lengua hablada. Estos corpus son muestras representativas de la producción oral en situaciones semiespontáneas, espontáneas o experimentales, elaboradas de acuerdo con distintos parámetros (temática, características lingüísticas de los informantes, situación social, situación de aprendizaje, conocimiento de otras lenguas, nivel de dominio de estas, entre otros aspectos). Los corpus que se describen a continuación son relevantes para la enseñanza del español, si bien de entre ellos se eligen los más convenientes para nuestro propósito: el análisis de la interlengua española en la interacción oral.

En este capítulo se describen, además de los corpus de E/LE seleccionados, sus metodologías tanto de recogida (que condiciona que sean espontáneos, semiespontáneos o totalmente experimentales), como de análisis (fonológico, morfológico, de errores, sintáctico, pragmático) y explotación en función de la disciplina en que se enmarquen (psicolingüística, adquisición infantil, análisis del discurso, adquisición de lenguas segundas, procesamiento del habla, etc.).

Este capítulo, por tanto, se organiza en tres secciones que reflejan la clasificación de corpus y metodologías de análisis de estos según los siguientes criterios. En la primera sección (2.1) se describen los corpus de producción oral nativa espontánea, semiespontánea o de entrevista semiguída, cuyos datos se han usado para aplicaciones de E/LE y/o cuya metodología ha servido de modelo en posteriores estudios y recogidas de datos, porque incorporan herramientas y códigos del marco de la pragmática y del análisis del discurso útiles para la enseñanza de lenguas (y del E/LE en particular). En la segunda sección (2.2) se presentan los corpus no nativos, sean orales o escritos, producidos por aprendices de español, pero circunscritos al contexto de instrucción formal (es decir, no espontáneos). En la sección tres (2.3) se presentan los corpus de nativos y no

nativos en situación de uso espontáneo, fuera de contexto de instrucción, diseñados fundamentalmente desde la psicolingüística: corpus de lenguaje infantil, adultos y niños bilingües, adultos multilingües, entre otros.

2.1. Modelos nativos: Metodología de corpus orales semiespontáneos y espontáneos para la descripción de variedades de español

En este apartado se presentan los diferentes corpus y metodologías de recogida de corpus de español oral producidos por hablantes nativos más usados en aplicaciones didácticas de ELE. Fundamentalmente, nos referimos a la extracción del léxico más frecuente del español, a las colocaciones más frecuentes por áreas temáticas, a los campos semánticos, etc. Entre estos corpus cabe destacar el *Corpus del Español*¹⁹ (**CdE**) de Mark Davis (2002). Se trata de un corpus diacrónico del español peninsular y de Latinoamérica con una base de datos que agrupa textos de los siglos XII al XX. La parte de textos contemporáneos contiene 20 millones de palabras y de estas 5 millones forman el corpus oral. Cabe destacar que se pueden realizar búsquedas desde una interfaz electrónica en la página web del corpus que permite consultar listas de palabras, frases, categorías gramaticales, etc. Además, ofrece la consulta de frecuencias y los contextos de uso.

Por otra parte, para reflejar y analizar la variedad dialectal de la lengua española hay numerosos corpus de corte sociolingüístico cuyos datos se han recogido tanto en España como en Latinoamérica. Entre los corpus que recogen las distintas variedades del español peninsular más recientes destacamos el *Corpus Oral y Sonoro del Español Rural (COSER)*²⁰ (Fernández-Ordóñez 2011) que empezó a confeccionarse en 1990 y actualmente sigue realizándose. COSER está coordinado por Inés Fernández-Ordóñez, de la Universidad Autónoma de Madrid. Se trata de un corpus dialectal que recoge entrevistas semidirigidas que se acercan a la conversación del habla de zonas rurales de la Península Ibérica para estudiar la variación gramatical. Cada entrevista ofrece las variables sociolingüísticas o contextuales de los informantes: edad (una

¹⁹ <<http://www.corpusdelespanol.org/>>

²⁰ <<http://www.uam.es/coser/>>

media de edad de 70,7 años), sexo, nivel sociocultural (generalmente con baja escolarización), de zona rural y temática acotada (tradicción etnográfica). Buena parte de los diálogos se han transcrito ortográficamente con algunas convenciones propias. Otro corpus dialectal, pero menos ambicioso geográficamente, y, en cambio, orientado sociolingüísticamente de forma más amplia (más niveles culturales y variedad de edades), es el **Corpus del habla de Almería**²¹ (Carbonero, Cortés y Bañón 2005). Este corpus fue recogido por el Grupo ILSE (Estudio del discurso oral en español) y coordinado por Luis Cortés Rodríguez, de la Universidad de Almería. Contiene diversos tipos de interacciones orales (96 horas en total): conversaciones cara a cara, distintos tipos de discursos monológicos y entrevistas semidirigidas (55 horas). En las entrevistas se recoge la información contextual relativa a edad, sexo y nivel sociocultural de los informantes y se transcriben siguiendo las convenciones ortográficas y las etiquetas TEI²².

El *Proyecto de Estudio Sociolingüístico del Español de España y América (PRESEEA)* (Moreno 2005) que comenzó en 1996 y actualmente sigue desarrollándose, ha confeccionado también un corpus oral de diversas zonas geográficas latinoamericanas y españolas, coordinado por Francisco Moreno Fernández, de la Universidad de Alcalá de Henares²³. El corpus PRESEEA integra las aportaciones de diferentes equipos²⁴ de PRESEEA, que se coordinan con una metodología sociolingüística común (grabación y técnica de las entrevistas). Los corpus de PRESEEA incluyen, generalmente, entrevistas semidirigidas de unos 45 minutos, donde aparecen el entrevistador y un informante seleccionado a partir de criterios sociolingüísticos para ofrecer una muestra estratificada (edad, sexo, nivel sociocultural). Este proyecto es heredero del *Macrocorpus de la norma lingüística culta de las principales ciudades*

²¹ <<http://www.grupoilse.org/>>

²² <http://www.tei-c.org/Vault/P4/Lite/teiu5_sp.html>

²³ <<http://preseea.linguas.net>>

²⁴ Por ejemplo, el corpus PRESSEA de Málaga que recoge el proyecto ESESUMA (Estudio Sociolingüístico del Español Urbano de Málaga); el grupo PRESSEA de la Universidad de Alcalá recopila el habla de Alcalá de Henares, Madrid (barrio de Salamanca) y Madrid (barrio de Vallecas); el corpus PRESSEA de Lérida; el grupo GIESCA (Grupo de Investigación para el Estudio Sociolingüístico del Caribe) de la Universidad del Atlántico (Colombia); etc.

de España y América (**MC-NC**)²⁵ (Samper et al. 1998), dirigido por José Antonio Samper y editado en CD-ROM, que también recoge 14 entrevistas libres con intervención del encuestador. El corpus está formado por la producción de 168 informantes de nivel culto de 12 ciudades hispanas (9 son ciudades latinoamericanas y 3 ciudades españolas). El corpus tiene una duración de 84 horas. Las muestras están distribuidas por generación y por sexo. Asimismo, los diálogos están transcritos ortográficamente y se han incorporado al *Corpus de Referencia del Español Actual (CREA)*²⁶ (Real Academia Española 2004).

Otro corpus que recoge una variedad dialectal del español es el *Análisis de la conversación-Alcalá de Henares (ACUAH)* (Moreno 2001). Recogido por Ana M^a Cestero durante 1993²⁷, forma parte del corpus de Alcalá de Henares y sigue las convenciones de análisis utilizadas por PRESEEA. El corpus se compone de 18 conversaciones semidirigidas cuyos informantes se seleccionaron siguiendo un criterio de carácter sociolingüístico: nivel sociocultural medio, tres rangos de edad y la misma cantidad de hombres y mujeres. Tiene una duración de 800 minutos y algo menos de 100.000 palabras. ACUAH se etiquetó según las normas TEI y también está incluido en el CREA.

Un enfoque distinto presenta el *Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)*²⁸ (Marcos Marín 1991), cuyo objetivo es obtener una muestra representativa de distintos géneros discursivos. CORLEC fue recogido entre 1989 y 1992 por el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid y dirigido por Francisco A. Marcos Marín. Se trata de uno de los primeros corpus de habla espontánea del español y se codificó según las propuestas TEI²⁹.

²⁵ También se conoce como corpus ALFAL porque fue cedido por la Asociación de Lingüística y Filología de América Latina.

²⁶ <<http://corpus.rae.es/creanet.html>>

²⁷ Sobre estas grabaciones Ana M^a Cestero realizó su Tesis Doctoral «Alternancia de turnos de palabra en lengua española», defendida en 1994.

²⁸ <<http://www.lilf.uam.es/ESP/Corlec.html>>

²⁹ Actualmente, el Laboratorio de Lingüística Informática de la UAM está convirtiendo el CORLEC a un formato similar a C-Oral_Rom, en XML.

El corpus CORLEC es una base de datos textual de lengua hablada que contiene textos de todo tipo, desde conversaciones espontáneas hasta presentaciones más académicas (no leídas) de diferentes ámbitos: administrativo, científico, familiar, educativo, humanístico, instrucciones por megafonía, jurídico, lúdico, político, periodístico, publicitario, religioso y técnico. Está constituido por alrededor de 1.100.000 palabras y aporta la transliteración de los textos grabados en cintas de audio del registro oral. En 1997, el corpus se analizó morfológicamente y se incorporó en la parte oral del CREA.

Algunos de los corpus descritos más arriba, como el ACUAH, el MC-NC y el CORLEC se incluyen en el **CREA oral**³⁰ (Rojo 2015), que a su vez está incorporado en el **Corpus de Referencia del Español**³¹, proyecto de la Real Academia Española, dirigido por el propio Guillermo Rojo. El corpus oral recoge compilaciones de carácter heterogéneo de un periodo temporal amplio (1995-2004) y funciona como corpus de referencia de la lengua española hablada. En cuanto a la proporción entre oral y escrito y a su procedencia geográfica, el CREA recoge textos escritos y orales (los orales representan un 10%), procedentes de España y de América Latina (el 50% de uno y el 50% del otro). Estos textos se encuentran en soporte informático y se acompañan de unas herramientas implementadas para el usuario que permiten acceder a las palabras y a sus contextos. Actualmente se está realizando su continuación, el *Corpus del Español del siglo XXI (CORPES XXI)* (versión actual 0.8) que también está formado por textos orales y escritos y contiene, hasta el momento, un total de 200 millones de palabras. Si bien este otorga más peso a la lengua escrita que a la oral, a diferencia del CREA, las transcripciones se encuentran alineadas con las producciones orales. Asimismo, el peso del español de América Latina es mayor que el peninsular (el 30% de las formas proceden de España y el 70% de Latinoamérica), para reflejar de manera más realista el peso de cada variedad geográfica en el conjunto del corpus de referencia.

³⁰ Otro corpus que se cedió al CREA oral es *el Corpus Oral de la Variedad Juvenil Universitaria del Español Hablado en Alicante (COVJA)* y que se tratará más adelante en este capítulo.

³¹ <<http://corpus.rae.es/creanet.html>>

A través de su página web se pueden realizar búsquedas de palabras concretas con opciones de filtrado (cronológico, geográfico, oral/escrito, etc.), obteniendo el número de veces que aparece y sus correspondientes contextos de ocurrencia.

Con otra perspectiva distinta de la variación, en la que la situación comunicativa es el eje, se recogió el *Corpus Valencia, Español Coloquial (Val.Es.Co.)*³² (Briz et al. 2002), elaborado por el grupo de investigación Val.Es.Co. y dirigido por Antonio Briz, de la Universitat de València. Este corpus se crea con el objetivo de caracterizar el registro oral coloquial en situaciones comunicativas muy definidas (por ejemplo: conversación familiar sobre un partido de fútbol o las vacaciones) con el objetivo de estudiar la estructura de la conversación desde el punto de vista pragmático y describir el funcionamiento de sus unidades. Para ello, se ciñe a la variedad lingüística de la zona geográfica de la Comunidad Valenciana y a las variables sociológicas de sus informantes. Este corpus se diferencia de los anteriores en que las conversaciones coloquiales son siempre espontáneas, 100% secretas o semisecretas y tomadas en situaciones comunicativas reales. Se grabaron en audio de manera secreta y en el entorno familiar del informante, por lo que en cada conversación pueden participar de 2 a 4 interlocutores.

El corpus Val.Es.Co. tiene una duración de 341 horas, de las que 6,5 horas (100.000 palabras) se transcribieron y se publicaron en papel y las otras 250.000 palabras están transcritas y alineadas con los archivos sonoros³³. Para la transcripción se utiliza un sistema de códigos propios y cada diálogo va acompañado de una ficha técnica que incluye datos contextuales (edad, sexo, lugar de la grabación, situación, temática y rasgos sociolingüísticos de los participantes). Es precisamente el sistema de transcripción y codificación propio el que más impacto ha tenido en los estudios posteriores tanto en España como en América Latina (Bravo y Briz 2004). También en el ámbito de la enseñanza de E/LE (Briz 2002, 2008), donde se describen los rasgos de la variedad coloquial y se hacen propuestas para trabajarla en la clase de E/LE.

³² <<http://www.valesco.es>>

³³ <http://fonocortesía.es/?page_id=47>

También con atención a una situación comunicativa concreta se ha desarrollado el corpus *Clinical Interview (CLINT)*³⁴ (González et al. 2010), que es el primer corpus oral bilingüe español-catalán de entrevistas médico-paciente, coordinado por M. Antònia Martí de la Universitat de Barcelona y desarrollado por el *Centre de Llenguatge i Computació (CLiC)*. Recoge 15 horas de conversación que corresponden a 40 entrevistas clínicas y está formado por archivos sonoros alineados con transcripciones a varios niveles que recopilan información ortográfica, fonética y morfológica, así como codificación lingüística y extralingüística.

A continuación, se presentan dos corpus, COLA y COVJA, que difieren de los anteriores en que recogen la producción de hablantes jóvenes (adolescentes y universitarios) y los registros y estilos propios de esos grupos.

El *Corpus Oral de Lenguaje Adolescente (COLA)* (Jørgensen 2008), dirigido por Annette Myre Jørgensen de la Universidad de Bergen (Noruega), incluye conversaciones espontáneas entre jóvenes de 13 a 19 años de Madrid y de otras capitales latinoamericanas. Dicho corpus se recogió entre 2002-2009, aunque el proyecto no está en sí terminado³⁵. Las conversaciones ofrecen los datos contextuales, están transcritas y alineadas con los ficheros sonoros mediante el programa *Transcriber*³⁶ (Barras et al. 2001). Esta herramienta permite la alineación entre el audio y la transcripción ortográfica, es decir, el material transcrito está sincronizado con el archivo de sonido. Y, además, se ha incluido un motor de búsqueda para navegar por los documentos.

El *Corpus Oral de la Variedad Juvenil Universitaria del Español de Alicante (COVJA)* (Azorín y Jiménez Ruiz 1997) forma parte de un corpus más amplio, el *Alicante Corpus Oral del Español (ALCORE)* (Azorín 2002), ambos coordinados por Dolores Azorín. En estos corpus se recoge el habla de la zona de Alicante a partir de entrevistas y conversaciones en grupo. El corpus COVJA está formado por la producción de 63 jóvenes estudiantes de nivel universitario, con edades comprendidas entre los 18 y los 24 años y con una mayoría de hablantes masculinos, seleccionados de entre distintas

³⁴ <<http://clic.ub.edu/corpus/es/clint>>

³⁵ Véase <http://www.colam.org/om_prosj-espagnol.html>

³⁶ <<http://trans.sourceforge.net/en/presentation.php>>

especialidades de estudios. La recogida de datos se basa en un mismo guion conversacional del cual surgen temas familiares y relaciones informales entre los participantes. ALCORE, a su vez, está formado por la producción de 76 informantes clasificados en tres niveles socioculturales, tres niveles de edad y de ambos sexos. Ambos corpus están transcritos ortográficamente y siguen las convenciones TEI .

Desde una orientación algo distinta, cabe destacar también los corpus orales multilingües que incluyen la producción de tareas similares producidas por hablantes nativos en varias lenguas: C-ORAL-ROM y SACODEYL.

El *Corpus Oral de Lenguas Románicas*, **C-ORAL-ROM** (Cresti y Moneglia 2005), incluye la producción en cuatro lenguas románicas: italiano, francés, portugués y español. Coordinado por Emanuela Cresti y Massimo Moneglia, de la Universidad de Florencia, el subcorpus de español fue recopilado por el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (equipo LLI-UAM) y coordinado por Antonio Moreno Sandoval (2000).

Las muestras de español de C-ORAL-ROM tienen una duración de 30 horas y cuentan con 300.000 palabras que corresponden a 183 documentos de una amplia variedad de géneros discursivos, además de conversaciones coloquiales y familiares. Por lo tanto, contemplan registro formal e informal, conversaciones cara a cara y telefónicas, monólogos, discursos políticos o legales, entre otros. Los textos están transcritos ortográficamente en formato *CHAT* (MacWhinney 2006) y anotados morfosintácticamente. Se han alineado los textos con los ficheros de audio mediante *WinPitch*³⁷. Los documentos también ofrecen información sobre las características contextuales (sexo, edad, profesión, educación, origen, etc.) e información sobre la grabación (fecha, lugar, situación y condiciones de grabación, tema, duración, etc.).

El corpus *System Aided Compilation and Open Distribution of European Youth Language* (**SACODEYL**)³⁸ se desarrolló en el marco de la Unión Europea dentro del proyecto *Minerva* (2005-2008), y recoge entrevistas realizadas a adolescentes en contextos de instrucción escolar en inglés, francés, alemán, italiano, lituano, rumano y español. El coordinador del subcorpus español es Pérez-Paredes, de la Universidad de

³⁷ <<http://www.winpitch.com>>

³⁸ <<http://sacodeyl.inf.um.es/sacodeyl-search2/>>

Murcia (Pérez-Paredes 2010). Las edades de los informantes están comprendidas entre los 13 y los 18 años. SACODEYL está formado por diferentes subcorpus que corresponden a los diferentes idiomas. A su vez, cada uno de ellos contiene de 20 a 25 grabaciones de entrevistas en vídeo de 10 minutos cada una. El hecho de disponer del vídeo constituye una novedad. Para recoger las entrevistas de vídeo de cada subcorpus se emplean las mismas preguntas, que cubren una serie de temas (información personal, familia y casa, vacaciones, escuela, etc.). Las transcripciones están disponibles en la web, agrupadas por áreas temáticas y anotadas siguiendo consideraciones pedagógicas (etiquetadas por temas, características gramaticales y léxicas, conectores del discurso y nivel de competencia lingüística según el *MCER*). Se pueden recuperar las grabaciones de vídeo con la transcripción a partir de las áreas temáticas.

SACODEYL es un corpus multilingüe diseñado para aplicación pedagógica en entornos virtuales, con alta interactividad e integrado en el sistema de las aulas TIC (con explotación didáctica en la web para el profesor, adicionalmente). Es un ejemplo de corpus diseñado para usarse en el aprendizaje de lenguas (*data driven learning*) en un macro-contexto de enseñanza y aprendizaje más innovador³⁹.

Relacionado con el proyecto SACODEYL existe otro corpus multilingüe, pero con informantes adultos de diferentes edades y no discentes. Es el corpus **BACKBONE**⁴⁰, que tiene como finalidad servir también como base de datos pedagógicos de otras variedades diafásicas, siempre en formato vídeo. Es un proyecto coordinado por Kurt Kohn, de la Universidad de Tübingen (Kohn et al. 2010). El corpus se estructura en siete subcorpus diferentes: seis constituidos por la producción oral nativa de hablantes de idiomas distintos (inglés, francés, alemán, polaco, español y turco), y uno constituido por la producción oral de aprendices de inglés como lengua extranjera, en el que los informantes son nativos de distintas lenguas maternas.

El subcorpus de lengua española está formado por entrevistas de temática variada; el habla refleja la de diferentes regiones de España y los informantes

³⁹ Veáse también la página web del departamento de filología inglesa de la Universidad de Murcia que explica el proyecto SACODEYL:

<<http://www.um.es/web/filologia-inglesa/contenido/investigacion/proyectos/sacodeyl>>.

⁴⁰ <<http://projects.ael.uni-tuebingen.de/backbone/moodle/>>

pertenecen a distintos ámbitos profesionales. Estas entrevistas están transcritas y contienen la anotación de elementos gramaticales, léxicos, funcionales, temáticos y, además, van acompañadas de materiales didácticos para su explotación pedagógica, como SACODEYL. Incluye, además, un programa tutor (*tutorial*) para enseñar a los docentes a usar las herramientas de los investigadores y poder así crear sus propios corpus y material didáctico.

Otros corpus vinculados al aprendizaje de lenguas son COGILA, el *Corpus del Español en Texas* y el *Corpus de conversaciones en español* de Columbia University. Los dos primeros se construyeron expresamente como recurso para el aula puesto que defienden que los materiales orales existentes en situaciones naturales se diseñaron con una finalidad investigadora y no educativa.

El *Corpus del Grupo de Investigación de Lingüística Aplicada (COGILA)* (2012), dirigido por el profesor Pedro Barros García de la Universidad de Granada (Barros; Morales y Barros 2015), incluye una selección de 10 transcripciones de discursos orales conversacionales en español por jóvenes andaluces. Esta recopilación, en formato libro (Barros 2012), parte de un proyecto anterior (de 1999 a 2008), cuyo objetivo era elaborar un corpus de conversaciones reales de español peninsular meridional. La muestra de 10 textos recoge la forma de conversar de jóvenes españoles de entre 17 y 35 años, la mayoría estudiantes universitarios, entre los años 2010-2011. Los diálogos, cuya duración oscila entre los 3 y los 21 minutos, suman un total de 134 minutos de grabación, se ofrecen transcritos y siguen el sistema propuesto por Val.Es.Co. Además, se incluye la versión oral en audio y vídeo, en formato CD_ROM. Cada grabación aporta información sobre los participantes y los datos técnicos de la grabación.

COGILA se construyó para uso tanto de hablantes nativos que desean conocer las reglas orales que subyacen en la interacción conversacional, como para uso de los no nativos, ofreciendo diálogos reales junto a referencias morfosintácticas, especializaciones léxico-semánticas (por ejemplo, modismos, colocaciones) y realizaciones fonéticas. Es decir, se trata de un corpus que tiene como finalidad poner a disposición de los estudiantes de español textos discursivos orales en los que previamente se han señalado los elementos del discurso que se van a trabajar. Este tipo de enfoque para los textos orales no es habitual en los manuales de E/LE.

El **Corpus del Español en Texas** está coordinado por Barbara E. Bullock y J. Almeida Toribio de la Universidad de Texas en Austin (UT Austin). Este proyecto se incluye dentro del programa del *Center for Open Educational Resources and Language Learning* (COERLL) entre los años 2010-2014 (Gilg, Larsen Serigos y Trovato 2014).

Este corpus tiene la finalidad de describir el español que se habla actualmente en Texas y de ofrecer herramientas de aprendizaje libres que ayuden a estudiantes, educadores e investigadores a analizar la variación lingüística del español, especialmente la variación local, a la hora de aprender dicho idioma y su cultura. El español que se habla en Texas, recordemos, se caracteriza porque la comunidad hispanohablante texana es bilingüe en mayor o menor medida y habla también inglés.

El proyecto del Corpus del Español en Texas desarrolla un corpus de entrevistas semidirigidas a partir de un guion de preguntas cuyos informantes tienen diversos perfiles: son estudiantes universitarios de la UT Austin, sus familiares o amigos y conocidos; tienen edades muy distintas; pertenecen a ambos géneros y proceden de diferentes regiones de Texas. El corpus incluye aproximadamente 92 vídeos en español o bien en español e inglés, con una duración entre 30-45 minutos. Los archivos se pueden descargar en formato vídeo y proporcionan las transcripciones ortográficas completas (contiene más de 500.000 palabras). Los textos se han alineado con las grabaciones y se han anotado morfológicamente.

Sobre la colección de los 92 vídeos de entrevistas del *Corpus del Español en Texas* se hizo una selección de videoclips, que se recogen en el *SpinTX: Corpus-to-Classroom Project* (2012-2013) como fuente de material pedagógico. Esto es, se parte del corpus más extenso, el *Corpus del Español de Texas*, y se seleccionan fragmentos de las entrevistas de los vídeos ya corregidos y anotados. Por lo tanto, está formado por 327 videoclips con una duración de 1 a 4 minutos de 33 hablantes distintos. Este corpus tiene las mismas características que el *Corpus del Español en Texas*: se ha transcrito (incluye 80.000 palabras), los textos se han alineado con los archivos sonoros, se ha realizado la anotación gramatical, si bien a éste se le ha añadido una anotación pedagógica. El recurso educativo SpinTX⁴¹ está organizado por temas y

⁴¹ <<http://www.coerll.utexas.edu/spintx/>>

contenidos gramaticales y es de libre acceso. Además, ofrece ideas para trabajar los diferentes vídeos: se detalla el nivel de lengua, las destrezas que se trabajan, la dinámica de clase y el tiempo necesario.

El *Corpus de conversaciones en español*⁴², coordinado por la profesora Guadalupe Ruiz-Fajardo de la *Columbia University*, está en construcción. El proyecto tiene la intención de recopilar conversaciones informales en formato vídeo entre hablantes nativos de español en situaciones de intercambio de información personal (desayuno, exámenes, viajes, tiempo, entre otros). Los diálogos se están recogiendo en diferentes lugares de España y Latinoamérica y se están considerando diferentes variables (la ubicación rural frente a la urbana; la edad; el sexo; el origen, la formación educativa de los participantes y la relación de interacción) para analizar las características dialectales y el tipo de discurso. Otros de los objetivos de este corpus son transcribir los diálogos y hacer la descripción técnica del vídeo (información sobre los participantes, lugar de grabación y sobre los datos técnicos relativos a la grabación).

Con unas características diferentes a los anteriores, está el *Corpus Oral Didáctico Anotado Lingüísticamente (C-Or-DiAL)*⁴³, coordinado por la profesora Carlota Nicolás Martínez de la Universidad de Florencia, y diseñado tanto para la enseñanza como para la investigación (Nicolás 2012b). En efecto, Nicolás propone no solo su uso didáctico sino llevar a cabo pequeñas investigaciones de corpus en el aula de E/LE, puesto que da las herramientas para crear un corpus similar con los alumnos. Además Nicolás, como metodología de aprendizaje en la propia aula de E/LE, enseña a los estudiantes la lengua oral a través de la transcripción de lo que escuchan y les hace reflexionar sobre cómo funciona el español hablado.

C-Or-DiAL se confeccionó entre 2001-2004 en el *Laboratorio Linguistico del Dipartimento di Italianistica dell'università di Firenze* y recogió 240 grabaciones espontáneas registradas en Madrid con hablantes madrileños —los informantes dieron permiso posteriormente—. Se transcribió ortográficamente y se anotó prosódicamente, según las etiquetas usadas en C-ORAL-ROM. Además, se etiquetó con

⁴² <<http://corpusdeconveraciones.com>>

⁴³ <<http://lablita.dit.unifi.it/corpora/cordial/>>

las funciones comunicativas recopiladas en cada texto y tomadas del *Plan Curricular del Instituto Cervantes*. Este corpus contiene 120.000 palabras y tiene una duración de diez horas de grabación, e incluye los siguientes géneros discursivos: por un lado, conversaciones y diálogos espontáneos (29%); por otro lado, textos no espontáneos y públicos: entrevista informal (51%), conversaciones con tema preestablecido (13%) y clases o conferencias (7%).

Los textos de C-Or-DiAL incluyen datos contextuales en cada transcripción: información de los hablantes y relación entre ellos, situación de la grabación, tema general del texto y el tema específico del fragmento transcrito. Incluye también información de carácter más concreto: nivel de dificultad del texto para el uso en el aula de E/LE, observaciones lingüísticas para dicha aplicación, funciones comunicativas anotadas en cada transcripción y palabras clave.

A continuación, la tabla 1 ofrece un resumen en orden alfabético de los corpus orales discursivos en español descritos en esta subsección, con excepción del corpus de la Universidad de Columbia, que está en construcción.

CORPUS	TIPO DE DISCURSO	CARACTERÍSTICAS INTERLOCUROS	SISTEMA DE REPRESENTACIÓN	FECHA DE GRABACIÓN	ÁREA GEOGRÁFICA	DISPONIBILIDAD	DIRECTOR DEL PROYECTO
ACUAH	Conversación semidirigida	Nivel sociocultural medio, tres rangos de edad y ambos sexos	Transliteración ortográfica y etiquetas TEI	1993	Alcalá de Henares	Cedido a CREA	Ana M ^a Cestero Universidad Alcalá de Henares
ALCORE Y COVJA	Entrevista-coloquio con guion conversacional. ALCORE: 120 h. COVJA: 800 min.	<i>ALCORE</i> : 76 informantes de tres grupos de edad, ambos sexos y tres niveles socioculturales, algunos bilingües valencianos <i>COVJA</i> : 63 informantes jóvenes universitarios	Transliteración ortográfica y etiquetas TEI	1996	Alicante (España), algunos inmigrantes de otras comunidades	<i>ALCORE</i> : en CD-ROM (2002) <i>COVJA</i> : incorporado al CREA	Dolores Azorín y Juan Luis Jiménez Ruiz, Universidad de Alicante
BACKBONE	Entrevistas de temática variada	Diferentes edades	Transliteración ortográfica, anotación pedagógica y enriquecido con actividades pedagógicas	2009-2010	Distintos países europeos. Parte de español: diferentes regiones de España	En la red	Pascual Pérez-Paredes, Universidad de Murcia, ha coordinado la parte de español
CdE oral	2000 entrevistas de temáticas distintas	—	Codificado por lema, partes de texto, sinónimos y otras informaciones semánticas	Creado en 2002, recoge textos del S.XIII-XX	Latinoamérica y España	Acceso electrónico gratuito	Mark Davies Universidad Brigham Young
CLINT	Entrevistas clínicas	Identificación del hablante, sexo y lenguas en que son competentes y lengua que usan en la entrevista	Transliteración ortográfica, siguiendo EAGLES; sincronización transcripción-sonido; anotación morfológica; transcripción fonética	2010	Barcelona (España)	No está disponible	M. Antònia Martí Universitat de Barcelona
COLA	Conversaciones espontáneas. Actualmente 720.000 palabras transcritas	Jóvenes de ambos sexos, de 13 a 18 años	Transliteración ortográfica y etiquetas TEI; sincronización transcripción-audio	2002-2009	Santiago de Chile, Madrid, Buenos Aires, Guatemala	Acceso electrónico gratuito (solicitud de contraseña)	Annette Myre Jørgensen, Universitat de Bergen

COGILE	Conversaciones espontáneas informales. 134 min., 36.000 palabras	Jóvenes, la mayoría estudiantes universitarios, de 17 a 35 años	Transliteración ortográfica con sistema de transcripción Val.Es.Co.	2010-2011	Variedad meridional del español peninsular	En libro con CD-ROM	Pedro Barros García, Universidad de Granada
C-ORAL-ROM (español)	Conversaciones cara a cara y telefónicas, diálogos públicos y familiares, diversos tipos de discursos monológicos y políticos. 30 h., 300.000 palabras	Diferentes edades, procedencias geográficas, profesiones. Se señalan los estudios	Transliteración ortográfica con anotación morfo-sintáctica; sincronización transcripción-audio.	2001-2004	Diversas zonas de Latinoamérica y España. La mayoría, Madrid y Segovia	Es de pago	Antonio Moreno Sandoval ha coordinado la parte del español
C-Or-DiAL	Dos tipologías de texto: -diálogos y conversaciones espontáneas; -entrevistas, clases, conferencias, etc. no espontáneas y públicas 120.000 palabras	Información sobre los hablantes y su relación	Transliteración ortográfica; texto y audio no alineado (en la página web).	2001-2004	Variedades de Madrid	Acceso electrónico gratuito y en texto con el CD	Carlota Nicolás Martínez <i>Università di Firenze</i>
CORLEC, Corpus Oral de Referencia de la Lengua Española Contemporánea	Conversaciones cara a cara, teléfono, diálogos públicos y familiares, discursos monológicos y políticos	Diferentes edades, niveles socioculturales, ambos sexos	Transliteración ortográfica y etiquetas TEI	1990-1992	España	En la red	Francisco A. Marcos Marín, Universidad Autónoma de Madrid
Corpus del habla de Almería	Entrevistas semidirigidas, conversaciones cara a cara y monólogos	Tres niveles socioculturales, tres rangos de edad y ambos sexos	Transliteración ortográfica y etiquetas TEI. 55 h.	2009	España, Almería	En la red (grabación sonora sin transcripción)	Luís Cortés Rodríguez Universidad de Almería

COSEER	Entrevistas semidirigidas con rasgos cercanos a la conversación	Edad (la media de edad de los informantes es de 72,5 años), sexo, nivel sociocultural (escasa escolarización) y zona rural	Transliteración ortográfica con algunas convenciones propias	1990-actualidad	España	En la red, aunque no están todos los textos	Inés Fernández-Ordóñez, Universidad Autónoma de Madrid
CREA oral	Conversaciones, entrevistas, documentales, reportajes, debates, concursos, retransmisiones deportivas, etc. 9 millones de palabras, 1600 documentos	Gran diversidad de hablantes, de distintas edades y niveles socioculturales; ambos sexos, pero no clasificados por estratificación sociolingüística	Transliteración ortográfica y etiquetas TEI; sincronización transcripción-audio	1979-2004 (últimos 25 años de la actualidad)	50% diversos dialectos peninsulares, 50% diversos dialectos latinoamericanos	En la red	Guillermo Rojo Universidad de Santiago de Compostela y miembro de la Real Academia Española
MC-NC	14 entrevistas libres con intervención del encuestador	Nivel culto, clasificados por estratificación sociolingüística	Transliteración ortográfica	1971-1992	9 ciudades latinoamericanas y 3 españolas	En CD-ROM Sí, los textos (no los audios)	José Antonio Samper Padilla Universidad de Las Palmas de Gran Canaria
PRESEEA	Entrevistas semidirigidas, normalmente con encuesta	Tres niveles socioculturales, tres rangos de edad y ambos sexos	Transliteración ortográfica, etiquetas TEI, algunos signos tipográficos y/o fonéticos	1996-actualidad	Diversas ciudades latinoamericanas y españolas	Algunas en la red o contactar con los equipos investigadores	Francisco Moreno Fernández, Universidad de Alcalá de Henares
SACODEYL (español)	Monólogos (información personal, vacaciones, escuela...) a partir de entrevistas	Jóvenes de ambos sexos entre 13 y 18 años	Transliteración ortográfica con anotación pedagógica; sincronización transcripción-vídeo. 4 horas de grabación aprox.	2005-2008	España	Acceso electrónico gratuito	Pascual Pérez-Paredes, Universidad de Murcia, ha coordinado la parte de español

Corpus del Español en Texas y SpinTX Video Archive Corpus	Entrevistas semidirigidas. Corpus Español Texas: 92 vídeos, SpinTX: 327 vídeo clips. Corpus Español Texas: más de 500.000 palabras, SpinTX: 80.000 palabras	Diferentes perfiles personales, edades, géneros y regiones de Texas.	Transliteración ortográfica con anotación pedagógica; sincronización transcripción-vídeo.	2010-2014	Regiones en Texas (Estados Unidos)	Acceso electrónico gratuito (solicitud de contraseña)	Barbara E. Bullock y Almeida Jacqueline Toribio, The University of Texas at Austin (UT Austin)
Val.Es.Co.	Conversación coloquial	Diferentes edades, niveles socioculturales, ambos sexos, algunos bilingües valencianos	Transliteración ortográfica y transcripción fonética. 341 h.	1989-2000	Comunidad Valenciana (España)	Acceso electrónico gratuito (solicitud de contraseña)	Antonio Briz Gómez, Universitat de València

TABLA 1. *Corpus orales discursivos de hablantes nativos de español*

Recapitulando, si tenemos en cuenta algunos de los corpus presentados en esta sección que reúnen muestras de español oral, vemos que, en principio, estos se construyeron principalmente para estudiar cuestiones sociopragmáticas de la lengua: unos se centran en el análisis de la conversación —los corpus ACUAH, PRESEEA y Val.Es.Co.—; otros, en los estudios de variación geográfica —el *Corpus del habla de Almería*, el COSER—. No obstante, es cierto que por sus características, se han utilizado con mayor o menor frecuencia en la didáctica del español como lengua extranjera. Algunos de estos grupos de investigación, interesados en los aspectos discursivos orales del español, además, han diseñado propuestas didácticas basadas en corpus, destinadas al aprendizaje del español como lengua extranjera: C-ORAL-ROM y el corpus de Val.Es.Co., fundamentalmente.

Campillos et al. (2007) adaptan la parte de español del corpus C-ORAL-ROM⁴⁴ para trabajar la destreza de la comprensión auditiva específicamente, así como también contenidos gramaticales concretos, de acuerdo con los niveles de dificultad según el nivel de competencia del *MCER*. El procedimiento consiste en diseñar los contenidos didácticos a partir del análisis de muestras de lengua oral y alejarse del *Data-driven Learning* —o aprendizaje dirigido a los datos— que utiliza los corpus en el proceso de enseñanza de las lenguas extranjeras. Resultado de esta investigación es el manual *Español oral en contexto* (Campillos et al. 2010), libro acompañado de CD, para profesores y alumnos de niveles B1 a C2⁴⁵.

Otra de las propuestas didácticas en E/LE la presenta Nicolás (2002) a partir de la explotación del corpus C-ORAL-ROM mediante una serie de actividades para trabajar el estudio de la lengua oral en contraste con la lengua escrita. El corpus C-ORAL-ROM, como se apuntó, se acompaña de la herramienta de análisis acústico *WinPitch*, que alinea el sonido con la transcripción escrita y, además, representa gráficamente la entonación; de este modo se estudian y se practican los elementos propios del español oral.

⁴⁴ En Campillos et al. 2007: 2-3, “la aplicabilidad de C-ORAL-ROM fue evaluada por el Departamento de Tecnología y Proyectos lingüísticos del Instituto Cervantes (...) destacando su aplicabilidad en el campo de ELE (enseñanza asistida por ordenador, aulas multimedia, autoaprendizaje, aplicación del enfoque comunicativo, práctica de aspectos sociolingüísticos, análisis de discurso oral, pragmática, pronunciación, vocabulario, sintaxis)”.

⁴⁵ Para leer la reseña hecha para este manual, consultar Muñoz, C. (2014).

Por otro lado, los investigadores de Val.Es.Co. y sus discípulos ofrecen propuestas didácticas para E/LE a partir de su corpus (Albelda y Fernández 2006, 2008); Albelda (2010) se centra, fundamentalmente, en los mecanismos de conectividad y atenuación; en los aspectos léxicos (registro, colocaciones, frecuencias) etc. Autores ajenos al proyecto también usan el modelo de transcripción así como los datos de Val.Es.Co. para explotaciones didácticas similares en el aula: Cordulino da Silva (2010) — para estudiantes brasileños de nivel B2— y Nica (2012) —para A1 y A2, para grupos multilingües. Secchi (2014), por su parte, propone actividades didácticas a partir del análisis y los criterios de los corpus Val.Es.Co. y C-ORAL-ROM.

2.2. Modelos no nativos: Metodología de corpus de aprendices (*learner corpus*)

En esta sección se presentan los corpus no nativos producidos por aprendices de español, sean orales o escritos, pero ceñidos al contexto de instrucción formal, sus criterios de recogida y su explotación. Primero se acota qué es un corpus de aprendices; a continuación se presentan algunos antecedentes y metodologías afines a nuestro estudio y, finalmente, se abordan específicamente los corpus de aprendices del español, y en concreto, los orales, en su tipología y metodología.

Se conocen como corpus de aprendices aquellos que recogen producción oral o escrita de aprendices de una L1 o de una L2, L3 (Meurers 2015). Los corpus de aprendices compilan textos para el estudio de la lengua en dos áreas diferentes: a) la adquisición de una L1 por parte de los aprendices nativos, ya sean niños o adolescentes; y b) el aprendizaje de una L2 —o una nueva lengua, L3, L4— diferente de la lengua materna. En el primer caso, los corpus sirven para estudiar el desarrollo de la adquisición infantil del lenguaje; mientras que en el segundo caso, los corpus permiten describir el desarrollo de la interlengua de los aprendices de una lengua extranjera.

En este subapartado se tratarán los corpus de aprendices de una L2, mientras que la sección 2.3. se centrará en los corpus de adquisición de una L1. La pertinencia de la presentación de ambos tipos de corpus y sus metodologías relacionadas en este capítulo responde a que la investigación suele compartir las herramientas

(sistema de transcripción, analizadores, criterios, etc.) y que, en ocasiones, los propios bancos de datos son comunes a estas situaciones de adquisición, como ocurre con *TalkBank* (ver sección 2.3).

Metodológicamente, Granger (2002, 2003) propone, además, que:

“Learner corpora provide a new type of data which can inform thinking both in SLA (Second Language Acquisition) research, which tries to understand the mechanisms of foreign/second language acquisition, and in FLT (Foreign Language Teaching) research, the aim of which is to improve the learning and teaching of foreign/second languages”.

Según Granger (2008) y Rosen et al. (2013), los corpus de aprendices (L2) se caracterizan por los siguientes rasgos:

- pueden recopilar textos escritos u orales, si bien estos últimos son más difíciles de recoger y, por ello, menos comunes;
- los informantes pueden tener la misma lengua materna o diferentes L1;
- la mayoría de corpus de aprendices de L2 o LE disponibles son para el inglés. Los corpus de aprendices en otras lenguas son menos; sin embargo, está empezando a aumentar su número, como ilustra el caso del español;
- algunos corpus recogen textos de estudiantes del mismo nivel de competencia en la L2; otros incluyen textos de estudiantes de diferentes niveles y la mayoría se centra en estudiantes de nivel avanzado (B2 en adelante);
- la gran mayoría de los corpus son transversales y agrupan textos de aprendices de distintas lenguas maternas; por el contrario, son pocos los corpus de L2 longitudinales. Muchos corpus de aprendices se recogen con unos criterios específicos (grupos homogéneos, diferentes niveles de lengua extranjera...) y se usan para la investigación en adquisición de segundas lenguas como corpus de aprendices casi-longitudinales; y
- algunos corpus de aprendices ofrecen los textos sin anotar con los datos en bruto, mientras otros recogen textos anotados con información lingüística (la más común, el etiquetado de las categorías gramaticales y, también, el etiquetado de la sintaxis, la

semántica y las funciones comunicativas); o bien solamente se anotan los errores con su información contextual correspondiente.

Si se revisan los estudios realizados en el ámbito del aprendizaje de lenguas, los primeros corpus de aprendices, que se ajustan ya a estos criterios mencionados, empiezan a aparecer entre la década de los 80 y los 90 dentro de la lingüística aplicada inglesa. No obstante, ya en los años 60 y 70 se empezaron a analizar manualmente los errores de aprendices y se almacenaron en una base de datos como fuente de tipos de errores (Nesselhauf 2004), si bien no era propiamente un corpus de aprendices.

El inicio de los corpus de aprendices en sentido estricto lo marca el proyecto de la Universidad de Lovaina: el *International Corpus of Learner English (ICLE)* (Granger et al. 2002, 2003) iniciado en los años 80. Este corpus consiste en textos argumentativos escritos por estudiantes, la mayor parte universitarios aprendices de inglés como L2 con diferentes lenguas maternas. La última versión del corpus, **ICLEv2** (Granger et al. 2009), consta de 3,7 millones de palabras producidas por estudiantes de hasta 16 lenguas maternas distintas. A continuación, se crea el *Louvain Corpus of Native English Essays (LOCNESS)*⁴⁶ que recoge un corpus equivalente de inglés pero integrado por hablantes nativos (británicos y norteamericanos, de aproximadamente 325.000 palabras), que les permite comparar y describir la interlengua de los aprendices respecto a la lengua nativa. Este mismo equipo desarrolla el corpus *Louvain International Database of Spoken English Interlanguage (LINDSEI)*, que recoge entrevistas informales de aprendices de inglés de nivel intermedio alto a nivel avanzado. El corpus está producido por aprendices de distintas lenguas maternas, entre ellas el español, y contiene 800.000 palabras que corresponden a la producción de los aprendices.

Otro estudio pionero fue el corpus de aprendices de inglés de distintas lenguas maternas de Longman, *Longman Learners' Corpus (LLC)*⁴⁷ (Gillard and Gadsby 1998), formado por 10 millones de palabras procedentes de ensayos escritos de todos los niveles de aprendizaje.

⁴⁶ <<http://www.learnercorpusassociation.org/resources/corpora/locness-corpus/>>

⁴⁷ <<http://www.pearsonlongman.com/dictionaries/corpus/learners.html>>

El objetivo de estos corpus es la descripción de las características de la IL de los aprendices, a partir de los rasgos léxicos, morfosintácticos, discursivos, etc., tomando como base los textos escritos de aprendices de inglés de diferentes lenguas maternas.

Después de ICLE, se han creado más corpus de aprendices bien de L2 inglés con L1 español, *Written Corpus of Learner English (WriCLE)*⁴⁸ de la Universidad Autónoma de Madrid y *Santiago University Learners of English Corpus (SULEC)*⁴⁹; bien L2 español con L1 inglés (acompañado de un corpus nativo similar), el *Corpus Escrito del Español L2 (CEDEL2)*⁵⁰ de la Universidad Autónoma de Madrid. CEDEL2 y WriCLE se desarrollan bajo el proyecto europeo WOSLAC (*Word Order in Second Language Acquisition Corpora* para estudiar el papel de interfaz que tienen el léxico y la sintaxis en el desarrollo del discurso en la interlengua⁵¹ española (Lozano 2009; Mendikoetxea et al. 2009; Lozano y Mendikoetxea 2013).

Hasta aquí se han citado algunos de los corpus más representativos, sobre todo corpus escritos de aprendices, tanto para el aprendizaje del inglés como del español. Como se desprende de la revisión bibliográfica (cf. la web *Learner Corpora around the world*⁵² (2015) de la Universidad de Lovaina), la gran mayoría de corpus de aprendices en español son recopilaciones de textos escritos. Además del ya mencionado corpus CEDEL2, destacan también el corpus *Aprender a Escribir en Lovaina (Aprescrilov)* (Buyse y González 2012) que recoge diferentes tipos de textos escritos (cartas, descripciones, argumentaciones, etc.) de L1 holandés; *The Corpus of Taiwanese Learners of Spanish* (Corpus de Aprendices Taiwaneses de Español, **CATE**) (Hui-Chuan Lu 2010), formado por 1058 redacciones (181,322 palabras) de aprendices taiwaneses de diferentes niveles de E/LE y universidades de Taiwan; *The Japanese learner corpus of Spanish* (Yoshihito Kamakura, 2005) con 83.400 palabras escritas por aprendices japoneses; *The Anglia Polytechnic University (APU) Learner Spanish Corpus* (Ife 2004) que reúne textos escritos (120.000 palabras) en los niveles iniciales de aprendizaje de E/LE, producidos por aprendices de diferentes lenguas maternas. Otros corpus que agrupan textos escritos, procedentes de aprendices que tienen diferentes lenguas maternas, son: el **Corpus para**

⁴⁸ < <http://web.uam.es/proyectosinv/woslac/WriCLE/>>

⁴⁹ < <http://www.sulec.es>>

⁵⁰ < <http://www.uam.es/proyectosinv/woslac/cedel2.htm>>

⁵¹ <<http://www.uam.es/proyectosinv/woslac/>>

⁵² <<http://www.uclouvain.be/en-cecl-lcworld.html>>

el análisis de errores de aprendices de E/LE (CORANE) (Cestero et al. 2002), por un lado, y el *Corpus de aprendices de español (CAES)* (Instituto Cervantes y Universidad de Santiago de Compostela 2014)⁵³ que comprende 573,718 palabras (versión 1.0, octubre de 2014), por otro.

No hay que olvidar tampoco los corpus multilingües en los que hay una sección dedicada al español y en los que los aprendices tienen la misma L1: *The MiLC Corpus*, cuya L1 es catalán (A. Andreu Andrés et al. 2010, de la Universidad Politécnica de Valencia) y *The Multilingual Learner Corpus (MLC)* (Tagnin 2002) de estudiantes universitarios que tienen como L1 el portugués de Brasil.

Otros corpus de aprendices procedentes de grupos homogéneos que comparten la L1, siempre de textos escritos, son: el *Corpus de textos escritos por universitarios italianos estudiantes de ELE* (Gutiérrez Quintana 2005); *The Spanish Corpus of Italian Learners (SCIL)* (Bailini 2013) y el corpus **SAELE** (Pino 2009), formado por aprendices suecos de E/LE.

En cambio, como veremos, los corpus orales de aprendices para el español son muchos menos. A continuación, se hace una descripción detallada de los disponibles.

2.2.1. Corpus orales de aprendices de español como LE y su metodología

Desde hace relativamente poco tiempo se cuenta con datos orales espontáneos o semiespontáneos producidos por aprendices de E/LE o E/L2. Los corpus de aprendices orales de español LE encontrados son 10 hasta donde hemos podido saber. La descripción y la metodología de estos se organiza en función de las características de los aprendices de español de los diferentes corpus.

Primero se presentan los corpus y metodologías constituidos por aprendices que comparten la misma L1 (por ejemplo, el corpus SPLLOC). A continuación, los corpus y metodologías constituidos por aprendices de diferentes L1s (por ejemplo, el corpus LUIS-ELE/L2) y por último, los corpus bilingües constituidos por sujetos que interaccionan en dos lenguas, siendo una la L1 de una parte del grupo y la L2 del otro y viceversa (por ejemplo, el caso de CORINÉI, de italiano y español).

⁵³ <<http://galvan.usc.es/caes>>

Uno de los primeros corpus orales de aprendices de español es el *Corpus oral del proceso de adquisición del español como L2 por una joven inmigrante búlgara* (Hernández et al. 2008a). Se trata de un estudio de caso longitudinal registrado a lo largo de un año, en 1997⁵⁴, para establecer el desarrollo de las categorías funcionales del español L2. Dicho corpus está transcrito ortográficamente (si bien los archivos de voz no están disponibles), no está anotado y aporta información contextual sobre la aprendiz: nacionalidad, fecha de las grabaciones y breve descripción de los temas de diálogo a través de entrevistas semidirigidas. Este es un ejemplo de corpus homogéneo formado, en este caso, por un único informante de una única L1.

También oral y homogéneo, de aprendices que comparten una L1, es el corpus *Spanish Learner Language Oral Corpora (SPLLOC) (2006-2010)*⁵⁵ (Mitchell et al. 2008), coordinado por Laura Domínguez de la Universidad de Southampton, aunque en el proyecto colaboran otras universidades. Dicho proyecto consiste en recopilar una base de datos oral para investigar la adquisición de E/LE por aprendices cuya L1 es el inglés.

SPLLOC está formado por dos subcorpus: uno en español L2 y otro en español L1 (de control), cuyos objetivos son la descripción de la interlengua en comparación con la L1⁵⁶. El estudio tuvo dos fases: la primera se centró en la descripción de la adquisición de los aspectos morfosintácticos en general; en la segunda se estudió el desarrollo de los tiempos verbales y el aspecto, específicamente. El corpus está formado por la producción de las grabaciones de 60 alumnos ingleses de español L2 de tres niveles de competencia distintos (veinte alumnos de A2, veinte de B1-B2 y veinte de C1-C2, niveles asimilados al *MCER* actual). Consiste en narraciones, entrevistas y descripción de fotos, en español L1 y L2 (como tareas de elicitación semidirigidas) con el fin de permitir comparar el español nativo y la interlengua española de los anglófonos, desde nivel básico (A2) hasta nivel avanzado (C2), en contextos institucionales. Como se ha señalado, para tener un término de comparación, un grupo de hablantes nativos de edades similares a los informantes realizó las mismas tareas.

⁵⁴ Incluido en el proyecto de investigación "Estudio Comparativo entre la adquisición del Español como Primera Lengua y la Adquisición del Español como Segunda Lengua para su aplicación metodológica en la enseñanza del español a Inmigrantes" (1999) (Hernández et al. 2008b).

⁵⁵ <<http://www.splloc.soton.ac.uk>>

⁵⁶ Este proyecto también se desarrolla para el francés como L2. Es el *French Learner Language Oral Corpus* (FLLOC), corpus de L1 inglés y L2 francés.

La base de datos resultante contiene las grabaciones digitales (*wave file* y *MP3 file*), además de las transcripciones ortográficas y archivos en XML. En algunos casos, los archivos están también disponibles en el estándar CHILDES⁵⁷ (*Child Language Data Exchange System*) (McWhinney 1995, 2000).

Otro corpus formado por aprendices que comparten la misma lengua materna (inglés) es *The Spanish Corpus Proficiency Level Training (SPT)*⁵⁸, proyecto dirigido por Dale Koike de la Universidad de Texas. Se trata de una colección de vídeos, formada por entrevistas semidirigidas sobre diferentes temas de interés, cuya finalidad es ofrecer una base de datos transversales de E/LE destinada a estudiantes de español de nivel inicial a avanzado, a estudiantes de lingüística aplicada a la adquisición de una lengua extranjera y a futuros profesores de E/LE.

Los vídeos van acompañados de la transcripción ortográfica y de preguntas interactivas que orientan al usuario para descubrir las características de la interlengua que permiten identificar el nivel de los alumnos grabados.

Otro corpus homogéneo de aprendices de E/LE desarrollado para el estudio del reconocimiento del habla es el ***Corpus Oral de Aprendices Japoneses de Español LE***⁵⁹ de Mario Carranza (2011, 2013), de la Universidad de Estudios Extranjeros de Tokio. Es un corpus oral de estudiantes universitarios japoneses, registrados en los cursos académicos 2010-2011 y 2011-2012, de primer y segundo año de español (correspondientes al nivel A1 y A2 del *MCER*). De un total de 150 estudiantes, seleccionaron la producción de 30 para hacer la transcripción ortográfica y fonética de las tres tareas llevadas a cabo por los estudiantes: una conversación semiespontánea, otra espontánea y la lectura de un texto, todas ellas grabadas en momentos diferentes a lo largo de los dos años académicos.

El corpus está transcrito, etiquetado y codificado para procesarlo como corpus de entrenamiento de sistemas de reconocimiento del habla con aplicación en el desarrollo de herramientas de enseñanza de la pronunciación asistida por ordenador para el aprendizaje de E/LE.

⁵⁷ Es un sistema de codificación y transcripción del habla infantil usado internacionalmente en los estudios de adquisición de la L1.

⁵⁸ <<http://www.laits.utexas.edu/spt/>>

⁵⁹ Sirve como complemento al Corpus CbLLE (<<http://cbllle.tufs.ac.jp/tag/es/search.php?menulang=en>>) de hablantes nativos de español realizado, también, en dicha universidad.

Un apartado específico merecen tres corpus homogéneos que, usando la metodología de PRESEEA, recogen la producción de aprendices de español italianos (CIELE), sinófonos (COBICE) y sinófonos taiwaneses (COET), que se describen a continuación.

El *Corpus de conversaciones en italiano y en español LE (CIELE)*⁶⁰, creado por Pascual (2015) de la Universidad L'Orientale de Nápoles⁶¹ recopila 45 conversaciones espontáneas en parejas, 30 en español y 15 en italiano, que llevaron a cabo 60 estudiantes universitarios italianos de E/LE sobre un tema libre. Las conversaciones están divididas en tres grupos, en función del nivel de los estudiantes: i) el grupo A incluye 15 conversaciones de español de nivel *B1*; ii) el grupo B, 15 diálogos de español de nivel *C1* y iii) el grupo C, 15 interacciones en italiano llevadas a cabo por los informantes del grupo B, por tanto de nivel *C2* (nativo). Las conversaciones no se transcribieron completas, puesto que se descartaron los inicios (aperturas, saludos, preámbulos), si bien aparecen sin turnos cortados; se dejaron los intervalos centrales porque eran el objeto de estudio fundamental para el análisis sobre la evolución de la interlengua. Para las transcripciones se siguieron las directrices de los proyectos Val.Es.Co. y PRESEEA.

En segundo lugar, el *Corpus oral del español en Taiwan (COET)* (Rubio 2009) recoge 8 conversaciones orientadas a tareas entre estudiantes universitarios taiwaneses de E/LE grabadas fuera del aula en 2004⁶². Las grabaciones se transcribieron según las convenciones del grupo PRESEEA y se usaron para estudiar las interrupciones en la conversación (Rubio 2013a, 2013b).

Finalmente, el *Corpus oral bilingüe chino español (COBICE)* (Pérez Ruiz 2011) está formado por 20 conversaciones semidirigidas, en parejas, cara a cara sostenidas en contexto académico por estudiantes universitarios taiwaneses de cuatro universidades diferentes. Una parte de este corpus procede de COET (Rubio 2009). Las conversaciones

⁶⁰ <http://www.linred.es/numero12_corpus-1.html>

⁶¹ Este corpus se incluye en la tesis doctoral titulada *El funcionamiento de los turnos de apoyo en la conversación de estudiantes italianos de E/LE*, dirigida por la Dr. Ana M^a Cestero de la Universidad Antonio de Nebrija.

⁶² Dicho corpus pertenece al trabajo de investigación de Pérez Ruiz tutelado para la obtención del DEA (no publicado) *La interrupción y la superposición de habla en la conversación de estudiantes taiwaneses de español como lengua extranjera* (2008), dirigido por la Dra. Ana M^a Cestero de la Universidad Antonio de Nebrija (Rubio 2009).

tienen una duración de aproximadamente 10 minutos cada una; sin embargo, se transcribieron solo 4 minutos de cada una de ellas. Para la transcripción y el etiquetado se siguieron las normas del grupo PRESEEA. Estos datos se utilizaron para investigar la producción de turnos de apoyo en las conversaciones de estudiantes taiwaneses.

A continuación, se presentan los corpus orales no homogéneos de español no nativo, es decir, que ofrecen la producción de aprendices del mismo nivel de competencia pero distintas L1.

Un ejemplo lo constituye el *Language Usage In Spanish L2/LE (Corpus LUIS-L2/LE)* de Lourdes Díaz (2007) de la Universidad Pompeu Fabra. Este corpus agrupa las producciones de hablantes de lenguas indoeuropeas (alemán, sueco, islandés) y no indoeuropeas (coreano y chino) que aprenden español en la Escuela Oficial de Idiomas y en la Universidad de Barcelona. Consta de entrevistas semiespontáneas y preguntas experimentales para describir las características sintácticas de la interlengua, en concreto el funcionamiento de los pronombres de sujeto, las relativas, las interrogativas y los determinantes. El corpus está transcrito ortográficamente según el estándar CHILDES y se anotaron los errores más frecuentes. Este corpus está integrado en el banco de datos abierto de adquisición bilingüe, *TalkBank* (McWhinney 1995, 2000), dentro del proyecto CHILDES (Véase apartado 2.3)⁶³.

Otro ejemplo de corpus heterogéneo es *A XML-tagged Spanish Learner Oral Corpus for Learner Language Research* o *Corpus Oral de Español como Lengua Extranjera (CORELE)* (Campillos Llanos 2014a, 2014b), de la Universidad Autónoma de Madrid (Laboratorio de Lingüística Informática). Se trata de un corpus de E/LE producido por un grupo de aprendices de L1 heterogéneo (portugués, italiano, francés, inglés, neerlandés, alemán, polaco, chino y japonés). El nivel de competencia de los aprendices de español abarca desde el nivel A2 al B1. El corpus recoge entrevistas semiestructuradas de 40 estudiantes universitarios (4 participantes de cada L1 de las lenguas consideradas) a fin de obtener una muestra comparable. Asimismo, se dispuso de un grupo de control formado por cuatro hablantes nativos de edad similar a los no nativos. Las entrevistas

⁶³ También, se ha publicado en formato libro acompañado de un CD (Díaz 2007) donde se seleccionan algunos datos orales del corpus.

semiestructuradas se transcribieron manualmente con el programa *Transana 2.05*⁶⁴ y se analizaron morfológicamente con el programa GRAMPAL (Moreno Sandoval y Guirao 2006)⁶⁵. Los criterios de transcripción siguen las convenciones del formato CHAT (herramienta de anotación del sistema CHILDES). La finalidad de este corpus es describir la interlengua de los aprendices y los errores más frecuentes por categorías. Se contemplan los errores de la producción oral en los diferentes niveles (léxico, gramatical, de pronunciación y pragmático-discursivo).

Por último cabe destacar el corpus comparable bilingüe español-italiano **CORINÉI** (González Royo 2011a, 2011b), coordinado por Carmen González Royo de la *Universitat d'Alacant*, bajo el proyecto *Teletándem* desde 2009. Recoge conversaciones espontáneas a distancia entre estudiantes universitarios italianos de E/LE (de nivel B2.2 a C1) de tercer y cuarto curso de la facultad de Lenguas y literaturas extranjeras de la Universidad de Salerno y de la Universidad Suor Orsola Benincasa de Nápoles, por un lado, y de estudiantes nativos españoles de la Facultad de Lengua y Traducción de la Universidad de Alicante, por otro. CORINÉI recopila la interacción oral mediante el programa de chat/videoconferencia *Skype* y *Pamela for Skype* (los participantes deciden si conectan o no la webcam). Se lleva a cabo en sus respectivas casas y a partir de un tema libre. Cada conversación tiene una duración de 15 minutos. La actividad didáctica de la grabación sirve para trabajar la producción, la recepción y la interacción oral entre aprendices de español e italiano como lengua extranjera. La actividad tiene una doble función: es una interacción real en una tarea extraescolar del curso y a la vez recopila datos para observar cómo gestionan la producción oral de la L2 en tiempo real, con informantes similares. Los dos grupos comparten parámetros sociolingüísticos (preparación, intereses, estudios, edad, etc.) y motivación.

CORINÉI cuenta con 86 participantes voluntarios (43 españoles y 43 italianos) que pueden incluir la actividad en su portafolio. Cada diálogo tiene asociada una ficha con los datos sociolingüísticos que los mismos aprendices participantes completan (sexo, edad, lugar de residencia, nivel de estudios, nivel de L2, L1, etc.). Tras la grabación los propios estudiantes realizan la transliteración ortográfica, como actividad de aula, lo

⁶⁴ Desarrollado en la universidad de Wisconsin. <www.transana.org>

⁶⁵ Una demostración de GRAMPAL está en: <<http://cartago.llf.uam.es/grampal/grampal.cgi>>

más fielmente posible y según su criterio, utilizando las etiquetas de funciones comunicativas a partir de las propuestas en el *PCIC* de nivel B facilitadas por el profesor, a partir del currículo.

A continuación, se presenta la tabla 2 en la que se recopila la información más relevante de los corpus orales de aprendices expuestos en esta sección, ordenados alfabéticamente.

CORPUS	L1	L2	TIPO DE DISCURSO/ TIPO DE TAREA	CARACTERÍSTICAS INTERLOCUROS/ DOMINIO DE LENGUA	SISTEMA DE REPRESENTACIÓN	FECHA DE GRABACIÓN	DISPONIBILIDAD	DIRECTOR DEL PROYECTO
CIELE	Italiano	Italiano, español	Conversación espontánea y tema libre	Nivel B1 y C1	Transliteración ortográfica, convenciones de Val.Es.Co y PRESEEA	Curso académico 2011-2012	Acceso electrónico gratuito: transcripción y audio	Consuelo Pascual Escagedo Universidad L'Orientale de Nápoles
COBICE	Chino	Español	Conversaciones semidirigidas por una lista temática	Nivel A2	Transliteración ortográfica, convenciones de PRESEEA	(no se indica la fecha)	No está disponible	Javier Pérez Ruiz Wenzao Ursuline College of Languages (Taiwan)
COET	Chino	Español	Conversación espontánea orientadas a tareas	Nivel B1 y B2	Transliteración ortográfica, convenciones de PRESEEA	2004	Acceso electrónico gratuito: transcripción	Miguel Rubio Universidad Tunghai (Taiwán)
CORINÉI	Italiano español	Italiano, español	Conversación espontánea y tema libre	L1 italiano: nivel B2.2 y C1 L1 español: nivel B1 y B2	Transliteración ortográfica y etiquetado de funciones	Cursos académicos 2009-2010 y 2010-2011	Acceso electrónico gratuito con permiso	Carmen González Royo Universitat d'Alacant
CORELE	Más de 9 lenguas: portugués, italiano, francés, inglés, neerlandés, alemán, polaco, chino y japonés.	Español	Entrevistas semiestructuradas; tenían que realizar la misma tarea: tareas de descripción y de narración	Nivel A2 y B1	Transliteración ortográfica con una adaptación del formato CHAT, sincronización de texto y audio, etiquetado de errores	(no se indica la fecha)	No está disponible	Leonardo Campillos Llanos Universidad Autónoma de Madrid

Corpus LUIS-L2/LE	Alemán, sueco, islandés, coreano, chino	Español	Semiespontáneo (entrevistas estructuradas) y experimental (cuestionarios estructurados)	Diferentes niveles	Transcripción ortográfica y anotación de errores	2007	Acceso electrónico gratuito	Lourdes Díaz Rodríguez Universitat Pompeu Fabra
Corpus oral de aprendices japoneses de español LE	Japonés	Español	Tres tareas: conversación semiespontánea, conversación espontánea y lectura	Primer y segundo curso de español en la Universidad de Estudios Extranjeros de Tokio (A1 y A2 del MCER)	Transcripción ortográfica y fonética	Cursos académicos: 2010-2011 y 2011-2012	No está disponible	Mario Carranza Universidad de Estudios Extranjeros de Tokio
Corpus oral del proceso de adquisición del español como L2 por una joven inmigrante búlgara	Búlgaro	Español	Entrevista dirigida	2º nivel de la Escuela Oficial de Idiomas	Transcripción ortográfica	1997	Acceso electrónico gratuito: transcripción	Carmen Aguirre Félix Villalba Maite Hernández Miriam Najt
SPLLOC	Inglés	Español	Entrevistas semiepontáneas, narraciones y descripciones de fotos	De nivel principiante a nivel avanzado	Transcripción ortográfica con el estándar CHILDES; los archivos de audio en distintos formatos	2006-2010	Acceso electrónico gratuito	Laura Domínguez University of Southampton, Reino Unido
SPT	Inglés	Español	Conversaciones semidirigidas por una lista temática	De nivel inicial a avanzado	Transcripción ortográfica	2010-2011	Vídeos disponibles	Dr. Dale Koike University of Texas

TABLA 2. *Corpus orales de aprendices de español como L2/LE*

2.3. Modelos nativo y no nativo: El enfoque psicolingüístico de corpus y su metodología

Tras el análisis de los corpus de aprendices, con especial atención a las bases de datos que recogen el habla de los estudiantes de una LE/L2, a continuación en esta sección se exponen los corpus lingüísticos agrupados desde otro enfoque metodológico: el psicolingüístico. Se tratará de los corpus procedentes del ámbito de la psicología y de la psicolingüística, normalmente de carácter longitudinal, cuyo objetivo principal inicialmente era describir el desarrollo de la competencia lingüística infantil (monolingüe, bilingüe y multilingüe) y que, posteriormente, se amplió a otras franjas de edad y a otras situaciones de contacto: de bilingüismo, multilingüismo (adultos y segundas lenguas incluidas), patologías (afasias) y pérdida de lenguaje (*attrition*). Un ejemplo muy conocido de este tipo de corpus lo constituye el corpus *Child Language Exchange System* (**CHILDES**) (McWhinney 1995, 2000) creado inicialmente para ilustrar y documentar la adquisición infantil monolingüe, bilingüe y multilingüe en distintos contextos y L1s, y que actualmente forma parte de un proyecto mucho más amplio, **TalkBank**⁶⁶ (Carnegie Mellon University y University of Pennsylvania), que incluye los otros supuestos que mencionamos antes (adultos, lenguas segundas, patologías, etc.). Estos corpus incluyen muestras procedentes de los intercambios verbales infantiles en situaciones naturales para observar las características del habla espontánea de los niños con sus cuidadores. Con esta finalidad se constituye una base de datos representativa del lenguaje real de los niños en situaciones de cotidianidad para establecer patrones adquisitivos de referencia, consignando las características más destacables a partir de las convenciones y normas de CHILDES y, así, poder recuperar los datos etiquetados (con sus propias herramientas). Una versión en español de estas normas la constituye el artículo “Registro, Transcripción y Análisis de Muestras de Habla Espontánea (RETAMHE)” (Diez-Itza et al. 1999). Las herramientas del sistema CHILDES son una base de datos informatizada; un sistema de codificación con unas convenciones específicas (CHAT) para obtener archivos

⁶⁶ <<http://talkbank.org>>

CHAT dentro de la Base de datos (DATABASE); y el paquete de programas CLAN para hacer recuentos automáticos o semiautomáticos de frecuencias, búsquedas de palabras, análisis de la interacción, entre otros.

Tanto CHILDES como *Talkbank* son una de las referencias centrales en psicolingüística, tanto para la adquisición infantil como para la multilingüe. Ambas bases de datos están integradas en una donde se pueden incluir transcripciones de habla infantil (L1), de habla bilingüe, de alteraciones del lenguaje y de una L2/L3/L4; en ocasiones los investigadores incluyen los audios e incluso los vídeos.

En este capítulo, hemos visto las características y las metodologías incorporadas en los corpus disponibles, que contienen producciones nativas y no nativas, contrastadas o no por lenguas, por niveles y por géneros textuales, pero queda un espacio por cubrir. Se trata del que ocuparía un tipo de corpus que incluyera simulaciones de tareas específicas realizadas tanto por hablantes nativos como por no nativos (aprendices de distintos contextos, L1, edades, etc.), en el transcurso de situaciones no simuladas (en L1 y L2) para poder contrastar y describir la variación tanto nativa como no nativa, por un lado, y la variación debida al carácter espontáneo o simulado de la situación, por el otro. Este último tipo de corpus que sugerimos crear constituiría lo que hemos denominado en este trabajo *corpus de aprendizaje*⁶⁷.

El corpus que desarrollamos en esta investigación, *Ferroviele*, y que presentamos en el siguiente capítulo (ver capítulo 3) sería un ejemplo de corpus de aprendizaje. Como veremos, *Ferroviele* está formado por distintos subcorpus —nativo espontáneo, nativo simulado, no nativo simulado no entrenado, no nativo simulado entrenado anglófono (estadounidenses) y no nativo simulado entrenado multilingüe— elicitados para resolver la misma tarea comunicativa (interacción para obtener información y compra-venta de billetes de tren) con el objetivo de obtener el modelo

⁶⁷ No confundir con el *aprendizaje automático* o “machine learning” que generaliza comportamientos a partir de una información no estructurada a partir de una serie de programas. Tiene aplicaciones en numerosas ramas de la ciencia (procesamiento del lenguaje natural, sistemas de recuperación de la información, análisis de imágenes, ciencias biológicas entre otras).

lingüístico de referencia nativa para los aprendices no nativos y, a la vez, recopilar los datos de los no nativos para caracterizar su interlengua.

Este tipo de corpus permite un enfoque de estudio que va desde una perspectiva *macro* —o de pragmática intercultural, interlingüística (Bardovi-Harlig y Hartford 1997; Blum-Kulka et al. 1989; Kasper 1992; Kramsch 1993)— a una perspectiva *micro* —sociopragmática entre hablantes de la propia L1 y en distintas circunstancias (Barron 2005)— en el área de la lingüística aplicada, destinado a describir la competencia pragmalingüística requerida en un escenario comunicativo concreto. En el escenario analizado en esta tesis, petición de información y compra-venta de billetes de tren, los datos de producción abarcan desde los niveles B1 a B2, correspondientes a la expresión oral no nativa, por un lado; y, por otro, al nivel C2 (por discutida que sea la etiqueta) reflejado por los nativos.

Dado que la finalidad de este trabajo es identificar el inventario básico de transacción de información y compra-venta de billetes de tren, el corpus creado tiene dos tipos de aplicación posible: una directa (como modelo para determinar los elementos que enseñar en una determinada situación comunicativa) y una indirecta (para evaluar la mayor o menor proximidad a la variación nativa, además del ajuste al nivel no nativo del *MCER* esperado). En los capítulos siguientes se presentará la metodología prevista para la construcción y análisis del corpus a partir del escenario comunicativo elegido y su explotación.

CAPÍTULO 3. METODOLOGÍA

En esta sección se presenta la propuesta metodológica basada en corpus llevada a cabo en esta tesis, de acuerdo con los presupuestos de los capítulos anteriores. Para acometer la descripción y posterior realización de un inventario de formas gramaticales correspondientes al escenario comunicativo *Transacción de información y compra-venta* como modelo para la producción no nativa en ELE, se ha creado el corpus oral *FerroviELE*, que se distribuye como se recoge en la figura 1. Dicho corpus está formado por un conjunto de conversaciones coloquiales (diálogos) sobre el ámbito del transporte cuyo estudio pragmático-funcional se aborda para descubrir el funcionamiento de esta situación comunicativa en el curso de la interacción cara a cara. En particular, se toma como punto de partida para el análisis funcional los repertorios del inventario básico propuestos por EAQUALS⁶⁸ (*European Association for Quality Languages Services*) del British Council (2010) para el inglés, la propuesta de descriptores del *Vantage Level* del Consejo de Europa (van Ek y Trim 2001: 27-59) para el inglés como segunda lengua; y el *Plan Curricular del Instituto Cervantes* (2006) y la *Guía de contenidos lingüísticos por niveles del español* (Díaz, Martínez y Redó 2011) para el español como lengua extranjera.

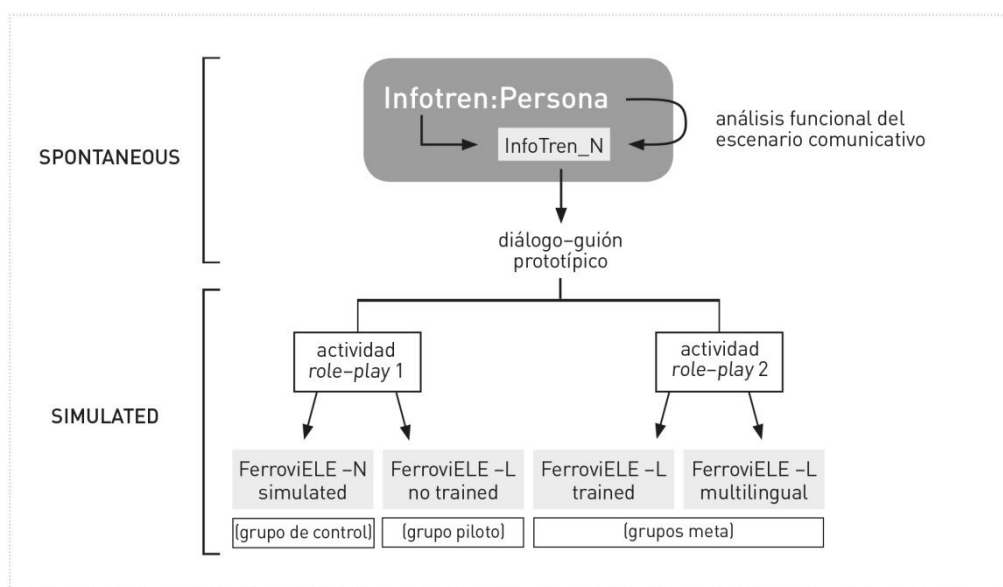


FIGURA 1. Distribución de los (sub)corpus y las metodologías empleadas

⁶⁸ El objetivo básico de esta agencia es garantizar la calidad en las instituciones de enseñanza de segundas lenguas, ligada al British Council.

Esta sección se organiza en cuatro apartados que reflejan las cuatro fases de desarrollo del corpus *Ferroviele*: 1) diseño de las pruebas y definición de grupos meta por niveles y lenguas maternas para la obtención del corpus (sección 3.1); 2) elicitación y obtención de datos orales para la creación del corpus (sección 3.2); 3) transcripción de las producciones nativas y no nativas (sección 3.3); y 4) anotación morfológica y pragmático-funcional del corpus (sección 3.4).

3.1. Diseño para la creación del corpus *Ferroviele*

Ferroviele es un corpus oral del español formado por producciones de hablantes nativos y no nativos (aprendices L2, L3 o L4) que incluye un total de 50 diálogos entre 100 interlocutores de edades, sexos y niveles socioculturales diferentes, en los que se produce una transacción de información sobre servicios ferroviarios, en concreto, la petición de información y compra-venta de billetes de tren para un recorrido de larga distancia. Dicha transacción siempre se produce entre dos interlocutores: un operador y un cliente. El número total de intervenciones es de 1684, de las que 851 corresponden al operador y 833 al cliente. La duración de los diálogos grabados es aproximadamente de dos horas.

Ferroviele está formado por cinco (sub)corpus diferentes que corresponden a tres metodologías distintas de obtención de los diálogos. Las diferencias se deben a que la primera consiste en una grabación secreta de la producción nativa espontánea, que corresponde al corpus *Infotren-N*, a partir del cual se elabora el diálogo prototípico entre operador y cliente (véase 3.1.1). La segunda corresponde a una simulación del escenario por parte de nativos, que da como resultado el corpus *Ferroviele-N*, que será utilizado como corpus de control (véase 3.1.2). La tercera consiste en una secuencia de preparación de la tarea como *role-play* en el aula con aprendices no nativos, que da como resultado los corpus *Ferroviele-L*, que son tres, *Ferroviele-L_no trained*, *Ferroviele-L_trained* y *Ferroviele-L_trained multilingual*⁶⁹ (véase 3.1.3). En la

⁶⁹ Se mantienen los nombres en inglés de los subcorpus así como de las etiquetas de identificación de los interlocutores porque fue como se marcaron previamente en la herramienta de anotación.

figura 1 se presenta de manera esquemática las metodologías empleadas para la creación de los subcorpus.

Asimismo, en la tabla 3, se detallan las características de cada uno de los cinco corpus mencionados, consignando la información acerca de los diferentes informantes y tipos de diálogo.

Corpus	Nº de diálogos	Nº de interlocutores	Tipo de discurso	Características de los informantes	Durac. subcorp. MIN:SEG	Fecha de grabación
InfoTren-N(atives)_spontaneous	16	32	Diálogo telefónico espontáneo	Nativos, diferentes edades, niveles socioculturales, ambos sexos	27:30	1997
FerrovieELE-N(atives)_simulated	6	12	Diálogo semiguado (<i>role-play</i>)	Nativos, diferentes edades, niveles socioculturales, ambos sexos	7:47	2012
FerrovieELE-L(earners)_no trained	6	12	Diálogo semiguado (<i>role-play</i>)	No nativos, estudiantes universitarios, ambos sexos, estadounidenses anglófonos B1	12:23	2012
FerrovieELE-L(earners)_trained	6	12	Diálogo semi-controlado, semiguado (<i>role-play</i>)	No nativos B1, estudiantes universitarios, ambos sexos, estadounidenses anglófonos B1	11:22	2013
FerrovieELE-L(earners)_trained multilingual	16	32	Diálogo semi-controlado, semiguado (<i>role-play</i>)	No nativos, estudiantes universitarios, ambos sexos, europes anglogermánicos, francófonos y lenguas orientales B2	53:40	2013
TOTAL	50	100			1h 52:42	

TABLA 3. *FerrovieELE corpus: resumen de las características de los subcorpus*

En principio, se recogieron 6 diálogos para cada subcorpus (muestra sub6) y, tras examinar algunos datos, se decidió ampliar dos de los subcorpus, *FerrovieELE-L_trained multilingual* y *Infotren-N_spontaneous*, con diez diálogos más en cada uno de ellos, porque pareció relevante asociar el multilingüismo a la experiencia experta o nativa. Es por este motivo, que estos corpus contienen 16 diálogos cada uno (muestra sub16).

Es cierto que la simulación (*FerrovieELE-L_no trained*) y el entrenamiento (*FerrovieELE-L_trained*) a monolingües en B1 (ambos estadounidenses) parecían dar alguna diferencia, si bien esta se disparaba en el subcorpus de los multilingües (B2). Por ello, se decidió ampliar la muestra en el subcorpus *FerrovieELE-L_trained multilingual* y *Infotren-N_spontaneous* para comparaciones específicas entre nativos expertos y no nativos no expertos entrenados multilingües, comparación que parecía más interesante que solo utilizar los datos de no nativos B1 y B2.

A continuación se presentan los corpus mencionados de forma más detallada.

3.2. Elicitación y obtención de datos orales para la creación del corpus *FerrovieELE*

En esta sección se describe de manera detallada el porqué y el cómo se ha obtenido cada uno de los distintos corpus que conforman *FerrovieELE*. Se hace referencia tanto a los mecanismos de elicitación, a las características de los informantes, a los roles que desempeñan en la interacción, al grado de espontaneidad o inducción de los datos (es decir, el grado de control para las simulaciones), así como a los guiones didácticos siempre que se han usado (según los destinatarios sean simulados o no), al uso y características de los cuestionarios de contacto lingüístico y a las pruebas de nivel de lengua extranjera cuando ha correspondido.

3.2.1 Modelo nativo: *InfoTren-N(atives)_spontaneous*

El corpus *InfoTren-N* parte de un corpus oral nativo espontáneo ya existente, *InfoTren:Persona* (Bonafonte et al. 2000). *InfoTren:Persona* incluye diálogos entre el operador de una compañía ferroviaria (RENFE) y un usuario. Este corpus forma parte

del proyecto *Sistema de diálogo para habla espontánea en un dominio semántico restringido*, también conocido como proyecto BASURDE⁷⁰. *Infotren:Persona* consiste en 204 transcripciones de llamadas telefónicas reales, grabadas secretamente, realizadas al servicio de información de RENFE (Barcelona), durante noviembre de 1997. En él prevalecen las conversaciones en español en las que el usuario pide información sobre las conversaciones de reservas. En los diálogos de reservas en el momento de hacer la transacción económica se corta el diálogo como garantía de confidencialidad. La mayoría de los diálogos hacen referencia a trenes de largo recorrido y la duración total de los diálogos es de aproximadamente 7 horas de habla espontánea.

El grupo de informantes lo constituye un total de 205 clientes anónimos y los operarios correspondientes que les atendieron en el servicio de atención al cliente. Los diálogos se organizan en turnos, dando como resultado un total de 7739 intervenciones, de las que 3949 corresponden al operador y 3790 al cliente. El número de palabras de media en el turno del operador es de 10; y en el del cliente, de 6 (Bonafonte y Lleida 1999: 2-4).

A partir de *Infotren:Persona* se realizó un análisis del escenario comunicativo, en concreto de los propósitos funcionales de éste, y se llevó a cabo una sistematización funcional de los diálogos, extrayendo los patrones recurrentes (fórmulas de inicio, cierre, negociación, petición, agradecimiento, expresión de preferencias, etc.). Se distingue si son diálogos de petición de información, de compra-venta o de reserva, y se sistematizan los contenidos funcionales y sociopragmáticos (negociación, confirmaciones, etc.) de cada uno de ellos, distinguiendo los turnos relativos al operador de los del usuario en dichos diálogos. Todo ello con el objetivo de extraer un guion o un diálogo prototípico entre operador-cliente que sirva para modelar la simulación y que permita llevar a cabo la posterior elicitación mediante la tarea, ya en la clase de lengua extranjera (con informantes no nativos); y, también, en la situación de laboratorio, con informantes nativos (grupo de control). Estos grupos de datos así obtenidos constituyen el modelo de competencia nativa real (no idealizada) en el escenario comunicativo de transacción de información y compra-venta de billetes.

⁷⁰ BASURDE URL: <http://gps-tsc.upc.es/veu/basurde>

De *Infotren:Persona* se han seleccionado 16 diálogos considerados representativos, de acuerdo con el análisis realizado, que constituyen el corpus *InfoTren-N_spontaneous*. Dicho subcorpus tiene una duración de 27'30", que contiene un total de 692 intervenciones (353 operador y 339 cliente). El promedio de palabras del operador es 7,03 y del cliente 4,54. A partir de *InfoTren-N_spontaneous* se ha extraído el guion prototípico de interacción operador-cliente para crear las situaciones simuladas de las que surgirán, posteriormente, los cuatro subcorpus no espontáneos simulados nativos y no nativos en el aula, a saber:

- a) *Ferroviele-N(atives)_simulated* (7'47");
- b) *Ferroviele-L(earners)_no trained* (12'23");
- c) *Ferroviele-L(earners)_trained* (11'22") y
- d) *Ferroviele-L(earners)_trained multilingual* (53'40").

3.2.2. Elicitación de datos: Diseño de la tarea por simulación

A partir del diálogo-guion prototipo obtenido a partir del corpus *InfoTren:Persona*, se elaboran dos guiones adaptados didácticamente del proceso de petición de información y compra-venta de billetes, a modo de actividad *role-play*.

La primera actividad *role-play 1* incluye: i) una muestra de interacción similar a la que han de llevar a cabo los estudiantes; ii) la atribución de roles —operador o cliente— especificando las características del cliente-viajero (estudiante con poco dinero que quiere ir a Granada o hombre/mujer de negocios que quiere viajar a Sevilla) e iii) información para que el operador pueda ofrecer el servicio (horarios, tipos de tren, tarifas, servicios, etc.) (véase anexo 2: actividad *role-play 1*_diálogo tren). Esta actividad *role-play 1* es la que se utilizó con el grupo de control (nativo), *Ferroviele-N_simulated*, y con el grupo de aprendices no entrenado (grupo piloto), *Ferroviele-no trained*. Esta actividad duraba aproximadamente 5 minutos.

La segunda actividad *role-play 2*, además de la atribución de roles, incluye también información sobre el ámbito del transporte ferroviario en España y contenidos socioculturales de este ámbito. Es decir, información sobre tipos de billetes, tipos de tren, con especial énfasis en los trenes de larga recorrido, servicios, etc., y toda esta información se apoya, además, con imágenes y videos (véase anexo 3: actividad *role-*

play 2_ para entrenar). Esta actividad es la que se administró a cada uno de los grupos no nativos de los distintos centros docentes que participaron en este experimento⁷¹, *FerrovieELE-L_trained* y *FerrovieELE-L_multilingual*. Esta actividad duraba entre 20 y 30 minutos.

Para controlar el impacto de la competencia sociolingüística y sociocultural de los informantes no nativos, previamente, se les pasó un cuestionario de contacto lingüístico (véase anexo 4: cuestionario contacto lingüístico_institución y alumnos). Este cuestionario permitió establecer los perfiles de alumnos y diferenciar su saber práctico —en el ámbito específico del transporte— del saber no declarativo⁷² que constituye la competencia lingüística en esta simulación, saberes que se pondrán en juego en las pruebas de simulación posteriores.

CUESTIONARIO DE CONTACTO

El cuestionario de contacto lingüístico (Dörnyei 2010) consiste en un conjunto de preguntas organizadas en cuatro bloques que recogen información sobre:

- (1) perfil general: nombre, edad, nacionalidad, estudios, etc.;
- (2) perfil lingüístico: lengua materna, lenguas que hablan, etc.;
- (3) información sobre la familiaridad con viajes: en tren, en avión, etc., y
- (4) petición de colaboración como voluntarios.

Los bloques 1 y 2 recogen información personal y experiencia lingüística previa. Los bloques 3 y 4 tienen relación directa con el uso de transporte público, el uso de taquillas con un trabajador de la compañía, el taquillero, o con la compra por Internet o por vía telefónica de productos o servicios de transporte.

⁷¹ Se agradece a los centros que participaron en esta investigación: *Arcadia University, The College of Global Studies* (<<http://studyaborad.arcadia.edu>>); *Universitat Pompeu Fabra* y *Escuela Oficial de Idiomas de Barcelona, EOI Barcelona Drassanes*. La recogida de datos estuvo supeditada al curso académico del centro (2012-2013).

⁷² El *MCER* distingue, por un lado, el conocimiento sociocultural y del mundo que se incluyen en el componente “Conocimiento declarativo” del apartado 5.1. *Las competencias generales*. Por otro lado, el conocimiento lingüístico, sociolingüístico y pragmático que se incluyen en el apartado 5.2. de *Las competencias comunicativas* de la lengua. Por lo tanto, se trata de una distinción entre ambos tipos de saberes (puede verse el capítulo 5: *Las competencias del usuario o alumno, del MCER*). Véase también Kramsch 2001.

El cuestionario se pilotó previamente con sujetos de niveles de español como lengua extranjera A2 y B1⁷³ para comprobar, por un lado, si la dificultad de las preguntas en español se ajustaba al nivel de lengua de los alumnos y no perder información relevante para la elaboración de los perfiles y, por otro lado, para la agrupación de los informantes en los diferentes subcorpus. Este cuestionario no se proporcionó al grupo de control nativo.

La realización del cuestionario de contacto lingüístico no presentó dificultades reseñables, aunque se constató una cierta reticencia ante la reiteración de algunas preguntas (redundancia buscada para asegurarnos la comprensión). Por ello, se decidió hacer una versión simplificada del bloque 3, más ligera y en un formulario más visual, que se pudiera responder con cruces u otras marcas gráficas y con menor coste de lectura y producción escrita del alumno (la versión recogida en el anexo 4, Cuestionario contacto lingüístico_institución y alumnos, es la definitiva).

3.2.3. *FerrovieLE-L(Learners)_no trained*

El corpus *FerrovieLE-L_no trained* incluye los diálogos resultado de la simulación de la actividad *role-play 1* al grupo piloto formado por alumnos de nivel B1. Se consideró que el nivel de dificultad en la producción estaba por encima de las posibilidades de los alumnos de nivel A2, especialmente en el rol del operador. Por esta razón finalmente se desestimó la participación de alumnos de nivel A2.

FerrovieLE-L_no trained es una muestra de 6 diálogos semiespontáneos o semiguidados producidos por aprendices no nativos anglófonos (estadounidenses) de nivel B1 de E/LE, a los que se les había administrado el cuestionario de contacto lingüístico anteriormente descrito. Por semiguidado se entiende que los informantes en la interacción tenían un modelo de diálogo previo, a partir del cual se realizó la actividad *role-play 1*, dividiendo a los alumnos en dos grupos según el rol de operador o de cliente. Todos tenían un modelo de diálogo básico, así como una lista de

⁷³ Nivel de aprendizaje que establece el *Marco Común Europeo (MCER)*. El *MCER* define la progresión en el aprendizaje de las lenguas, diferenciando tres grandes etapas y cada una de ellas subdivididas en dos: A1, A2; B1, B2; C1, C2.

supuestas necesidades de transporte y exigencias de horarios y servicios en la que basar las preguntas y las respuestas (estructuras modelo, como por ejemplo, ¿podría informarme...?, ¿Quieres viajar...?, hay un tren a las 11h...) (ver roles y listas para operador y cliente en el anexo 2: actividad *role-play 1*_diálogo tren).

Antes de la grabación, los alumnos trabajaron durante 10 minutos aproximadamente con un modelo de diálogo y, además, con el manual de clase (que solo toca tangencialmente aspectos del viaje, sobre todo las anécdotas y el relato del viaje en sí, pero no la obtención del servicio). Realizaron la actividad de simulación en clase y se les grabó con una grabadora SONY-IC Recorder ICD-UX512. El tutor que dirigió la actividad fue la profesora habitual del grupo, Mónica Caballero (que es la investigadora).

Se registró en audio un total de seis simulaciones telefónicas correspondientes a seis parejas de estudiantes. La duración de los diálogos oscila entre un mínimo de 1'56" y un máximo de 2'23". La duración total del conjunto de producciones es de 12'23", que contiene un total 103 intervenciones (67 operador y 63 cliente). El promedio de palabras del operador es 6,85 y del cliente 4.44.

El corpus *FerrovieLE-L_no trained*, si bien desarrolla diálogos semiguidados, no entrena a los estudiantes en el ámbito del transporte en tren y no les adiestra en el conocimiento declarativo⁷⁴; es decir, no les ofrece información sobre cómo está organizado el mundo de los servicios ferroviarios en España, ni conocimiento sociocultural alguno, ni favorece la reflexión intercultural en este ámbito.

Estas simulaciones de situación de interacción en una taquilla se grabaron con estudiantes voluntarios del centro *Arcadia University, The College of Global Studies*. A estos estudiantes, previamente, se les pasó una prueba de nivel estandarizada para

⁷⁴ Es uno de los saberes que tiene en cuenta el *MCER*. Este concepto deriva de las teorías cognitivas de la psicología de aprendizaje. Se considera conocimiento declarativo siempre que pueda comunicarse a través del lenguaje verbal.

En muchas ocasiones el éxito de un intercambio comunicativo reside en la forma en que los interlocutores representan la realidad, así que los significados compartidos entre los miembros de una misma comunidad cultural, que se comunican a través de una lengua en común, son fundamentales. Es difícil sistematizar el tipo de conocimiento del mundo que resulta relevante en una lengua extranjera. El *MCER* ofrece parámetros para que se seleccione el conocimiento declarativo que sea relevante para el aprendiz según sus necesidades de aprendizaje (Llorián 2007: 146).

confirmar el nivel de lengua extranjera⁷⁵. Todos ellos eran nativos estadounidenses, universitarios, de edades comprendidas entre los 20 y los 23 años, anglófonos y de ambos sexos.

El corpus *FerrovieLE-L_no trained* funcionó de piloto para la creación de los corpus *FerrovieLE-L_trained*.

3.2.4. *FerrovieLE-L(earners)_trained* y *FerrovieLE-L(earners)_trained multilingual*

Se presentan en un único apartado los dos subcorpus de aprendices no nativos entrenados, *FerrovieLE-L_trained* y *FerrovieLE-L_trained multilingual*, porque comparten todas las características metodológicas y divergen solo en la homogeneidad lingüística del grupo.

El corpus *FerrovieLE-L_trained* es una muestra de 6 diálogos semiguidados y semicontrolados de aprendices no nativos anglófonos (estadounidenses) de nivel B1, con una duración de 11'22", que contiene un total de 113 intervenciones (56 operador y 57 cliente). El promedio de palabras del operador es 9,17 y del cliente 5,13.

El subcorpus *FerrovieLE-L_trained multilingual* es una muestra de 16 diálogos, también, semiguidados y semicontrolados de aprendices no nativos europeos anglogermánicos, francófonos y de lenguas orientales, de nivel B2. El subcorpus *trained multilingual* tiene una duración total de 53'43", que contiene un total de 644 intervenciones (323 operador y 321 cliente). El promedio de palabras de operador es 7,85 y del cliente 5,01.

Los dos corpus son semiguidados, como ocurría con el subcorpus *FerrovieLE-L_no trained*, pero además son, también, semicontrolados. Por semicontrolado se entiende que los informantes fueron entrenados por un profesor con una propuesta de explotación didáctica completa durante una sesión previa a la simulación. En los subcorpus de los aprendices *trained* y los aprendices *multilingual* de lenguas orientales la tutora fue la profesora Mónica Caballero y, para el resto de participantes del subcorpus *multilingual* la profesora fue Lourdes Díaz.

⁷⁵ Cada centro tiene sus pruebas de nivel. Sin embargo, suelen ser muy estandarizadas porque se rigen según los niveles del *MCER*.

La tarea de entrenamiento tuvo una duración de 30 minutos aproximadamente. El material consta de 4 hojas. La primera hoja da información sobre conocimiento del mundo ferroviario en España, del conocimiento sociocultural y permite trabajar la concienciación acerca de las diferencias interculturales. Se les explican los tipos de trenes (de largo recorrido —AVE, grandes líneas, regionales, cercanías—; los tipos de asientos—individual, cabinas—; y otros servicios disponibles: clase (por ejemplo, clase turista, preferente); forma de pago (en metálico, con tarjeta), etc. Se presentan acompañados de los términos específicos (por ejemplo, coche-bar, coche-restaurante, revisor, litera, andén, etc.). A continuación, en la segunda hoja, se incluye una dirección web, extraída de *Youtube*; se proyecta el vídeo en el aula y se les muestra el interior del AVE, con el que algunos podrían no estar familiarizados en su experiencia en España con trenes regionales y locales. Con toda la información ofrecida, tienen que completar dos ejercicios: a) una comprensión lectora en que el texto es un billete de RENFE y para la que deben contestar unas preguntas y b) un ejercicio de selección múltiple que consiste en revisar el nuevo vocabulario⁷⁶. Tras haberles introducido en el ámbito del transporte ferroviario, se continúa con la actividad de grabación del diálogo de compra de billetes de tren basado en un *role-play 2*, tercera hoja (ejercicio 3) (ver anexo 3: actividad *role-play 2*_para entrenar).

Todos los diálogos se grabaron con sus móviles y se enviaron posteriormente por correo electrónico a la profesora, ya que se dio el ejercicio como tarea para casa a fin de no tomar tiempo de las actividades oficiales ordinarias. Esto fue así para todos, excepto para los estudiantes orientales.

Para realizar la actividad *role-play 2* se les asignó aleatoriamente el rol de cliente y operador y se le dio al viajero (cliente) una lista con sus supuestas exigencias y necesidades y al operador otro guion con los servicios que podía ofertar. El mecanismo de la tarea de elicitación es la negociación entre los diferentes roles, por eso se les da unas características de la situación de información y compra-venta de billetes de tren (ir a un concierto y tener poco dinero), pensadas para que hubiera negociación y que la

⁷⁶ Para el entrenamiento sobre el sistema ferroviario en España (información y ejercicios 1 y 2) se toma la propuesta de *Bellver y San Andrés* (2010) con alguna variación. La actividad 3 para grabar el diálogo es nuestra.

interacción pudiera dar pie a la aparición de diferentes elementos propios de la conversación coloquial. La actividad no tenía, pues, una finalidad didáctica, sino estrictamente la de recabar datos. Por todo ello, el diálogo previsto era semiguado y semicontrolado y no se trató de producción libre ni espontánea, ni tuvo finalidad de evaluación.

A diferencia del corpus *FerrovieELE-L_no trained*, que sirvió de piloto a la hora de realizar el experimento de aprendices, en los subcorpus de aprendices *trained*, al estar entrenados, valga la redundancia, se trataba de un grupo cerrado o más controlado de modo que no se les pasó el cuestionario de contacto lingüístico.

Los informantes del corpus *FerrovieELE-L_trained* pertenecían al centro *Arcadia University* y la mayoría de los participantes del corpus *FerrovieELE-L_trained multilingual* procedían de la *Universitat Pompeu Fabra*, a excepción de los estudiantes orientales, que eran voluntarios de la *Escuela Oficial de Idiomas* de Barcelona, Drassanes.

La composición de los corpus de aprendices está condicionada por el acceso a alumnos en período lectivo y a su disponibilidad y predisposición para participar como voluntarios. También, a la homogeneidad lingüística o no del grupo (multilingüe u homogéneo).

A los estudiantes que participaron en la elaboración de los corpus *FerrovieELE-L_trained* y *FerrovieELE-L_no trained* se les administró una prueba objetiva de nivel o clasificación (la misma) para garantizar que los niveláramos a todos por igual.

3.2.5. *FerrovieELE-N(atives)_simulated*

El objetivo del corpus *FerrovieELE-N_simulated* es servir de referencia para comparar las diferencias y las similitudes entre las interacciones de parejas no nativas, siempre en situación artificial de simulación, porque la interacción no se produce en RENFE. También sirve para comparar las interacciones de parejas nativas en una situación espontánea.

Los participantes en la simulación son profesionales en otras áreas (diseñadores gráficos, secretarías, contables). No son operadores cualificados, ni estudiantes de lenguas extranjeras, ni tampoco multilingües.

Este corpus lo constituye una muestra de 6 diálogos semiguidados de nativos no vinculados a la docencia (no aprendices). La duración total de los diálogos de este grupo de control es de 7'47", que contiene un total 103 intervenciones (52 operador y 51 cliente). El promedio de palabras del operador es 8,32 y del cliente 5,94.

A este grupo se le administró el mismo guion de situación de petición de información y compra-venta de billetes de tren con los roles de operador y cliente. Se les colocó espalda con espalda para hacer más real la simulación telefónica y se les grabó en audio. En este grupo, obviamente, no fue necesario pasar el cuestionario de contacto lingüístico.

3.3. Grabación, transcripción y anotación de los diálogos

Tanto los estudiantes como los nativos de control sabían que estaban siendo grabados y nos dieron su consentimiento para utilizar dichas grabaciones para ser analizadas y procesadas con fines académicos.

Si bien la mayoría de las grabaciones se han realizado con una grabadora SONY-IC Recorder ICD-UX512, en algunos casos se han registrado con la grabadora de voz de dispositivos telefónicos móviles.

Una vez obtenidas las grabaciones de los 50 diálogos, se procedió a la transcripción ortográfica de los ficheros de audio MP3 de manera manual y a la anotación de información lingüística y extralingüística en formato XML, utilizando el programa *Transcriber*⁷⁷ (Barras et al. 2001). Esta herramienta permite la alineación entre el audio y la transcripción ortográfica, es decir, el material transcrito está sincronizado con el archivo de sonido.

La transcripción y la anotación se han realizado manualmente por un único anotador que ha seguido las recomendaciones del *Expert Advisory Group on Language Engineering Standard* (EAGLES 1996) y las convenciones del *Text Encoding Initiative* (TEI 2007, 2015).

En resumen, el corpus *Ferroviele* está formado por los archivos de voz alineados con la transcripción ortográfica, enriquecidos con anotación lingüística y

⁷⁷ *Transcriber*: <<http://trans.sourceforge.net/en/presentation.php>>

extralingüística. A la hora de etiquetar se aprovechan algunas etiquetas predefinidas de *Transcriber*, aunque también se han creado de nuevas. A continuación se describen de forma detallada los criterios y las convenciones usados para la transcripción ortográfica (3.3.1) y la anotación del corpus FerroviELE (3.3.2).

3.3.1. La transcripción ortográfica

Para la transcripción ortográfica de los diálogos, se usan los signos ortográficos convencionales. La transcripción discursiva se estructura en párrafos o frases que se corresponden con las intervenciones de los interlocutores (intercalando los turnos de palabra) para reflejar la estructura dialogada. La segmentación del discurso oral se rige por criterios sintáctico-semánticos, es decir, se segmenta en oraciones o sintagmas con sentido completo y los signos de puntuación convencionales (interrogantes, exclamaciones, etc.) se usan para diferenciar los rasgos prosódicos de entonación (una estructura exclamativa, interrogativa, etc.).

Se realiza una transcripción literal donde las palabras se transcriben tal como se producen oralmente, y a veces no tienen por qué seguir las formas prescriptivas, como por ejemplo, palabras mal pronunciadas o incorrectas, palabras que no son propias del español (mayoritariamente en inglés), palabras inacabadas o vacilaciones, autocorrecciones (1), especialmente en los corpus de aprendices.

(1)

- Palabras mal pronunciadas o incorrectas: **prefinta** *cama* en vez de *prefiero cama*;
- Palabras que no son propias del español: **or**, **okey**, **so**;
- Palabras inacabadas: *hay **posi** hay la posibilidad*;
- Autocorrecciones: *¿Y cuál y qué es y qué son **cuáles son**?*

En el caso de la información relacionada con números (horarios, precios, fechas), se han transcrito en letras. Los nombres propios (antropónimos, topónimos, nombres de empresa, organizaciones e instituciones) se escriben con la inicial en mayúscula. Las siglas y los acrónimos se transcriben en mayúsculas sin puntos intercalados.

Los alargamientos vocálicos que indicaban vacilación y tenían una duración prolongada (aproximadamente más de un segundo) se han marcado con puntos suspensivos, como en el ejemplo (2).

(2)

O8⁷⁸: *Eh y... después de Linares a Granada también, y sólo cuesta eh sesenta y un euro, como lo más barato* [FerroviELE-L_trained-multilingual_4]

En el corpus *FerroviELE* no se recogen los rasgos prosódicos del tipo inflexiones entonativas ascendentes o descendentes, aspiraciones y duración de las pausas⁷⁹. Estas últimas no se transcriben si bien sí se marcan con la etiqueta <pause> para posteriores tratamientos (véase sección C de este mismo apartado: Etiquetas para fenómenos suprasegmentales, paralingüísticos y pragmático).

Las interjecciones y las onomatopeyas se transcriben siguiendo la representación ortográfica que aparece en los diccionarios. Las onomatopeyas transcritas (*eh, um, mm, ah*, etc.) se corresponden principalmente con los rellenos de pausa (que también se han etiquetado como <filler hesitation pause>).

En los corpus de los aprendices, las onomatopeyas, cuando se producen en su lengua materna, son transcritas según la representación ortográfica de su lengua, aunque el diálogo lo estén produciendo en español, especialmente en el caso de los aprendices de nivel B1. En otras ocasiones, cuando el estudiante que actúa como interlocutor tiene un nivel B2, las pausas de relleno pertenecen ya al español y se transcriben como tales, como por ejemplo (3).

(3)

- estadounidenses /anglogermánicos: *um, uh, em*;
- nativos español: *eh, ajá, ajá, mm mm, uhum, hum-hum*;
- indistintamente: *ah*.

Asimismo, se observa que en el corpus *InfoTren-N* (nativos espontáneos) las pausas de relleno son inexistentes, excepto los *hum-hum* de confirmación.

⁷⁸ La etiqueta O8 indica el operador no nativo europeo anglogermánico de nivel B2 y *FerroviELE-L_trained-multilingual_4* indica el corpus del cual se extrae el ejemplo y el número del diálogo.

⁷⁹ Véase el artículo de Enríquez, Díaz y Taulé (2015) donde se analiza el empleo de las pausas y de las vacilaciones en los diferentes subcorpus de *FerroviELE*, cada uno de ellos integrado por seis diálogos.

En cuanto a los aspectos éticos de la grabación, en los diálogos en los que aparece el nombre y, en ocasiones, el apellido del informante, estos se anonimizan sustituyéndolos por las palabras NOMBRE y APELLIDO, escritos en mayúscula. En el caso en que se deletee el nombre o el apellido, también se cambian las letras para que no se identifique al informante. Por este motivo, solo están disponibles *online* las versiones transcritas y anotadas, pero no el archivo de voz correspondiente, precisamente para garantizar el anonimato.

3.3.2. La anotación del corpus *FerrovieLE*

Una vez realizada la transcripción ortográfica se ha procedido a la anotación morfosintáctica (sección 3.3.2.1) y a la anotación pragmático-funcional (sección 3.3.2.4). También, se ha codificado la información sobre las características de los diálogos, sobre el contexto comunicativo (sección 3.3.2.2) y sobre los interlocutores (sección 3.3.2.3). A continuación, la figura 2 presenta de forma esquemática el proceso de anotación realizado.

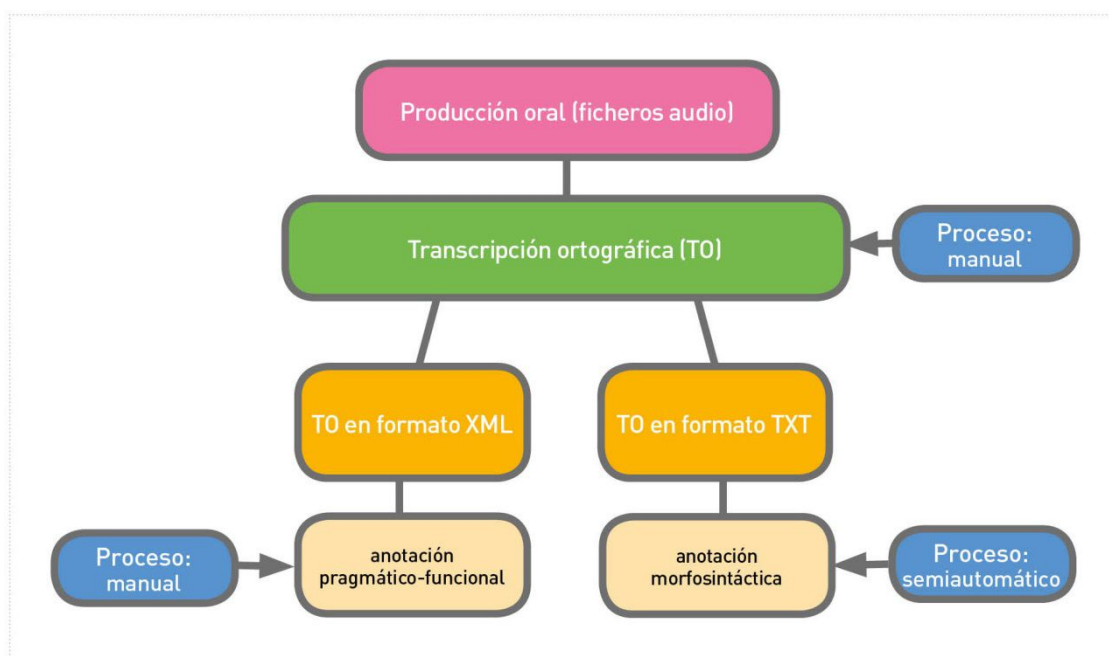


FIGURA 2. Desarrollo del corpus y procesos de anotación

3.3.2.1. Codificación morfosintáctica y lematización

La transcripción ortográfica (TO) en formato *.txt* del corpus *FerroviELE* se usa para llevar a cabo la anotación morfosintáctica. Para ello se ha utilizado el analizador morfológico de *Freeling*⁸⁰ para el español. Este analizador asigna de forma automática a cada unidad léxica (palabra o multipalabra) el lema y la categoría morfosintáctica correspondiente (basado en la propuesta del grupo EAGLES⁸¹). Las etiquetas morfosintácticas constan básicamente de ocho dígitos: los cuatro o seis primeros categorizan morfosintácticamente la palabra, mientras que se ha reservado el último dígito para categorizar errores, vacilaciones, repeticiones, autocorrecciones, etc. Las etiquetas morfosintácticas además de especificar la categoría léxica (N, V, A, R...) también incluyen información sobre el género y el número en el caso de los nombres y los adjetivos, el tipo de nombre (común o propio) e información sobre el tiempo, modo persona y número en el caso de los verbos. También se incluyen categorías formadas por un solo dígito para casos especiales, como son: I, interjección (*muchas_gracias*); W, fechas y horas (*quince_de_febrero* y *dos_de_la_tarde*) y Z, numerales (*nueve_y_ocho*).

En los ejemplos (4) y (5) se ofrecen casos ilustrativos de la codificación de las distintas categorías extraídos del corpus.

(4)

– *podría VMIC3S0*: el primer dígito indica la categoría léxica (verbo, V); el segundo el tipo de verbo (principal, M⁸² o auxiliar, A); el tercero y el cuarto el modo, en este caso indicativo (I) y el tiempo, condicional (C); el quinto la persona (3, tercera); el sexto el número (S de singular o P para el plural) y, por último, el séptimo el género (en este caso 0, no se especifica).

(5)

– *servicios del mi portátil*, la preposición *del* en vez de *para* SPCMSE: el primer dígito indica la categoría léxica (S, adposición); el segundo el tipo (P, preposición); el tercero la forma (S, simple o C, contraída); el cuarto el

⁸⁰ Padró, L. (s.f.): <<http://nlp.lsi.upc.edu/freeling/demo/demo.php>>

⁸¹ Etiquetas EAGLES para el español: en <<http://nlp.lsi.upc.edu/freeling/doc/tagsets/tagset-es.html>>; 09/15]

⁸² M corresponde a *main* del inglés.

género, siempre masculino, M; por tratarse de la preposición “del” el quinto el número, siempre singular (S) y, por último, el sexto el dígito propio E que se ha incorporado en la anotación de este corpus para indicar que hay un error.

Para garantizar la calidad del análisis morfosintáctico, se valida de forma manual la información obtenida automáticamente por un experto del grupo CLiC (*Centre de Llenguatge i Computació* de la Universitat de Barcelona). En este proceso de revisión no se corrigieron las formas con errores producidas por los alumnos, pero sí se les asignó el lema correcto y en la etiqueta morfosintáctica se marcó éste con el dígito/código E correspondiente para poder recuperar las ocurrencias erróneas posteriormente. Hay que señalar que en este corpus no se ha realizado una clasificación de los errores encontrados, simplemente se ha marcado a nivel de palabra si hay un error. De esta manera, se pueden obtener en las búsquedas bien las palabras con errores (a través de las palabras con la etiqueta E), bien a partir de los lemas, todas las formas, correctas o incorrectas asociadas a cada lema. Por lo tanto, se dispone de dos vías de acceso independientes que dan información complementaria. Nótese que los errores de los que nos ocupamos son los que se producen a nivel de palabra y están limitados al alcance de la transcripción ortográfica.

Además de la etiqueta E, se han marcado también, con sus códigos correspondientes, las vacilaciones, autocorrecciones, repeticiones y los cambios de interlocutor (O1, C1, TEACHER) a lo largo del diálogo. A continuación se listan las etiquetas con un ejemplo ilustrativo:

- E = error; palabra incorrectamente utilizada: **prefinta** (< *prefiere*)
- V= vacilación: **enla enlace**. Se marca la primera palabra
- A= autocorrección: **dola euros**. Se marca la primera palabra
- R= repetición: **cada cada dirección**. Se marca la segunda palabra
- Z= los interlocutores (O3, O5, O6, O10, C3, C5, C6, C10, TEACHER)
- C= clíticos: **beberlo**
- D= interferencia lingüística: **adéu** (< *adiós*) catalán-castellano, pero también con otras lenguas

El resultado del análisis morfosintáctico final es un fichero *.txt* con la información organizada en tres columnas. En la primera columna aparece la unidad léxica original

(palabra o multipalabra); en la segunda, el lema; y en la tercera, la categoría morfosintáctica. El analizador trata de manera automática como multipalabras las fechas (*el_ocho_de_febrero*), los horarios (*las_ocho_y_cinco_de_la_mañana*), los precios (*ciento_setenta_y_seis_euros*), porcentajes (*diez_por_ciento*) y algunas fórmulas o interjecciones (*buenas_tardes, hasta_luego, de_nada*). En el caso de que una multipalabra incluya una vacilación (*mm, eh, um...*), la multipalabra se separa (*diez por ciento vs. diez mm por_ciento*).

En el caso de los pronombres clíticos, se separan manualmente para identificar los errores de referencia (deixis y anáfora): *informarme > informar/me*. Se le añade una C a la etiqueta morfosintáctica del pronombre para poder recuperarlos posteriormente.

Las interjecciones y vacilaciones (*ehm, hum, emm, eh, ah, um...*) se tratan como una sola categoría: interjecciones.

Finalmente, como por razones éticas se anonimizan los participantes, aparece la etiqueta de nombre propio atribuida a personas y, también, a lugar (*Granada*) u organización o marca (*Renfe, AVE, Intercity*) (ejemplo 6).

(6)

```
NOMBRE_APELLIDO NP00000
Granada NP00000
AVE NP00000
```

El criterio general aplicado, como ya se ha dicho, es que no se modifica el texto original. Sin embargo, se identifican las palabras con errores asignando el lema correcto e indicando que hay un error con la etiqueta E en la categoría. A continuación, se presentan los criterios de lematización que se han seguido.

Palabras con error y soluciones adoptadas:

1. Errores debidos a que la palabra no existe, pero se deduce de qué palabra se trata.

Se interpreta y se le asigna el lema correcto del castellano:

*prefieres un **sentado** > prefieres un asiento* (palabra: sentado, lema: asiento).

2. Palabras en otro idioma. Se le asigna el lema correcto en castellano:

***ticket** > billete* (palabra: ticket, lema: billete)

3. Errores que reflejan una producción incorrecta (por su forma o su sentido). Se asigna el lema correcto en castellano:

– **parabra** > *palabra* (palabra: parabra, lema: palabra)

– **con base** > *con origen* (en la frase: *tren con base Barcelona*, para referirse a *tren cuya estación de salida es Barcelona*. Base se lematiza como *origen*) (palabra: base, lema: origen)

De los datos obtenidos de la lematización del corpus es posible extraer información especialmente útil para poder identificar qué tipo de léxico utilizan los aprendices según el nivel y combinado con el tipo de error que realizan.

La anotación morfosintáctica, además, permite obtener el número total de palabras (*tokens*), de palabras distintas (*types*) y de lemas de todo el corpus y de cada subcorpus. Así pues, *Ferroviele* está formado por 13,610 *tokens* que corresponden a 2,972 *types* y 1,849 lemas distintos. En la tabla 4 se presenta de forma detallada los *tokens*, *types* y lemas para cada subcorpus, según si la muestra incluye 6 diálogos (sub6) o 16 diálogos (sub16):

Corpus sub6/sub16	tokens	types	lemas
<i>InfoTren-N_spontaneous</i> (sub6)	1,335	336	276
<i>Ferroviele-N_simulated</i> (sub6)	1,093	336	278
<i>Ferroviele-L_no trained</i> (sub6)	1,024	226	172
<i>Ferroviele-L_trained</i> (sub6)	1,214	314	248
<i>Ferroviele-L_trained multilingual</i> (sub6)	2,720	536	370
TOTAL¹	7,386	1,748	1,344
<i>InfoTren-N_spontaneous</i> (sub16)	4,470	1,018	615
<i>Ferroviele-L_trained multilingual</i> (sub16)	5,809	1,078	536
TOTAL²	13,610	2,972	1,849

TABLA 4. Recuento de *tokens*, *types* y lemas

¹ Este total corresponde a la suma de los cinco subcorpus sub6.

² Este total es la suma de los dos subcorpus sub16 y de tres de los subcorpus sub 6, *Ferroviele-N_simulated*, *Ferroviele-L_no trained* y *Ferroviele-L_trained*.

3.3.2.2. Información sobre los diálogos

Cada diálogo tiene asociadas tres etiquetas⁸³ que sirven para identificar el tipo de diálogo realizado (<mode>), la calidad de la grabación (<fidelity>), y el contexto de grabación (<channel>).

El atributo <mode> tiene dos valores posibles <mode: planned/spontaneous>, planificado o espontáneo, que indica si se trata de un diálogo planificado o espontáneo.

El atributo <fidelity> tiene tres valores: alto, medio o bajo (<fidelity: high/medium/low>), para indicar la calidad de la grabación. Los cinco subcorpus poseen una calidad alta.

El atributo <channel> tiene dos valores: estudio o teléfono (<channel: studio/telephone>) para indicar si la grabación se ha realizado a través del teléfono o en estudio. En el caso del corpus *InfoTren-N*, el modo fue espontáneo y grabado por vía telefónica, mientras que en los corpus de aprendices *FeroviELE-L* y de nativos de referencia, *FeroviELE-N*, fueron planificados y grabados en estudio.

3.3.2.3. Identificación de los interlocutores

En el proceso de transcripción se han identificado las intervenciones de cada interlocutor según los roles adoptados: el cliente con la etiqueta <C> y el operador con la etiqueta <O>. Además, estas etiquetas se han acompañado de un número (O1-C1, O2-C2, O3-C3, etc.) para identificar tipos de interlocutores. Cada tipo de interlocutor <O> o <C> puede ser: hombre, mujer o desconocido; nativo/no nativo; en el caso de no nativo, se indica si es americano, asiático, inglés, francés o alemán (<American, Asian, English, French, German>) y el nivel de español como LE (<A2, B1, B2>).

De modo que cada uno de estos interlocutores tiene asignada una etiqueta que especifica su género y lengua nativa.

La etiqueta de género (<type>) tiene tres valores posibles: masculino, femenino o desconocido (<type:masculine/femenine/unknown>). Respecto a la lengua del interlocutor la etiqueta es <dialect> y los valores posibles son solo nativo y no nativo

⁸³ Estas etiquetas son las que incorpora la herramienta *Transcriber*.

(<dialect:native/nonnative>). A los interlocutores de lengua no nativa, además, se asocia la etiqueta de nivel de lengua según el *MCER* (Council of Europe, 2001), (<A1, A2, B1, B2 Level>) y la lengua materna (<American, Asian, English, French, German>), distinguiendo el inglés americano del británico (<American> <English>). En cambio, a los interlocutores de lengua nativa no se les vincula la etiqueta de nivel de lengua, aunque si se siguiesen los parámetros del *MCER*, a los hablantes nativos les correspondería un nivel C2. Tampoco se especifica su lengua materna. Así pues, la etiqueta <native> debe leerse como hablantes nativos de español peninsular y según los criterios del *MCER* equivaldría a un nivel C2.

Así, podemos encontrar, por ejemplo, las combinaciones: <American_B1 Level>, <Asian_B2 Level>, <English_B2 Level>, <German_B2 Level>, <French_B2 Level>.

A continuación, en la tabla 5 se presenta resumida la información sobre las características de los cinco subcorpus que configuran *FerrovieLE*. En la primera columna se recoge el tipo de corpus; en la segunda el tipo de interlocutor (en total 27 participantes distintos); en la tercera el género de los informantes; en la cuarta, si son nativos o no y el nivel de lengua y, en la quinta, si el diálogo es espontáneo (spontaneous) o simulado (planned).

CORPUS	INTERLOCUTOR	GÉNERO	LENGUA	MODO
InfoTren-N(atives)_ spontaneous	O1 C1	unknown	native	spontaneous
FerrovieLE-N(atives)_ simulated	O2 O4 C2 C4	male female male female	native	Planned
FerrovieLE-L(earners)_ no trained	O3 O5 C3 C5 TEACHER	male female male female female	nonnative: American_B1 Level native	Planned spontaneous

FerrovieELE- L(earners)_ Trained	O6	male	nonnative: American_B1 Level	Planned
	O10	female		
	C6	male		
	C10	female		
FerrovieELE- L(earners)_ trained multilingual	O7	male	nonnative: Asian_B2 Level	planned
	O11	female		
	C7	male		
	C11	female		
	O8	male	nonnative: English_B2 Level nonnative: German_B2 Level	planned
	O12	female		
	C8	male		
	C12	female		
	O9	male	nonnative: French_B2 Level	planned
	O13	female		
	C9	male		
	C13	female		

TABLA 5. Características de los diálogos e interlocutores⁸⁴

3.3.2.4. Codificación pragmático-funcional

En las páginas que siguen, se presenta el conjunto de etiquetas (*tag set*) que se han utilizado para anotar pragmática y funcionalmente el corpus *FerrovieELE*, en concreto las funciones comunicativas y las nociones léxico-semánticas, adecuadas al escenario comunicativo⁸⁵ y propias de esta tarea de la transacción de información en un contexto de compra-venta de billetes de tren.

Se entiende por función comunicativa o funciones comunicativas los elementos lingüísticos que necesita cualquier hablante que “*wishes to use a language (in our case English) to transact the business of daily living and enter into relations with the other speakers of that language in the situations which bring them together.*” (van Ek y Trim 2001, chapter 5: 27).

Se entiende por noción la categoría léxico-semántica asociada a cada situación comunicativa. Las nociones pueden ser generales y específicas. Las nociones

⁸⁴ Se han incluido los valores en inglés, porque el corpus *FerrovieELE* se ha codificado en dicha lengua.

⁸⁵ Recordemos que en la sección 1.2., Contexto de la investigación, se han presentado los dos tipos de análisis (de la forma a la función y de la función a la forma) que se aplicarán en el capítulo 4 y 5, a partir de los datos del corpus *FerrovieELE*.

generales hacen referencia a ámbitos como existencia, espacio, tiempo, cantidad, calidad, nociones mentales, relacionales y deícticas y se definen como:

“[...] a list of concepts or semantic categories derived from a consideration of what, in general, people deal with by means of language, of what concepts they may be like to refer to whatever the specific features of a particular communication situation may be.” (van Ek y Trim 2001, chapter 6: 60).

Las nociones específicas, en cambio, se definen como:

“By definition, the specific notions are transaction-related and/or topic-related, the ones particularly needed by the learners can only be determined once the transactions and topics concerned have been listed, together with indications as to what the learners are required to do with regard to each of them.” (van Ek y Trim 2001, chapter 7: 72).

En este grupo de las nociones se incluye léxico directamente relacionado con los detalles resultantes de la concreción, aquí y ahora, de la situación comunicativa elegida y la secuencia de funciones que intervienen en ella. Así, en la transacción de la (macro) función *comprar un billete de tren*, nociones generales serían: *horarios, precio, cantidad, recomendación, pago*; mientras que serían nociones específicas (léxico específico de la situación): *hora de llegada, billete de ida, billete de ida y vuelta, doscientos euros, un billete/un bono, visa, tarjeta, metálico, abono*, entre otras.

Los repertorios de referencia para niveles de aprendizaje (como el *MCER*) no incluyen las nociones específicas debido, precisamente, a la falta de concreción que, por definición, tienen como obra de referencia general.

La anotación del corpus *FerrovíELE* usa una selección de etiquetas procedentes de diversos marcos sociopragmáticos de análisis de la lengua. Todas las funciones de esta investigación están basadas en el *MCER*, en los descriptores del nivel *Vantage Level* (van Ek y Trim 2001: 27-59) para el inglés como segunda lengua, en las propuestas de *EAQUALS* del British Council para el inglés (North et al. 2010), en el *Plan Curricular del Instituto Cervantes* (2006) y la *Guía de contenidos lingüísticos por niveles del español* (Díaz, Martínez y Redó 2011) para el español como lengua extranjera.

3.3.2.4.1. Definición de las etiquetas

La definición de las etiquetas funcionales y nocionales se ha realizado en dos etapas:

- 1) En una primera etapa, se ha analizado una muestra de 12 diálogos —6 diálogos nativos espontáneos (*InfoTren-N*) y 6 diálogos de no nativos y no entrenados (*FerrovieELE-L_no trained*)— con el objetivo de obtener las etiquetas que mejor describen la situación de transacción. Los 12 diálogos seleccionados muestran las diferentes maneras de gestionar la información por parte de hablantes nativos y de hablantes no nativos. El resultado es una primera propuesta de etiquetas con las que se empieza a etiquetar el corpus.
- 2) En una segunda etapa, se redefinen etiquetas, cuando es necesario, y se amplían las etiquetas funcionales y nocionales como resultado de analizar el resto de diálogos del corpus, especialmente en el caso de los subcorpus en los que los informantes fueron entrenados (*FerrovieELE-N_simulated*; *FerrovieELE-L_trained*; *FerrovieELE-L_trained multilingual*). Además, se establecen los criterios de anotación definitivos, dando como resultado la información incluida en la guía de anotación⁸⁶, que se describe en la siguiente sección.

A continuación, se detallan el conjunto de etiquetas funcionales seleccionadas (A) y las etiquetas nocionales (B). También, se describen las etiquetas para fenómenos suprasegmentales, paralingüísticos y pragmáticos-discursivos (C).

A. ETIQUETAS DE LAS FUNCIONES COMUNICATIVAS

En la figura 3 se muestra el repertorio de funciones comunicativas utilizadas en la anotación del corpus. Estas funciones se inspiran en el *MCER* y se basan en los descriptores del *Vantage Level* y el *Core Inventory* de EAQUALS. El etiquetario consiste en 41 funciones distintas que constituyen una lista cerrada.

⁸⁶ Disponible en <<http://clic.ub.edu/es/publicaciones>>

– Advise	– Giving precise information / information / product printing ⁸⁷
– Agreeing	– Initiating conversation
– Apologising for making a mistake	– Interrupting
– Asking for confirmation or denial	– Inviting agreement
– Checking availability ⁸⁸	– Making a mistake
– Check-task	– Obligation and necessity
– Closing conversation	– Offer
– Confirming	– Paying
– Correcting	– Refusing
– Counting	– Repeating what one has said / other interlocutor has said
– Describing plans	– Request
– Encouraging and inviting another speaker to continue	– Seeking identification
– Enquiring about	– Socialising
– Expressing degrees of probability, doubt	– Socialising: asking for repetition
– Expressing feelings	– Socialising: asking someone to wait / Socialising: asking someone to wait: empathic offer
– Expressing knowledge	– Socialising: asking whether one is heard and understood
– Expressing opinion	– Socialising: supplying a word, repetition or expression
– Expressing preference	– Spelling
– Expressing reaction	– Suggestions
– Expressing reason, cause, excuse	– Synthetising information
– Filler hesitation pause	

FIGURA 3. Repertorio de etiquetas funcionales

⁸⁷ *Giving product printing*: Categoría híbrida entre función y noción específica, tomada del *Marco*, del *Vantage Level* y de EAQUALS. La función corresponde a *giving*, pero en la situación es relevante distinguir el tipo de producto y si éste implica una acción específica: *product printing*, resultado de la concreción del escenario.

⁸⁸ *Checking availability, check-task, counting, making a mistake, paying, socialising: asking someone to wait: emphatic offer* y *spelling* son las funciones que no aparecen recogidas en los repertorios usados como fuente (*Marco, Vantage Level* ni EAQUALS).

A continuación, se describen brevemente cada una de las funciones comunicativas utilizadas y se acompaña con un ejemplo de esta.

– **Accepting:** aceptar información anteriormente dada. Se realiza principalmente en el rol del cliente (C).

O5: *um para esta billete cuesta, oh este billete es um uh... cinco cientos dos euros*

C5: **Vale...**

[FerroviELE-L_no trained_1]

– **Advise:** dar consejo.

O1: ***pues no lo deje mucho más***

[InfoTren-N_spontaneous_12]

– **Agreeing:** expresar acuerdo.

C3: *Puah, porque no tengo tanto dinero*

O3: **Ah sí**

[FerroviELE-L_no trained_3]

– **Apologising for making a mistake:** pedir disculpas por equivocarse.

O1: *Dieciséis del [pron=pa]*⁸⁹

C1: **No, perdona, el quince por la noche**

[InfoTren-N_spontaneous_29]

– **Asking for confirmation or denial:** preguntar para confirmar o rechazar.

O12: *Y llega a las ocho y cinco de la mañana*

C12: *Muy bien, perfecto*

O12: **¿Sí?**

C12: *Perfecto*

[FerroviELE-L_trained-multilingual_8]

– **Checking availability:** comprobar la disponibilidad. Suele ir junto a la función <Socialising: asking someone to wait>.

O1: ***A ver que miro a ver [pronn=pa]***⁹⁰

[InfoTren-N_spontaneous_3]

⁸⁹ [pron=pa]: Etiqueta para identificar las pausas.

⁹⁰ *ibidem*

– **Check-task:** los interlocutores dejan su rol de operador (O) y cliente (C) para ser alumnos. Ocurre únicamente en un diálogo.

C9: *Gracias* [<Closing conversation>]

O11: *Gracias* [<Closing conversation>]

C9: **¡Está bien!**

O11: **¿Sí?**

[FerroviELE-L_trained-multilingual_12]

– **Closing conversation:** finalizar la conversación con diferentes fórmulas de cierre.

C5: **Gracias**

O5: **Hasta luego**

[FerroviELE-L_no trained_1]

– **Confirming:** confirmar información anteriormente dada. Se realiza únicamente en el rol de operador (O).

O5: *Um ¿necesitas esta uh billete para um sólo ida o para ida y vuelta?*

C5: *Ida y vuelta para mí, por favor*

O5: **Um bueno**

[FerroviELE-L_no trained_1]

– **Correcting:** corregir una información ofrecida.

O5: (...) *um hay un descuento del diez um per ciento*

C5: *Um...*

O5: **Por cierto**⁹¹

[FerroviELE-L_no trained_6]

En algunos casos, en que la información es más detallada, se especifica la noción de esta función.

O1: **Son dos horas, ¿eh?, a Castellón, que le había dicho dos horas y media y son dos horas. ¿Ida y vuelta sería?**

<Correcting: time>

[InfoTren-N_spontaneous_37]

– **Counting:** calcular la tarifa en voz alta.

O1: **Cuatrocientas. O sea que en principio este queda mejor. A ver que no me haya confundido en hacer alguna tasa. Son treinta y dos mil novecientas el principal, cuatro mil cien cada complementario. Quince y tres, cuatro. Cuarenta y cinco mil doscientas**

⁹¹ Se autocorrigió pero sigue siendo un error.

por dos son las noventa mil cuatrocientas, es correcto. O sea que en principio esto queda mejor [InfoTren-N_spontaneous_3]

– **Describing plans:** describir planes futuros.

C3: *Pues, voy a volver y comprar después*

[FerroviELE-L_no trained_3]

– **Encouraging and inviting another speaker to continue:** animar e invitar al otro interlocutor a continuar su intervención e indicar atención (función fática).

O11: *En na Línarés, Línarés, Sí. Este tren eh sale ah sale de Barcelona*

C7: *Mm*

O11: *De la mañana*

[FerroviELE-L_trained multilingual_14]

– **Enquiring about:** preguntar por algo con la finalidad de informarse.

C1: *¿Y alguna combinación para llegar antes de las doce a Murcia?*

[InfoTren-N_spontaneous_12]

– **Expressing degrees of probability, doubt:** expresar grados de probabilidad, duda.

O1: *¿Qué? ¿Va en este?*

C1: *Sí, puede ser [...]*

[InfoTren-N_spontaneous_38]

– **Expressing feelings:** expresar las emociones, actitudes o estados del interlocutor.

C4: *Pues mire, necesitaba viajar hoy mismo a Sevilla, y tengo algo de prisa (...)*

[FerroviELE-N_simulated_2]

– **Expressing knowledge:** expresar conocimiento, previo o compartido.

C1: *Sale, por lo que tengo aquí puesto, de Sants a las diez de la noche*

[FerroviELE-N_simulated_18]

– **Expressing opinion:** expresar una opinión.

C1: *Depende del precio, a ver*

[InfoTren-N_spontaneous_18]

– **Expressing preference:** expresar una preferencia.

C8: *Tengo que llegar antes de la diez*

[FerroviELE-L_trained multilingual_13]

Algunos turnos podrían interpretarse como <Request a service>, pero se descarta al haber en dicho turno más negociación y por la posición que ocupa en el diálogo es <Expressing preference>.

C5: **Um quiero comprar el billete más barato** [pron=unin]⁹²

<Expressing preference: fare>

[FerroviELE-L_no trained_4]

– **Expressing reaction**: expresar una reacción, ya sea de sorpresa, insatisfacción, etc.

O8: *¡Ah, vale! ¡Qué guay! ¿Viajas sola?* [FerroviELE-L_trained-multilingual_10]

– **Expressing reason, cause, excuse**: expresar una razón, causa o excusa.

C4: *Pues, va a ser por la mañana, porque tengo bastante trabajo allí*

[FerroviELE-N_simulated_1]

– **Filler hesitation pause**: relleno de pausas mientras se piensa o se busca una expresión o una palabra. Los dos “um” (pausas) del O5 no se etiquetan como <filler hesitation pause> por aparecer dentro de una producción, mientras que el “um” del ejemplo C5 se marca con dicha etiqueta porque es la única producción del turno.

O5: (...) *um hay un descuento del diez um por ciento*

C5: **Um...**

O5: *Por cierto* [FerroviELE-L_no trained_6]

– **Giving precise information / information / product printing**: dar información precisa de un ámbito concreto / dar información de ámbito personal / dar el producto (en este caso al ser un billete es un producto impreso).

<Giving precise information>

O1: **Vamos a ver, de momento quedan veintiséis plazas en primera**

[InfoTren-N_spontaneous_18]

<Giving information>

C5: **Soy una estudiante y no tengo mucho dinero**

[FerroviELE-L_no trained_4]

<Giving product printing>

O13: *Perfecto, aquí están sus billetes*

[FerroviELE-L_trained-multilingual_11]

– **Initiating conversation**: iniciar la conversación con diferentes recursos de apertura.

O5: **Hola**

⁹² [pron=unin]: Etiqueta para identificar palabra o palabras que no se pueden entender.

[FerroviELE-L_no trained_1]

– **Interrupting**: interrumpir para hacer una nueva pregunta.

O12: *Muchas gracias y aquí es su billete. ¿El siguiente? ¿Dónde tienes que ir? ¿Qué día es?*

C12: **Eh...**

O12: *Significa que eh aquí dice que tienes que ir a la vía cuatro*

[FerroviELE-L_trained-multilingual_8]

– **Inviting agreement**: confirmar o aceptar la información.

O1: *¿A Sevilla, **no**?*

[InfoTren-N_spontaneous_18]

– **Making a mistake**: equivocarse.

O1: *Sí, hay bastantes. Permita un momentito que le confirme el precio de Castellón, **a ver, que me parece que no se lo he dado bien***

[InfoTren-N_spontaneous_37]

– **Obligation and necessity**: expresar una obligación o necesidad.

O1: *Eh, no. **Tienes que pagar con la tarjeta de crédito***

[FerroviELE-L_trained-multilingual_1]

– **Offer**: ofrecer ayuda, alternativas, etc.

C10: *Bueno... aquí es mi tarjeta de crédito*

C10: *¿**Necesitas algo más?***

[FerroviELE-L_trained_1]

– **Paying**: pagar el producto.

C5: ***Aquí tienes***

[FerroviELE-L_no trained_2]

– **Refusing**: rechazar la información anterior, el ofrecimiento, etc.

C3: ***No, no, no. No quiero comprar***

[FerroviELE-L_no trained_3]

En algunos casos se especifica la noción de esta función.

O1: ***No, de momento no hay bonos para este, todavía***

<Refusing: transport season tickets>

[InfoTren-N_spontaneous_37]

– **Repeating what one has said / other interlocutor has said:** repetir lo que uno mismo o el otro interlocutor ha dicho en la intervención anterior (eco, reformulación, etc.). En algunas ocasiones, termina las frases del otro interlocutor como estrategia de colaboración.

<Repeating what one has said>

C1: *Ah, o este también me iría bien, vale y **si cogiera este de viajar de noche***

O1: *¿Perdón?*

C1: **Si cogiera el de viajar de noche**

[InfoTren-N_spontaneous_70]

<Repeating what other interlocutor has said>

O3: *Ah sí, sentado y...*

C3: *Porque no tengo*

C3: **No tengo bastante dinero**

[FerroviELE-L_no trained_3]

– **Request:** pedir un servicio u otra cosa.

C1: **Mire, le quiero hacer una pregunta. Si me podría decir, ¿si hay billete para el día cuatro por la tarde? ¿Por la noche?**

[InfoTren-N_spontaneous_18]

– **Seeking identification:** identificar una información previa mediante preguntas (aclaración).

O1: *Llevando el perro, sí. En Renfe, sí. En renfe, ¿eh?, luego si quieren mirar ya otra cosa, eso ya*

C1: **Otra cosa ¿qué sería?**

[InfoTren-N_spontaneous_68]

–**Socialising:** establecer relaciones sociales, expresar toma de contacto (saludos, despedidas), expresar empatía, etc. Suele aparecer en el inicio o en el cierre del diálogo.

INICIO

C3: **¿Qué tal?**

[FerroviELE-L_no trained_3]

CIERRE

O5: **Buen viaje**

[FerroviELE-L_no trained_2]

Los diálogos del subcorpus espontáneo, al ser grabaciones telefónicas reales, tienen funciones propias de este tipo de discurso. A continuación, se detallan las funciones específicas de <socialising> de este subcorpus. Estos subtipos de <socialising> no aparecen en los subcorpus de aprendices.

– **Socialising: asking someone to wait:** al teléfono, para pedir que esperen. Suele ir junto a la función *Checking availability*.

C1: *No, no, esta noche*

O1: *Esta noche, **espere**, le miro lo que, lo que hay en un principio y qué tal anda de plazas, **un momentico, por favor***

[InfoTren-N_spontaneous_81]

Asimismo, hay la variante **Socialising: asking someone to wait: empathic offer**. Tiene el mismo valor que la anterior, lo único que la diferencia es que en ésta aparecen pronombres como “te”, “a ti”, etc.

O8: ***Eh voy a hacerlo para ti** y sólo necesito em tu eh tu informaciones de la de tu tarjeta. Eh, por favor, dime tu titular*

[FerroviELE-L_trained-multiling_4]

– **Socialising: asking for repetition:** pedir que repitan una información.

C1: *¿Cada cuánto salen?*

O1: ***¿Eh?***

C1: *¿Cada cuánto sale este tren?*

[InfoTren-N_spontaneous_37]

– **Socialising: asking whether one is heard and understood:** verificar el canal.

O1: *Renfe, bon dia, **hola, Renfe, bon dia***

[InfoTren-N_spontaneous_95]

– **Socialising: supplying a word, repetition or expression:** facilitar al interlocutor una palabra, repetición o expresión para cooperar.

O11: *Y si no, si si viajas, si cambia otro tipo de... tren hay precio más barato, por ejemplo, mm Talgo pero con enlace, eh tienes que haz hacer u una parada y tienes que...*

C11: ***Cambiar***

[FerroviELE-L_trained-multilingual_16]

– **Spelling:** deletrear.

C8: ***Be como Barcelona y o, eme, e, a y eme***

[FerroviELE-L_trained-multilingual_4]

– **Suggestions:** proponer sugerencias.

O1: *Pues no lo sé, **alguna compañía, que le puedan llevar el animal, de transportes***

[InfoTren-N_spontaneous_68]

– **Synthetising information:** resumir la información ya dicha.

O1: ***Vamos a ver, usted quiere saber el precio***

InfoTren-N_spontaneous_29]

B. ETIQUETAS DE LAS NOCIONES ESPECÍFICAS O LÉXICAS

En la figura 4 se presenta el conjunto de etiquetas que constituyen las nociones léxicas para describir la situación comunicativa concreta de petición y suministro de información y venta de billetes de tren. Estas etiquetas son el resultado del refinamiento de las propuestas para el español del *Plan Curricular del Instituto Cervantes* (B1, B2). Constituyen un repertorio semicerrado de 37 etiquetas, susceptible de ser ampliado.

1. alternatives	15. making a reservation
2. arrival date	16. passengers
3. arrival time	17. payment conditions
4. availability of seats	18. personal
5. availability of tickets	19. platform
6. buying a ticket	20. preference
7. credit card	21. price
7.1. CVV number	22. product printing
7.2. expiration date	23. quantity of tickets
7.3. holder's name	24. seats
7.4. number	25. time
7.5. PIN	26. train services
8. date	27. train stops
9. departure place	28. train traffic
10. departure time	29. train traject
11. destination place	30. train type
12. fare	31. transport season ticket
13. ID document	32. travelling with pets
14. information	

FIGURA 4. *Repertorio de etiquetas nocionales*

A continuación, se describen brevemente cada una de las *nociones* léxicas utilizadas. Para cada etiqueta se marca en negrita la etiqueta de noción léxica.

1. alternatives: otras opciones.

C1: *Diecisiete y ¿entre medio tienes algo?*

<Enquiring about **alternatives**>

[InfoTren-N_spontaneous_3]

2. arrival date: fecha de llegada.

C1: *Hum, hum, vale. ¿Y para volver el domingo?*

<Enquiring about **arrival date**>

[InfoTren-N_spontaneous_36]

3. arrival time: hora de llegada.

O13: *Y ¿a qué hora le gustaría llegar a Granada?*

<Enquiring about **arrival time**>

[FerroviELE-L_trained-multilingual_11]

4. availability of seats: disponibilidad de plazas, asientos, etc.

O1: *Vamos a ver, de momento quedan veintiséis plazas en primera*

<Giving precise information: **availability of seats**>

[InfoTren-N_spontaneous_18]

5. availability of tickets: disponibilidad de billetes.

O1: *Vamos a ver lo que queda en este Talgo*

<Giving precise information: **availability of tickets**>

[InfoTren-N_spontaneous_18]

6. buying a ticket: pasos para comprar el billete.

C1: *Nada más podrás sacar el primer tramo, ¿eh?, porque el otro lo tienes que comprar allí. Entonces, en la estación de Sants, Paseo de Gracia o la Estación de Francia*

<Giving precise information: **buying a ticket**>

[InfoTren-N_spontaneous_12]

7. credit card: información sobre la tarjeta de crédito.

O8: *¿Me puede dar la tarjeta?*

<Enquiring about **credit card**>

[FerroviELE-L_trained-multilingual_10]

7.1. holder's name: nombre del titular de la tarjeta de crédito.

O13: (...) [*lex=iq-*]⁹³ entonces voy a tomar **la títula**, (...) [*-lex=iq*]
<Enquiring about credit card: **holder's name**>
[FerroviELE-L_trained-multilingual_11]

7.2. number: número de la tarjeta de crédito.

C10: **Cuatro, dos, cinco, seis, tres, seis, siete, uno, nueve, uno, cinco, tres, dos, siete, ocho, cuatro**
<Giving information: credit card **number**>
[FerroviELE-L_trained_2]

7.3. CVV number: número CVV (número de seguridad de la tarjeta de crédito).

C10: **Seis, seis, seis**
<Giving information: credit card **CVV number**>
[FerroviELE-L_trained_2]

7.4. expiration date: fecha de caducidad de la tarjeta de crédito.

C10: *Um hasta **nueve de... dos mil quince***
<Giving information: credit card **expiration date**>
[FerroviELE-L_trained_2]

7.5. PIN: número PIN.

O12: *Um ¿es posible poner tu **PIN**?*
<Enquiring about: credit card **PIN**>
[FerroviELE-L_trained-multilingual_13]

8. date: día del viaje (hoy, mañana, lunes, martes, miércoles, etc.), el día concreto de vuelta o el día concreto de salida, a diferencia de **time** que es general.

O1: *¿Me ha dicho **el día**?*
<Enquiring about **date**>
[InfoTren-N_spontaneous_18]

9. departure place: lugar de salida o punto de partida.

O1: *¿**De Barcelona**?*
<Enquiring about **departure place**>
[InfoTren-N_spontaneous_18]

⁹³ [*lex=iq*]: Etiqueta para identificar las preguntas indirectas.

10. departure time: hora de salida.

O1: *¿A las siete en punto?*

<Enquiring about **departure time**>

[InfoTren-N_spontaneous_19]

11. destination place: lugar de destino, mientras que la noción **train traject**, en este corpus, se refiere solo al origen-destino del tren.

C1: *Sí, para Sevilla*

<Expressing preference: **destination place**>

[InfoTren-N_spontaneous_18]

12. fare: tarifa o precios. Incluye especificaciones del tipo de billete (1) o de tren (2) y descuentos:

1. Tipo de billete:

- *de ida, de ida y vuelta, billete turista, business;*
- *juntos los pasajeros, en un pack de cuatro;*
- *primera clase, segunda clase, preferente, etc.,* conectado con las intervenciones anteriores en que la transacción era la tarifa (*fare*).

Dado que el concepto de tarifa suele incluir varios pares adyacentes en la interacción en torno al ámbito, y que estos presentan una secuencia flexible (precio, clase, tipo de tren, categoría, asientos, etc.).

O1: *Pero, ¿cómo quiere viajar?*

<Enquiring about **fare**>

C1: *Bueno, pues lo normal, lo menos posible; no me quiero gastar mucho dinero*

<Expressing preference: **fare**>

O1: *[lex=iq-] Primera clase [-lex=iq]*

<Enquiring about **fare**>

[InfoTren-N_spontaneous_18]

Por tanto, al hablar de *clase de tren*, de *categoría*, esa intervención se incluye dentro de *fare* (tarifa): tren de primera clase o de segunda clase; clase preferente, etc.

2. Tipo de tren:

C1: *Mire, quería saber **el precio del billete del Talgo del Antonio Machado, Barcelona-Sevilla***

<Request a service: information_fare>
[InfoTren-N_spontaneous_29]

- trenes relacionados con la tarifa, como por ejemplo: *Alaris, AVE, Euromed, Talgo*, etc. y
- servicios aceptados con la tarifa y sus especificaciones, por ejemplo, *en la misma cabina por ese precio*.

Si aparecen las palabras *precio*, *tarifa* y las expresiones *ir en cama/sentado* también se etiquetan como *fare*.

O11: *Sí, por supuesto, y **hay tarifas para cama individual, turista o preferente individual***

<Giving precise information: fare>
[FerroviELE-L_trained-multilingual_8]

13. ID document: documento personal de identificación.

O11: *A ver, ah entonces ah ¿puedo ver **tu... tu NIE o el pasaporte**, por favor?*

<Enquiring about ID document>
[FerroviELE-L_trained-multilingual_15]

14. making a reservation: reservar.

C1: *¿hoy lo **puedo reservar**?*

<Enquiring about making a reservation>
[InfoTren-N_spontaneous_2]

15. **information**: información del ámbito de estudio, es decir, sobre trenes. En principio, suele aparecer ligada a la petición de un servicio.

C5: *eh, podría **informarme eh del tren Barcelona-Sevilla***

<Request a service: information_traject>
[FerroviELE-L_no trained_1]

16. passengers: pasajeros.

O1: *¿Cuántas personas viajan?*

<Enquiring about **passengers**>

[InfoTren-N_spontaneous_29]

17. payment conditions: forma de pago, en metálico o con tarjeta de crédito.

O5: *Y, uh, ¿vas a pagar con tar tarjeta de empresa o con tarjeta personal?*

<Enquiring about **payment conditions**>

[FerroviELE-L_no trained_2]

18. personal: información personal.

C5: *Soy un estudiante y no tengo mucho dinero*

<Giving information: **personal**>

[FerroviELE-L_no trained_4]

19. platform: andén.

O12: *significa que eh aquí dice que tienes que ir a la vía cuatro*

<Giving precise information: **platform**>

[FerroviELE-L_trained-multilingual_8]

20. preference: preferencias en general.

O5: *¿Qué opciones quieres?*

<Enquiring about **preference**>

[FerroviELE-L_no trained_5]

21. price: importe. Suele estar al final del diálogo y, en ocasiones, junto a la función *paying* o a la noción **payment conditions**.

O5: *Um para esta billete cuesta, oh este billete es um uh... cinco cientos dos euros*

<Giving precise information: **price**>

[FerroviELE-L_no trained_1]

22. product printing: billete impreso.

O10: *No nada más. Espera un momento, por favor. ¡Ah, aquí es tu billete! Buen viaje*

<Giving **product printing**>

[FerroviELE-L_trained_1]

23. quantity of tickets: cantidad de billetes que comprar.

O6: *Uh pues, ¿cuántos tiquetes?*

<Enquiring about **quantity of tickets**>

[FerroviELE-L_trained_5]

24. seats: descripción de las plazas, ya sean sentadas o en litera.

O9: *Sí, sí, hay hay hay **cuatro camas en una cabina***

<Giving precise information: **seats**>

[FerroviELE-L_trained-multilingual_12]

25. time: horario, duración del viaje, partes del día (por la mañana/por la tarde) y tipo de tren (nocturno, diurno).

O5: *¿**Qué hora es el primer tren?***

<Enquiring about **time**>

[FerroviELE-L_no trained_1]

También se aplica <time> a circunstancias del ámbito personal —noción 17. *Personal*—. En el caso de la simulación en el subcorpus *FerroviELE-L_trained multilingual*, el guion les obligaba a hablar sobre la participación en un concierto (su horario específico, inicio, fin, etc.).

O10: *¿**Qué hora es el concierto?***

<Enquiring about **time**>

C10: ***A las veintidós***

<Giving information: **time**>

[FerroviELE-L_trained_1]

26. train services: servicios que ofrece el tren, como por ejemplo, los servicios de cafetería, restaurante, conexión wifi, etc.

C: *[lex=iq-]uh **tiene un restaurante y servicios del mi portátil y wifi** [-lex=-iq]*

<Enquiring about **train services**>

[FerroviELE-L_no trained_1]

27. train stops: paradas que realiza el tren.

C1: *Que pasa por Tarragona, Castellón, **¿qué paradas hace? ¿Hace en Tarragona, verdad?***

<Request a service: information_ **train stops**>

[InfoTren-N_spontaneous_37]

28. train traffic: circulación de trenes.

C1: *Vamos que hoy **sí que circula**, ¿no?*

<Enquiring about **train traffic**>

[InfoTren-N_spontaneous_81]

29. train trajet: trayecto del tren (origen-destino); si es directo o con enlace, ya sea otro tren o autobús; pueden aparecer palabras como *de regreso, dirección Murcia*; la ciudad de trasbordo se incluye en esta noción.

C5: *uh, ¿quiero quieres **directos** o no?*

<Enquiring about **train trajet**>

[FerroviELE-L_no trained_4]

30. train type: tipo de tren (*AVE, Alaris, Tren Hotel*, etc.) y si es rápido o lento.

O1: *¿Qué tipo de tren prefiere, Alaris o AVE?*

<Enquiring about **train type**>

[FerroviELE-L_simulated_4]

31. transport season tickets: abono de transporte.

C1: *Dos horas y media, jolín. Esto, es que, es que, es, es por mi trabajo y esto se ¿se pueden sacar **bonos**?*

<Enquiring about **transport season tickets**>

[InfoTren-N_spontaneous_37]

32. travelling with pets: viajar con animales.

C1: *Llevando un perro, ¿cómo se hace?*

<Request a service: information_ **pets**>

[InfoTren-N_spontaneous_68]

Así como las etiquetas funcionales no son combinables, las etiquetas nocionales sí pueden combinarse entre sí (7). Es decir, junto a una función puede incluirse una noción o varias nociones léxicas.

(7)

<Giving precise information>

O2: *Si prefiere viajar en tarifa preferente, **<amount.fare>** le informo que dispondrá de muchas más ventajas: servicio de cafetería, desayuno incluido, etc.*

<product.train.services> pero es obvio que el precio es diferente **<amount.fare>**

[FerroviELE-N_simulated_3]

En el ejemplo 7 se puede observar como dentro de la función <Giving precise information> aparecen dos nociones y no una única noción (<amount.fare> y <product.train.services>). Son las nociones combinadas en una única función comunicativa (en el ejemplo 7). Por

ello, hay un conjunto de 26 etiquetas de nociones léxicas que no siguen el mismo formato que las anteriormente presentadas por cuestiones técnicas de la herramienta *Transcriber*. Se detallan en la figura 5.

A. amount.fare	N. prod.buying.a.ticket
B. amount.price	O. prod.making.a.reservation
C. fac.cvv.number	P. prod.train.services
D. fac.credit.card	Q. prod.train.stops
E. fac.expiration.date	R. prod.train.traffic
F. fac.holder.name	S. prod.train.traject
G. fac.number	T. prod.train.type
H. fac.payment.conditions	U. prod.printing
I. loc.departure.place	V. prod.seats
J. loc.destination.place	W. time
K. loc.platform	X. time.arrival
L. pers.passengers	Y. time.date
M. pers.travelling.with.pets	Z. time.departure

FIGURA 5. Etiquetas nocionales especiales

Las nociones de la figura 5 aparecen combinadas, dos o más, en una misma función, como en el ejemplo (8); mientras que las nociones de la figura 4 se presentan sin combinarse: una única noción por función.

(8)

<Enquiring about>

C1: <prod.train.traject-> ¿Y alguna combinación <-prod.train.traject>
 <time.arrival-> para llegar antes de las doce <-time.arrival>
 <loc.destination.place-> a Murcia <-loc.destination.place>?
 [InfoTren-N_spontaneous_12]

C. ETIQUETAS PARA FENÓMENOS SUPRASEGMENTALES, PARALINGÜÍSTICOS Y PRAGMÁTICO-DISCURSIVOS

Con el objetivo de reproducir de la manera más fiel posible las producciones orales, la transcripción también incorpora elementos contextuales, como pausas, risas, ruidos, solapamientos, etc. en forma de etiquetas.

En esta sección se presentan las etiquetas utilizadas para identificar los fenómenos suprasegmentales y paralingüísticos más característicos, que se dan en este tipo de transacción.

Se han organizado en tres grandes grupos:

A) Ruidos *aislados* (*Isolated noise*), en el que se incluyen diferentes ruidos: llamada de teléfono, ruido indeterminado, risas, vocalizaciones o chasquidos que preceden a una locución. También se ha considerado como **ruido** en la conversación el cambio de idioma o la interferencia de la conversación privada entre los alumnos (o participantes), que normalmente aparece al principio, como etapa previa de ensayo, para ponerse de acuerdo o cuando se dan el pie para intervenir. Veamos algunos ejemplos:

– **Llamada de teléfono:** En algunos de los diálogos simulados, los participantes simulan el sonido de una llamada telefónica: <pc>, que se lee como *phone call*.

(no speaker): [pc-] Ring, ring, ring [-pc]
[FerroviELE-N_simulated_2]

– **Ruido indeterminado:** incluye portazos, ruidos ambientales, caída de objetos, etc. Se etiquetan con: <un>, que se lee como *undetermined noise*.

C12: Sí, [un] perfecto
[FerroviELE-L_trained-multilingual_8]

– **Risa:** las risas no se transcriben sino que se introduce directamente la etiqueta <l>, que se lee como *laugh*.

O12: Eh vale, cuesta... em mil cua mil, no [l] ciento cuarenta y seis euros y sólo se aceptan tarjetas
[FerroviELE-L_trained-multilingual_3]

– **Ruido de boca:** remite a ruidos involuntarios como chasquidos, fundamentalmente. Se etiquetan como <mn>, que se lee como *mouth noise*.

C10: **[mn]** *Tres para mí y dos amigas*
[FerroviELE-L_trained-USA_5]

– **Fragmento no transcrito:** conversación privada entre alumnos en su lengua materna antes de realizar el *role-play*. Se etiquetan como **<nontrans>**, que se lee como ***nontrans scripted segment***.

O10: *¡Hola!* **[nontrans]**
[FerroviELE-L_trained-USA_4]

B) *Pausas y vacilaciones:* Se incluyen las pausas silenciosas, los solapamientos entre dos interlocutores y los fragmentos incomprensibles (mayoritariamente debido a problemas de pronunciación, etc.). Las etiquetas asociadas a estos fenómenos son valores del atributo **[pron]**, correspondiente a ***pronunciación***.

– **Pausa:** bajo esta etiqueta se codifican los silencios más largos de lo esperado (mínimo 2 segundos) entre dos palabras o unidades léxicas y se han etiquetado como **<pa>**, que se lee como ***pause***.

C13: *Vale...* **[pron=pa]** *Aquí tienes*
[FerroviELE-L_trained-multilingual_7]

– **Solapamiento:** se da cuando dos interlocutores hablan al mismo tiempo. Se etiquetan como **<overlap>**. Al producirse un solapamiento, la parte final de la intervención de uno de los interlocutores se solapa con el principio de la siguiente intervención del otro interlocutor.

C1: *Quisiera preguntarle, ¿un tren que sale para dirección Murcia?*
[pron=overlap-]
O1: *Sí* **[-pron=overlap]**
[InfoTren-N_spontaneous_12]

En este ejemplo, **[pron=overlap-]** O1: *Sí* **[-pron=overlap]** indica que *sí* se solapa con *Murcia*. Es decir, aparece el solapamiento con la última palabra del interlocutor cliente y la primera palabra del otro interlocutor, el operador.

– **Palabra o fragmento incomprensible:** se da cuando no se puede entender la palabra o palabras que dice el interlocutor. Se etiqueta como **<unin>** y se lee como ***unintelligible***.

O8: *No, creo que* **[pron=unin]** *un plazo en coche cama*
[FerroviELE-L_trained-multilingual_4]

C) *Estrategias pragmático-discursivas:* Se refieren a actuaciones del interlocutor para resolver un problema comunicativo de tipo discursivo o de gestión o control de la información (repeticiones y, en ocasiones, la reformulación con preguntas indirectas) y los cambios de función para reconducir, asegurar,

ejecutar más eficazmente la tarea. Las etiquetas asociadas a estos fenómenos son valores del atributo **[lex]**, correspondiente a **léxico**.

– **Cambio de función en la misma intervención:** en algunas intervenciones puede haber alojadas dos o tres funciones comunicativas. Se separan con la subetiqueta **<change function>** que se lee como **same intervention_change function**.

O10: *Vale, [lex=change function-]*

O10: *[-lex=change function] y ¿tu número de ce be?*

[FerroviELE-L_trained-USA_2]

– **Pregunta indirecta:** se marca con la etiqueta **<iq>** y se lee como **indirect question**.

O1: *[lex=iq-] Sé que hay un Intercity a las diez y media de la mañana Barcelona-Madrid [-lex=iq]*

[InfoTren-N_spontaneous_2]

– **Repeticiones:** se distingue entre las repeticiones a corta distancia (casi efecto “eco”) de las que remiten a enunciados que están a mayor distancia en turnos de interacción. Se etiquetan como: **<check_short distance>** y **<check_long distance>** respectivamente.

O1: *Éste llega a [sale para] Alicante de las siete de la mañana; llegaría a las once cuarenta y cinco*

C1: *[lex=check_short distance-] Llegaría a las once y cuarenta y cinco [-lex=check_short distance]*

[InfoTren-N_spontaneous_19]

O1: *En el Talgo de las diez*

(...)

O1: *[lex=check_long distance-] ¿En el Talgo de las diez? [-lex=check_long distance]*

[InfoTren-N_spontaneous_18]

3.3.2.5. Proceso de anotación

La anotación del corpus *FerroviELE* se ha realizado por un único anotador y diálogo por diálogo. Para ello, se ha tenido como referencia la modelización de la transacción (tabla 8, capítulo 5).

En el proceso de anotación, antes de asignar una etiqueta funcional hay que tener en cuenta si se trata del O o del C, porque el rol y el orden de intervención puede

condicionar el tipo de etiqueta funcional, como se puede ver en la tabla 8. Se procede, por tanto, horizontalmente.

Una vez asignada la etiqueta funcional, se asocia la etiqueta nocional para llevar a cabo el contenido expresado en la función comunicativa según el rol, como se puede ver en las tablas 9, 10 y 11 (capítulo 5).

A partir de los resultados obtenidos en el análisis y la anotación del corpus *FerrovíELE* se pueden inferir elementos caracterizadores de los nativos frente a no nativos como veremos en los siguientes capítulos 4 y 5.

CAPÍTULO 4. ANÁLISIS DE DATOS: MORFOLOGÍA

En este capítulo se presenta el análisis cuantitativo y cualitativo de los datos obtenidos tras el análisis morfológico del corpus *FerrovíELE*. Nos acercamos a estos datos desde la perspectiva de análisis propuesta por Ellis-Baruiken (2005) (véase capítulo 1), de la forma a la función. Esto es, se estudian las formas gramaticales concretas y, a continuación, se pasa a verificar el significado específico que estas formas desarrollan en la función comunicativa. En el capítulo 5, se presentarán los datos del corpus *FerrovíELE* desde la perspectiva contraria, de la función a la forma.

Este análisis morfológico nos permite, por un lado, observar el tipo de léxico o vocabulario que utilizan los aprendices de español en esta muestra; y, por el otro, analizar las categorías morfosintácticas que usan y con qué frecuencia aparecen en ella. Los resultados nos permiten observar, también, i) si hay diferencias entre los distintos niveles de competencia lingüística de los aprendices, es decir, entre los no nativos de B1 y de B2; ii) estudiar el efecto de la simulación en hablantes nativos y no nativos frente al grupo formado por hablantes nativos en una situación no simulada, es decir, en un contexto real, de comunicación espontánea y, finalmente, iii) analizar las diferencias entre hablantes nativos y no nativos.

El capítulo se organiza en dos secciones. En la primera se presentan los resultados obtenidos en el análisis morfológico del corpus *FerrovíELE* (sección 4.1), compuesta por la descripción del etiquetario morfológico utilizado (4.1.1), el análisis de la muestra sub6 (4.1.2) y el análisis de la muestra sub16 (4.1.3). Junto a la descripción de los resultados, se destacan las categorías que permiten caracterizar el discurso nativo frente al no nativo en función de su peso cuantitativo en la producción.

En la segunda sección se realiza el análisis del vocabulario utilizado en este escenario comunicativo (4.2). Se presenta el léxico disponible en los hablantes nativos y no nativos y se recoge el léxico central o nuclear (*core*) del escenario comunicativo de estudio.

4.1. Análisis de los datos morfológicos

El análisis morfológico que presentamos a continuación se ha realizado a partir de dos muestras de datos distintas: i) en primer lugar, se analizará la muestra sub6, que consiste en una muestra homogénea en la que cada subcorpus de *FerrovieLE* contiene 6 diálogos (tabla 6) (*FerrovieLE-L_no trained-USA*, *FerrovieLE-L_trained-USA*, *FerrovieLE-N_simulated*)⁹⁴; ii) en segundo lugar, se analizará la muestra sub16 integrada por los subcorpus *FerrovieLE-L_trained-multilingual* y *FerrovieLE-N_spontaneous* que contienen 16 diálogos cada uno (tabla 7).

La primera muestra (sub6) nos permite realizar un análisis comparativo del uso de las diversas categorías en cada subcorpus, haciendo especial hincapié en las categorías que permiten diferenciar niveles de competencia lingüística (entre aprendices B1 y B2), y entre hablantes (nativos y no nativos), así como estudiar el efecto de la simulación en la tarea. La segunda muestra (sub16) nos permite realizar comparaciones entre las producciones realizadas por hablantes nativos expertos (en una situación no simulada) y las realizadas por hablantes no nativos multilingües (nivel B2), que no son expertos y que han sido entrenados (situación simulada).

En las tablas 6 y 7 se presentan los datos morfológicos obtenidos para cada una de las muestras analizadas (sub6 y sub16 respectivamente).

En la primera columna de cada tabla se indican las categorías morfosintácticas utilizadas (13 en total) y, a continuación, el recuento de ocurrencias o palabras etiquetadas bajo estas categorías (*tokens*)⁹⁵; el número de palabras diferentes (*types*); los lemas (la forma canónica o de citación en el diccionario) y la ratio entre *types* y *tokens* (TTR o *type/token ratio*). Por último, se muestra la media de repetición (Mrep) para cada uno de los subcorpus (es decir, la *token/type ratio*).

En las columnas siguientes se presentan los datos obtenidos para cada subcorpus: además del nombre del subcorpus y del número de diálogos recogidos, se indica la descripción cuantitativa para cada categoría, detallando en primer lugar el número de ocurrencias o *tokens*; en segundo lugar, el porcentaje de aparición de

⁹⁴ El criterio seguido para la selección de los diálogos es que tuvieran todos una longitud homogénea.

⁹⁵ Se incluye en palabras, también, a las multipalabras (fechas, horarios, interjecciones complejas y fórmulas). Véase el apartado: 3.3.2.1. Codificación morfosintáctica y lematización.

tokens respecto a su subcorpus, es decir, qué representa en el total de su grupo cada una de las categorías usadas y, en tercer lugar, el porcentaje de errores para cada categoría (respecto al total de errores en cada subcorpus). Esta información se da siempre para cada subcorpus. La presentación de los subcorpus se ha organizado siguiendo el criterio de competencia lingüística: primero, *Ferroviele-L_no trained-USA*; segundo, *Ferroviele-L_trained-USA*; tercero, *Ferroviele-L_trained-multilingual*; cuarto, *Ferroviele-N_simulated* y, por último, *Ferroviele-N_spontaneous*.

Por ejemplo, en la tabla 6, el subcorpus *Ferroviele-L_no trained-USA-sub6* contiene un total de 1024 *tokens* (el total de las ocurrencias del subcorpus) de los cuales 46 *tokens* son adjetivos, que representan el 4,49% de las ocurrencias de dicho subcorpus. La tercera descripción cuantitativa es el porcentaje de errores de cada categoría respecto del total de errores detectados en el subcorpus, que en el caso de la categoría de adjetivos, se detectan 2 errores que representan el 4,35% de adjetivos incorrectos.

En las cuatro últimas filas se incluye el número de *tokens*, *types*, lemas, la TTR y la Mrep para cada subcorpus. Por ejemplo, *Ferroviele-L_no trained-USA-sub6* contiene un total de 1024 *tokens* que corresponden a 226 *types* y 172 lemas distintos, con una TTR del 0,22 y una Mrep del 4,55.

La relación entre los *tokens* y los *types* (TTR), indica la riqueza o densidad léxica de los textos: a mayor número de palabras distintas mayor riqueza léxica, es decir, cuanto más cerca esté la ratio de 1, mayor diversidad léxica; cuanto más cerca de 0, habrá más repeticiones y menor variedad de palabras. Cabe destacar, que para el corpus *Ferroviele* los índices de riqueza léxica en cada subcorpus son proporcionales porque se tuvo en cuenta que los diálogos tuvieran la misma longitud. Por último, se indica la cifra de la media de repetición de cada palabra (Mrep), que se obtiene a partir de la división entre *tokens* y *types*.

	FerroviELE-L_ no trained-USA -sub6 (Nivel B1)			FerroviELE-L_ trained-USA -sub6 (Nivel B1)			FerroviELE-L_ trained-multilingual -sub6 (Nivel B2)			FerroviELE-N_ simulated -sub6 (Nativos)			InfoTren-N_ spontaneous -sub6 (Nativos)			TOTAL
	tokens	%text	%error	tokens	%text	%error	tokens	%text	%error	tokens	%text	%error	tokens	%text	%error	
Adj.	46	4,49	4,35	56	4,61	12,5	99	3,64	12,12	37	3,39	0	39	2,92	0	
Adv.	136	13,28	4,41	111	9,14	4,5	323	11,88	5,57	119	10,89	0	152	11,39	0	
Conj.	85	8,3	7,06	88	7,25	7,95	259	9,52	5,79	84	7,69	0	102	7,64	0	
Det.	95	9,28	13,68	141	11,61	7,8	280	10,29	13,93	119	10,89	0	170	12,73	0	
Interj.	41	4	0	36	2,97	5,56	59	2,17	0	36	3,29	0	55	4,12	3,64	
Int-pau.	108	10,55	0	31	2,55	0	169	6,21	0	22	2,01	0	13	0,97	0	
NC	172	16,8	9,3	222	18,29	4,05	436	16,03	10,78	194	17,75	0	150	11,24	0	
NP	36	3,52	5,56	53	4,37	5,66	96	3,53	17,71	36	3,29	0	75	5,62	0	
Pron.	23	2,25	8,7	47	3,87	2,13	99	3,64	22,22	66	6,04	0	86	6,44	0	
Prep.	105	10,25	12,38	149	12,27	8,05	291	10,7	13,75	141	12,9	0	203	15,21	0,49	
V	148	14,45	11,49	196	16,14	5,1	464	17,06	19,4	198	18,12	0	203	15,21	0	
F/H	8	0,78	62,5	19	1,57	31,58	44	1,62	25	19	1,74	0	59	4,42	0	
Cifra	21	2,05	9,52	65	5,35	6,15	101	3,71	21,78	15	1,37	0	48	3,60	0	
Tokens	1024			1214			2720			1093			1335			7386
Types	226			314			536			336			336			1748
Lemas	172			248			370			278			276			1344
TTR	0,22			0,26			0,20			0,31			0,24			
Mrep	4,55			3,87			5,07			3,25			4,03			

TABLA 6. *Corpus FerroviELE-sub6: datos morfológicos*

	FerroviELE-L_trained-multil -TOTAL:16 (Nivel B2)			InfoTren-N_spontaneous -TOTAL:16 (Nativos)			TOTAL
	Tokens	%text	%error	tokens	%text	%error	
Adj.	211	3,63	11,37	130	2,91	2,31	
Adv.	680	11,71	6,32	565	12,64	4,78	
Conj.	551	9,49	5,44	383	8,57	1,04	
Det.	579	9,97	13,3	474	10,6	1,27	
Interj.	202	3,48	5,45	195	4,36	4,62	
Int-pau.	322	5,54	0	66	1,48	0	
NC	911	15,68	9,22	548	12,26	0,36	
NP	191	3,29	12,04	187	4,18	0	
Pron.	221	3,8	17,19	329	7,36	2,18	
Prep.	594	10,23	11,45	566	12,66	1,41	
V	1013	17,44	12,14	733	16,4	0,68	
F/H	96	1,65	17,71	150	3,36	0,67	
Cifra	238	4,1	16,39	164	3,67	1,22	
Tokens	5809			4470			10279
Types	1078			1018			2096
Lemas	536			615			1151
TTR	0,19			0,23			
Mrep	5,39			4,39			

TABLA 7. *Subcorpus FerroviELE-L_trained-multi-sub16 e InfoTren-N_spontaneous-sub16: datos morfológicos*

4.1.1. Breve descripción del etiquetario morfológico utilizado

En relación con cada una de las 13 categorías morfosintácticas utilizadas⁹⁶, hay que indicar lo siguiente:

- a) **Adjetivo**. Incluye dos tipos de adjetivos:
 - calificativos (AQ) (*prefiero un tren directo*) y
 - ordinales (AO) (*puede tomar el primero de la mañana*).En *FerrovíELE* se da esta restricción sobre el repertorio de Parole, debida a las características comunicativas del corpus.
- b) **Adverbio**. Se distingue entre adverbios:
 - generales (RG) (*prefiero comprar solo el preferente sentado*) y
 - negativos (RN) (*non es*).
- c) **Conjunción**. Abarca dos clases:
 - las coordinantes (CC) (*un billete sencillo o ida y vuelta*) y
 - las subordinantes (CS) (*prefiero comprar solo el preferente sentado que incluye desayuno y la cena*).
- d) **Determinante**. Contempla cinco tipos:
 - artículo (DA) (*me da igual el tipo de tren*),
 - demostrativo (DD) (*las otras opciones em de este tren son reservar un cama*),
 - indefinido (DI) (*¿es un viaje de noche?*),
 - posesivo (DP) (*aquí están sus billetes*) e
 - interrogativo (DT) (*¿cuántos em camas hay en una...?*).
- e) **Interjección**. (I) Incluye onomatopéyicas, fáticas, lexías usadas con valor interjectivo, etc. (*¡qué guay!/buenos días*).
- f) **Int-pausa**. (P) Corresponde a interrupción o pausa y es una etiqueta para fenómenos suprasegmentales de creación propia para este corpus (*eh/em/ah/mm*).

⁹⁶ El conjunto de etiquetas se basa en la propuesta del grupo EAGLES (1996) para el español: <<http://nlp.lsi.upc.edu/freeling/doc/tagsets/tagset-es.html>>

g) **Nombre.** Los nombres pueden ser de dos tipos: nombre común y nombre propio. En este caso se ha contabilizado por separado las ocurrencias porque para el estudio nos pareció pertinente, dado que en la simulación se cerraban los destinos y el incremento de número de destinos en la no simulada no responde solo a la competencia, sino a la no restricción de la situación real:

- nombre común (NC) (*¿a qué hora le gustaría llegar a Granada?*) y
- nombre propio (NP); no se subdivide. Será una categoría interesante para ver tipos de tren (*AVE, Euromed, Intercity...*) y origen-destino que son lo específico de este corpus (*Granada, Barcelona, Sevilla, Málaga*). Se recuerda que no aparecen nombres de persona ni apellidos en el corpus (anonimización garantizada por razones éticas).

h) **Pronombre.** Se distinguen cinco tipos:

- demostrativo (PD) (*este me interesa*),
- indefinido (PI) (*¿Y en el otro?*),
- personal (PP) (*Y yo quiero llegar allí y... antes de posible*),
- relativo (PR) (*lo que te propongo...*) e,
- interrogativo (PT) (*¿adónde vas?*).

No se incluyen los numerales y los posesivos porque no se utilizan en este tipo de situación comunicativa⁹⁷.

i) **Preposición.** Se distinguen:

- las contractas (SPC) (*del/al*) y
- las no contractas o simples (SPS) (*a/de/en*).

j) **Verbo.** Se diferencian entre tres subgrupos:

- los verbos predicativos (VM) (*tendré/quiero/gustaría*),
- los copulativos (VS) (*es/hay/está*) y
- los auxiliares (VA) (*he.../has...*).

k) **Fecha y hora.** (W) No se subdivide (*ocho y media por la mañana; 3 de marzo*).

⁹⁷ No se han encontrado tampoco formas como *el mío, el tuyo*, etc., porque la compra se hace siempre individualmente y no es necesario este tipo de aclaración/atribución en la situación comunicativa.

- l) **Cifra.** Se subdivide en dos:
- la cifra en sí (Z) —para cantidad (somos cuatro)— y
 - la moneda (ZM) (doscientos sesenta euro).

Todos los ejemplos ofrecidos para estas categorías se han ilustrado con el subcorpus *FerrovieLE-A_trained-multilingual_B2*.

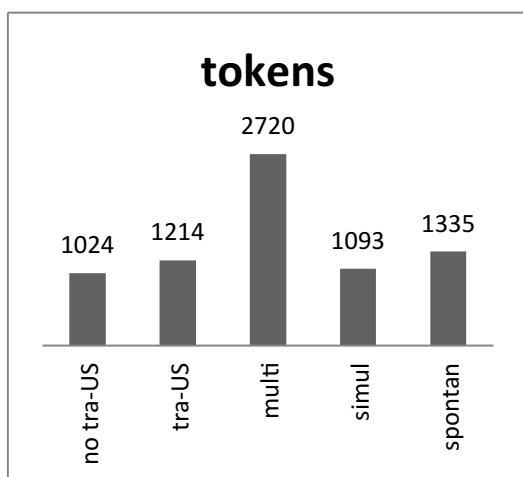
4.1.2. Análisis de la muestra sub6

En este apartado se realiza un análisis cuantitativo y cualitativo de los datos obtenidos en la muestra sub6 (tabla 6) con dos objetivos concretos:

- 1) El primero, comparar las producciones realizadas por los distintos grupos de aprendices (B1 y B2) y respecto a los nativos, para analizar separadamente el efecto del nivel de competencia, por un lado y el efecto del entrenamiento, por otro.
- 2) El segundo, comparar las producciones realizadas por los hablantes nativos en una situación simulada respecto a las producciones nativas en situación espontánea, para evaluar el efecto de simulación en la tarea.

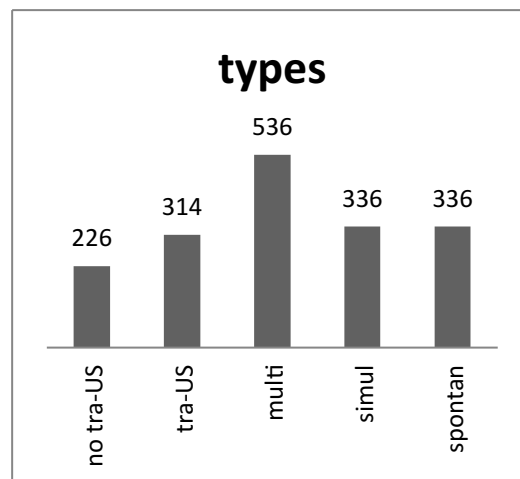
La muestra sub6 tiene un total de 7386 *tokens* que corresponden a 1748 *types* y a 1344 lemas distintos. Se trata, por tanto, de una muestra modesta pero suficiente para poder poner de manifiesto las diferencias entre los distintos grupos y condiciones (entrenado y no entrenado, simulado y espontáneo) que se pretenden estudiar. En las figuras 6 y 7⁹⁸ se recoge la distribución de *types* y *tokens* en la muestra sub6.

⁹⁸ En las figuras 6 y 7 así como en las figuras 8 y 9, los nombres de los diferentes subcorpus aparecen abreviados: *no tra-US* corresponde al subcorpus *FerrovieLE-L_no trained-USA*; *tra-US* al subcorpus *FerrovieLE-L_trained-USA*; *multi* al subcorpus *FerrovieLE-L_trained-multilingual*; *simul* al subcorpus *FerrovieLE-N_simulated* y *spontan* al subcorpus *FerrovieLE-N_spontaneous*.



Total: 7386 tokens

FIGURA 6. Número total de tokens en sub6



Total: 1748 types

FIGURA 7. Número total de types en sub6

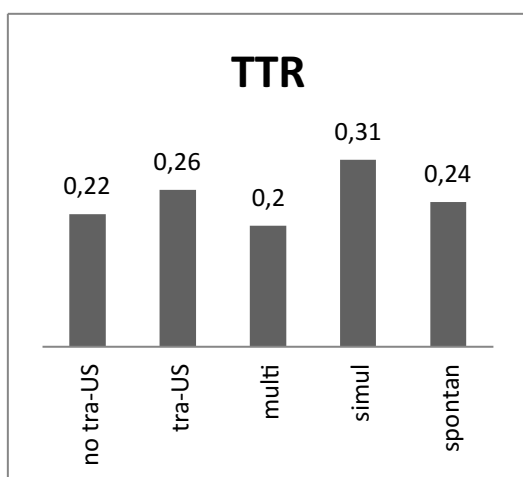


FIGURA 8. TTR en sub6

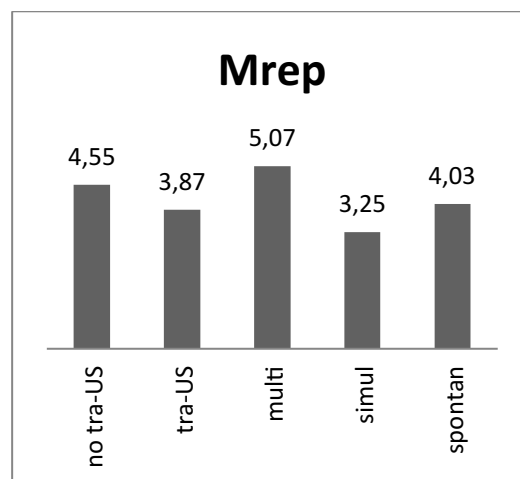


FIGURA 9. Mrep en sub6

En lo que se refiere al número total de palabras, pues, existe una diferencia visible entre los tres subcorpus de aprendices: el valor de *tokens* aumenta progresivamente (figura 6) ; el valor más bajo aparece en los alumnos no entrenados de nivel B1 (1024), el número de ocurrencias crece con los alumnos entrenados de nivel B1 (1214) y todavía más en B2 (2720). En cambio, en cuanto al número de *types* (figura 7), éste también incrementa pero no en la misma distribución. Como se ve en la TTR y la Mrep de las figuras 8 y 9, la TTR muestra que los de B1 entrenados tienen más riqueza léxica y menos repeticiones que los de B1 sin entrenar. Curiosamente, también que los B2 entrenados. La explicación para los B1 es que el efecto

entrenamiento produce la mejora dentro del nivel, en el ámbito de precisión y variedad léxica. En cambio, en B2, donde la producción es mayor, la repetición tiene que ver con la aparición de una estrategia de coherencia, cognitiva y comunicativamente más compleja, que no pueden llevar a cabo los aprendices de B1. Por lo tanto, la lectura cuantitativa tiene que matizarse: dentro del mismo nivel (B1) el salto cualitativo es explicable por el entrenamiento; mientras que en B2, se atribuye al uso de una estrategia comunicativa-discursiva de recapitulación de información o de creación de coherencia (a base de léxico), propia de este nivel. Estos aspectos se verán con más detalle en el capítulo 5 (apartado 5.5.3.3. La cooperación como indicador de cortesía, en particular, el bloque C.1. Las repeticiones).

Por otro lado, podemos observar los resultados de los hablantes nativos, donde hay una diferencia entre la cantidad de *tokens* de los nativos simulados (1093) y los espontáneos (1335). En cuanto a los *types*, coinciden exactamente (336). De manera que la diferencia reside en la TTR (0,31 nativos simulados y 0,24 nativos espontáneos) y la Mrep (3,25 nativos simulados y 4,03 nativos espontáneos). En la situación simulada los hablantes nativos presentan una densidad léxica mayor, debido al efecto simulación, a la preparación de la tarea, a la participación en la tarea y a la familiaridad con la situación. No obstante, cuantitativamente, los nativos espontáneos (los profesionales de RENFE) al encontrarse en una situación real (la suya laboral) se diferencian de los simulados porque utilizan más ocurrencias.

Es interesante destacar dentro de la comparación entre aprendices y nativos, que el número de *tokens* de los aprendices B1 (ambos grupos) se acercan más a los de los nativos que los de B2. Numéricamente: B1 (1024 y 1214, no entrenados y entrenados respectivamente), nativos (1093 y 1335, simulados y espontáneos respectivamente) y B2 (2720). Estos últimos (B2) casi triplican las ocurrencias de los otros grupos. Una posible interpretación es que sobreactúan porque su competencia y su conocimiento de la tarea les permiten priorizar la cantidad de producción (y la exhaustividad en las fases de la tarea), mientras que en los otros grupos, por razones distintas (en B1 por competencia; entre los nativos porque la evaluación de la lengua no está en juego) son más directos y operativos.

En relación a los *types* —o palabras diferentes que se utilizan— se observa que los subcorpus de aprendices siguen la misma distribución que los *tokens*. Así pues los aprendices de B1 sin entrenamiento presentan el valor más bajo (226) seguidos de los B1 entrenados (314) y de los B2 (536); frente a los 336 que usan los nativos (igual en ambos grupos). Sin embargo, si nos fijamos en la TTR y la Mrep, el grupo de B1 entrenado es, una vez más, el que se acerca más a los nativos (ver tabla 6).

A continuación, se comparán las categorías morfosintácticas presentes en cada subcorpus, destacando las categorías que caracterizan cada subgrupo. Para ello tenemos en cuenta, en este caso, los porcentajes y no el número de ocurrencias. No obstante, se puede recurrir a ellas en la tabla 6, si así se desea. Primero, se listan las categorías más frecuentes de cada subcorpus y, después, se comenta detalladamente cada una de ellas.

Las categorías caracterizadoras de cada subcorpus son las siguientes:

- a) la abundancia de las *pausas* caracteriza al grupo B1 sin entrenar y al grupo B2;
- b) la abundancia de *adverbios* caracteriza al grupo B1 sin entrenar;
- c) la abundancia de *adjetivos* caracteriza a los grupos B1;
- d) la abundancia de *conjunciones* caracteriza al grupo B2;
- e) la abundancia de *pronombres* y *preposiciones* caracteriza a los hablantes nativos frente a los aprendices, tanto a los B1 como a los B2;
- f) la abundancia de F/H (*fechas y horarios*) caracteriza a los hablantes nativos en una situación espontánea;
- g) los hablantes nativos en una situación simulada utilizan más *adjetivos, verbos, nombres comunes* y *pausas* que los hablantes nativos que en una situación espontánea;
- h) los hablantes nativos en una situación espontánea usan más *adverbios, nombres propios, preposiciones, interjecciones, fechas-horarios* y *cifras* que los simulados y
- i) el grupo B2 se caracteriza en que para todas las categorías excepto *conjunción* y *fecha-horarios* tiene un porcentaje de error mayor que los demás no nativos (con y sin entrenamiento).

En general, las diferencias se podrían explicar por las siguientes razones. En relación con (a), el uso de pausas en los aprendices de B1 (10,5%) sin entrenar es atribuible a la falta de competencia lingüística y de conocimiento específico del escenario comunicativo: falta de vocabulario, gestión de información, etc. Con el entrenamiento, sin embargo, el grupo B1 entrenado reduce muy significativamente la producción de pausas y vacilaciones (2,5%). En cuanto a los aprendices B2, la abundancia de pausas (6,2%) se debe probablemente a que asumen más riesgos: más producción, cuantitativamente; complejidad, cualitativamente. Recordamos que no disponemos del grupo B2 sin entrenar para ver el efecto del entrenamiento y matizar esta interpretación.

En relación con (b), la abundancia de adverbios en los aprendices de B1 sin entrenar (13,2%) puede atribuirse a la necesidad de precisar localización, tiempo, etc. léxicamente, lo que parece confirmar la más baja frecuencia de verbos (14,4%) respecto al resto de grupos (asociada, a su vez, a un alto número de errores en estos en la flexión (11,4%)).

En relación con (c), la presencia de adjetivos es mayor en ambos grupos de aprendices B1 (4,4%). Esto se debe a que valoran la situación, el servicio, que es algo bastante ajeno a lo esperado en la situación. Es, probablemente, una estrategia de compensación léxica que no va acompañada de un dominio efectivo de la adjetivación (puesto que tienen muchos errores, especialmente los entrenados (12,5%)).

En relación con (d), la abundancia de conjunciones caracteriza a B2 (9,5%). Es un grupo que asume más riesgos a nivel sintáctico (que no pueden asumir, en cambio, los aprendices de B1) y que tienen posiblemente interiorizado un modelo más escrito de lengua que uno oral. Esto hace que los dos grupos de nativos tengan una producción en número de conjunciones más baja (casi exacta entre sí (7,69% los simulados y 7,64% los espontáneos), y más cercana a los B1 entrenados (7,25%)).

En relación con (e), el mayor uso de pronombres y preposiciones caracteriza a los hablantes nativos frente a los aprendices. Por razones distintas en cada caso: los pronombres porque implican gestionar la información discursiva y usar las anáforas recuperando referentes y usando la correferencia con agilidad. Esto se escapa a los hablantes no nativos en general y podemos encontrar tanto estrategias de evitación

(en todos los grupos de aprendices, que producen equilibradamente poco), como indicios claros de dificultad, confirmable por el número de errores (especialmente alto en B2 (22,2%)). Esto hace de los pronombres un indicador de “natividad” inequívoco. Las preposiciones son puramente un aspecto sintáctico y son difíciles en el aprendizaje de todas las lenguas. La diferencia con los hablantes nativos (12,9% los simulados y 15,2% los espontáneos) no es tanto cuantitativa como, sobre todo, cualitativa (errores presentes en porcentaje alto y en todos los niveles; aunque B1 entrenado es sensible al entrenamiento y reduce el número de errores (8,05%)).

En relación con (f), la presencia de fechas y horarios más elevada en nativos y, especialmente en los espontáneos (4,4%), tiene que ver con el dominio de la situación (de los nativos en general frente a los no nativos) y con el carácter abierto de la situación real (que permite más repertorio de destinos, fechas, horarios, cifras, trenes, etc.) frente al cerrado por el propio guion de la tarea.

En relación con (g), las diferencias entre nativos simulados y espontáneos recogidas en este apartado responden al efecto simulación. Los nativos que simulan, preparan, planifican, dan variación deliberadamente a su producción y por tanto usan más nombres comunes distintos, adjetivos y verbos. En la situación real, en cambio, la inmediatez y la prioridad del objetivo comunicativo frente al cuidado por la forma hacen que estas categorías tengan menor presencia.

En relación con (h), los hablantes nativos espontáneos utilizan más adverbios, nombres propios, preposiciones, interjecciones, fechas-horarios y cifras que los simulados, porque la situación real amplía los referentes⁹⁹ y porque la interacción real no está encorsetada ni guiada.

En relación con (i), B2 se caracteriza por producir notablemente más errores en las categorías señaladas, —esto es, adjetivos 12,12%, adverbios 5,5%, determinantes 3,9%, nombres comunes 10,7%, pronombres 22,2%, preposiciones 13,75%, verbos 19,4%, cifras 21,78%— excepto en *fecha y hora* (25%) donde su porcentaje de error está más de 6 puntos por debajo de B1 entrenado (31,5%) y casi

⁹⁹ Hay que recordar que la situación simulada parte de un guion donde se delimita el destino, se ofrece una lista reducida de horarios y se encorseta la situación. Esto tiene un efecto directo en el repertorio posible de nombres propios (lugares de destino), precios y horarios (cifras y fechas) y en las vacilaciones que pueden tener los clientes al elegir o valorar las posibilidades que les ofrece el operador respecto a sus intereses reales de viaje.

tres veces por debajo de B1 sin entrenar (62,5%). También hay menor porcentaje de errores en conjunciones (5,7%) (aunque las cifras no son tan llamativas como en el caso anterior).

A modo de cierre destacamos una anomalía: la presencia de errores en nativos y no nativos en la categoría de interjección. Se trata de la intervención de las lenguas maternas (en el caso de los nativos, el catalán) en las respuestas automáticas y no controladas en el curso de la interacción, relacionadas con reacciones instintivas, emocionales, etc.

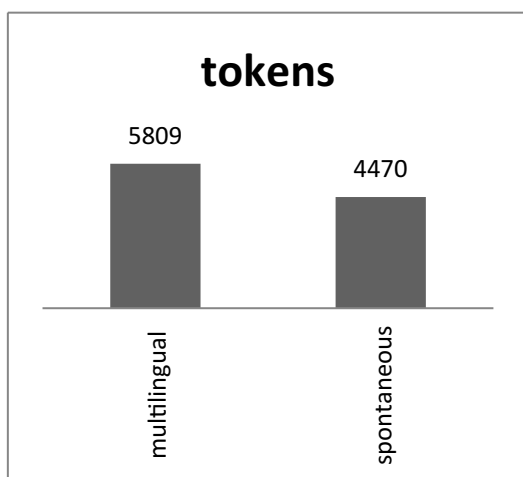
4.1.3. Análisis de la muestra sub16

Con el fin de comprobar si las coincidencias entre los expertos no nativos (B2) y nativos (en una situación espontánea) se mantenían al aumentar la muestra —como explicamos en el capítulo 4— se tomó una muestra sub16 que analizamos a continuación. Recordemos que sub16 está formado por la muestra sub6 de ambos grupos (*FerrovieLE-L_trained-multilingual* y *FerrovieLE-N_spontaneous*) que se incrementa con otros diez diálogos que constituyen esta segunda muestra.

El objetivo de este apartado es (a) confirmar los hallazgos del epígrafe anterior y (b) destacar las diferencias que se hayan puesto de manifiesto con la ampliación de la muestra.

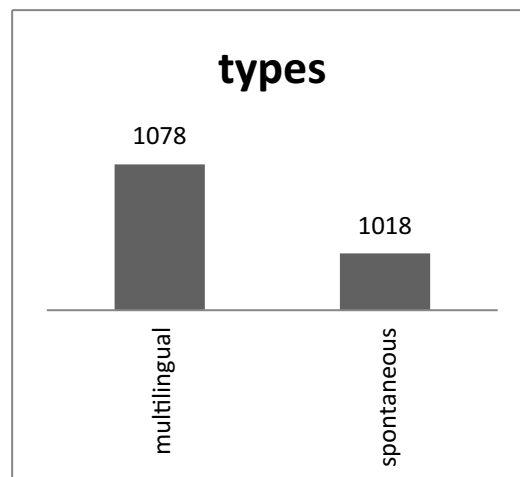
La muestra sub16 tiene un total de 10279 *tokens* que corresponden a 2096 *types* y a 1151 lemas distintos. En las figuras 10 y 11¹⁰⁰ se recoge la distribución de *types* y *tokens* para cada subcorpus.

¹⁰⁰ En las figuras 10 y 11 así como en las figuras 12 y 13, los nombres de los diferentes subcorpus aparecen abreviados: *multilingual* corresponde al subcorpus *FerrovieLE-L_trained-multilingual* y *spontaneous* al subcorpus *FerrovieLE-N_spontaneous*.



Total: 10279 tokens

FIGURA 10. Número total de tokens en sub16



Total: 2096 types

FIGURA 11. Número total de types en sub16

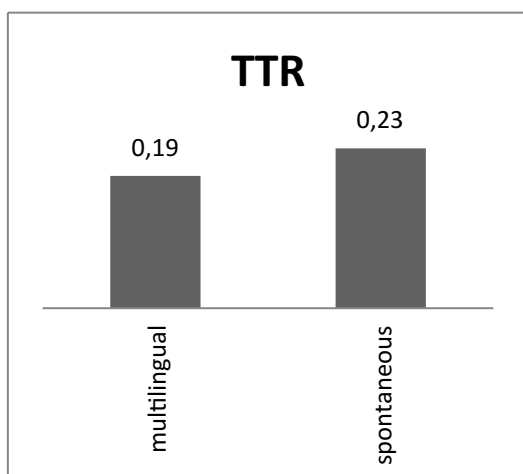


FIGURA 12. TTR en sub16

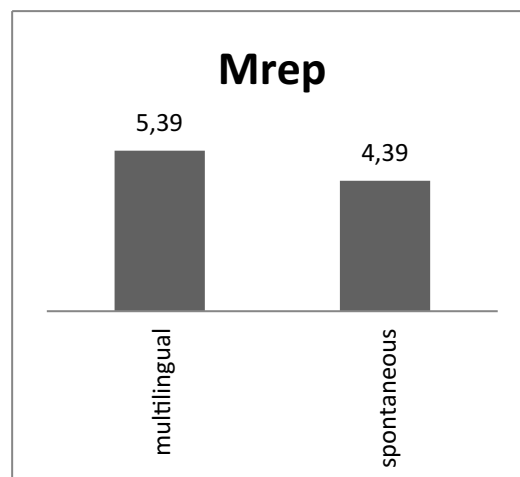


FIGURA 13. Mrep en sub16

En cuanto al número total de palabras existe una diferencia visible entre los dos subcorpus: los aprendices multilingües producen más ocurrencias que los hablantes nativos (5809 y 4470 respectivamente). Lo mismo sucede, respecto al número de *types*, el grupo multilingüe tiene más formas de palabras diferentes que el grupo de hablantes nativos espontáneos (1078 y 108 respectivamente). Si bien se observa en la TTR, figura 12, que el grupo de hablantes nativos espontáneos tiene más riqueza léxica y contiene menos repeticiones, como vemos en la figura 13.

Como vimos en el apartado anterior para las muestras sub6, el entrenamiento tiene efecto en los no nativos de B1, pero no se pudo verificar qué ocurría sin

entrenamiento con los aprendices de B2 puesto que no pudimos disponer de una muestra sin entrenamiento.

No obstante, y todavía también en las muestras sub6, vimos que sí había diferencias claras entre los no nativos de B2 y los nativos espontáneos que quizá convenía explorar más con los datos disponibles para estos dos grupos. Más interesante todavía resultaba constatar que había errores en los nativos espontáneos en categorías como la preposición. Aunque eran poco apreciables en la muestra, parecía prometedor recurrir a una muestra mayor donde serían más visibles los errores y donde las diferencias y las coincidencias que se apuntaban serían más evidentes.

Por estas razones expuestas, decidimos ampliar la muestra con un total de 16 diálogos (sub16) con el objeto de ver:

- a) Las características de la espontaneidad en los nativos que, entre los hablantes nativos espontáneos y los nativos en situación de simulación en el corpus *FerrovíELE*, los espontáneos son más expertos que los simulados y que no borran ni maquillan los errores en el transcurso de la producción¹⁰¹. Al ampliar la muestra vemos que, en efecto, hay errores en más categorías. En concreto, en todas excepto en los nombres propios (por ejemplo, no los hay en los nombres de trenes: AVE, ALVIA ni en los de ciudades).
- b) Si se confirman las características privativas de B2 que vimos en el corpus sub6; esto es, el rasgo (d) de arriba que corresponde a la abundancia de conjunciones y al rasgo (a) que corresponde a la presencia de pausas/interrupciones.

¹⁰¹ Recordemos que se trata de grabaciones secretas en situación real, no planificada y que, por tanto, la producción contiene las vacilaciones, repeticiones, etc. propias del discurso espontáneo. En cambio, los que tienen tiempo para planificar, en algunas ocasiones hacen incluso guiones de sus intervenciones, léxico, etc. que no siempre podemos controlar. De ahí que consideremos más relevante tener en cuenta los nativos espontáneos como la meta (o *target*) de la prueba, como el tipo de producción natural o modélica que se pretende imitar.

	FerroviELE-L_trained-multil -TOTAL:16 (Nivel B2)			InfoTren-N_spontaneous -TOTAL:16 (Nativos)			TOTAL
	tokens	%text	%error	tokens	%text	%error	
Adj.	211	3,63	11,37	130	2,91	2,31	
Adv.	680	11,71	6,32	565	12,64	4,78	
Conj.	551	9,49	5,44	383	8,57	1,04	
Det.	579	9,97	13,3	474	10,6	1,27	
Interj.	202	3,48	5,45	195	4,36	4,62	
Int-pau.	322	5,54	0	66	1,48	0	
NC	911	15,68	9,22	548	12,26	0,36	
NP	191	3,29	12,04	187	4,18	0	
Pron.	221	3,8	17,19	329	7,36	2,18	
Prep.	594	10,23	11,45	566	12,66	1,41	
V	1013	17,44	12,14	733	16,4	0,68	
F/H	96	1,65	17,71	150	3,36	0,67	
Cifra	238	4,1	16,39	164	3,67	1,22	
Tokens	5809			4470			10279
Types	1078			1018			2096
Lemas	536			615			1151
TTR	0,19			0,23			
Mrep	5,39			4,39			

TABLA 7. *Subcorpus FerroviELE-L_trained-multi-sub16 e InfoTren-N_spontaneous-sub16: datos morfológicos*

Los datos de sub16 de la tabla 7 permiten confirmar que los aprendices de B2 utilizan más conjunciones también (9,4%), como en el caso (d) de los rasgos de análisis de la muestra sub6. Pero para (a), rasgo también incluido en el análisis de la muestra sub6, que corresponde a la presencia de pausas, en la muestra sub16 hay tanto en los aprendices de B2 (5,5%) como en el subcorpus de hablantes nativos espontáneos (1,4%). Esto era esperable en una conversación espontánea (nativa), si bien la proporción sigue siendo muy alta para B2 porque se suman, como es esperable, problemas de competencia.

En relación con las demás categorías, la ampliación de la muestra con diez diálogos más a sub16 no arroja diferencias, esto es, se mantiene la proporción de sub6 con pocas oscilaciones, que no parecen significativas.

Sin embargo, más revelador resulta considerar los errores que antes no aparecían en los hablantes nativos espontáneos. En sub6, solo había errores en las

categorías de interjecciones (por *code-switching* o cambio de lengua) y de preposiciones. En sub16, en cambio, están afectadas por errores todas las categorías, excepto los nombres propios.

En cuanto a la comparación de los errores de los aprendices de B2 con los nativos en situación espontánea de la muestra sub16, se constata que se acercan mucho en dos categorías:

- adverbios: 6,32% y 4,78% e
- interjecciones: 5,45% y 4,62%.

Curiosamente, los adverbios era lo que diferenciaba entre todos los subcorpus sub6 a los no nativos de B1 sin entrenamiento.

Aunque no vamos a detenernos en ello, es llamativo también observar el 2,18% de errores en los pronombres que encontramos en los nativos espontáneos (por concordancia, leísmo, etc.) propios de la cohesión y/o coherencia discursiva. Ambos rasgos, los adverbios y los pronombres, son mecanismos lingüísticos difíciles del discurso y no dan el 100% de “limpieza” nunca en el habla espontánea nativa. La presencia de estos errores, por tanto, permite intuir que la muestra ya es suficientemente representativa puesto que aparecen estos rasgos propios de la espontaneidad.

4.2. Análisis del vocabulario

Para analizar el vocabulario que utilizan los hablantes nativos y los aprendices en este escenario comunicativo nos fijaremos, en principio, en los lemas de la muestra sub 6, con la finalidad de recoger el léxico central o nuclear (*core*). En esta sección se presentan los lemas de la muestra sub6, organizados en i) lemas compartidos entre todos los subcorpus y ii) lemas exclusivos de cada subcorpus.

El corpus *Ferroviele* contiene un total de 1849 lemas distintos. De dicho total 1344 lemas pertenecen a la muestra sub6, sin olvidar los 1151 lemas de la muestra sub16. No obstante, en esta sección no se van a tener en cuenta los lemas de la muestra sub16.

En la figura 15 podemos comprobar la relación de lemas entre los grupos no nativos y nativos en la muestra sub6. Se observa que el vocabulario en los aprendices

aumenta progresivamente con el entrenamiento (entre el B1 y el B2) y con el nivel de aprendizaje (de B1 a B2). Mientras que los hablantes nativos ya sea en una situación simulada o espontánea producen prácticamente la misma cantidad de lemas. Gráficamente, la distribución de lemas en los subcorpus sub6 queda de la siguiente forma:

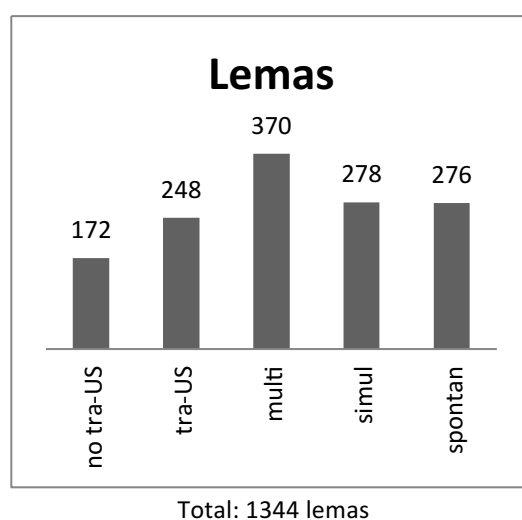


FIGURA 14. Número total de lemas en sub6

Con el objetivo de analizar el léxico que utilizan los diferentes grupos y poder establecer qué vocabulario es el nuclear de la situación comunicativa, primero, se identifica el léxico compartido por todos los grupos y, después, el léxico exclusivo de cada subgrupo. Esta organización se compone de 6 listas distintas y, además, se incluye la frecuencia en que se utiliza este vocabulario:

- lemas compartidos por los cinco subcorpus;
- lemas exclusivos de los nativos simulados;
- lemas exclusivos de los nativos espontáneos;
- lemas exclusivos de los aprendices B2 entrenados multilingües;
- lemas exclusivos de los aprendices B1 entrenados USA y
- lemas exclusivos de los aprendices B1 no entrenados USA.

Asimismo, con la finalidad de caracterizar el vocabulario del escenario comunicativo a partir de un inventario básico que recogerá todas estas listas de lemas, a su vez, estas se presentan distribuidas por categorías gramaticales (discursivo, nombre, adverbio, adjetivo, verbo, preposición conjunción, pronombre y

determinante) y por otras realizaciones lingüísticas (expresiones fijas, fórmulas de cortesía, interjecciones y onomatopeyas).

En relación con (a), la figura 15 incluye el repertorio **léxico compartido por los subcorpus sub6** que constituye el léxico central o nuclear (*core*) del escenario comunicativo de estudio.

a (268)	hola (41)	primero (16)
ah (45)	hora (32)	pues (37)
aquí (15)	ida (97)	que (97)
bien (58)	ir (64)	qué (62)
billete (66)	llegar (51)	querer (96)
bueno (40)	mañana (30)	salir (55)
cómo (9)	más (45)	ser (160)
comprar (23)	muchas_gracias (19)	si (31)
con (66)	muy (28)	sí (174)
costar (19)	no (99)	sólo (49)
cuánto (16)	o (74)	también (26)
de (229)	otro (26)	tener (100)
del (30)	para (74)	tren (116)
descuento (15)	perfecto (21)	tú (16)
el (469)	pero (45)	uno (191)
en (109)	poder (52)	vale (108)
estar (53)	por (65)	viajar (29)
este (39)	por_favor (25)	viaje (19)
gracias (24)	porque (23)	vuelta (61)
haber (87)	posible (7)	y (333)
hasta_luego (13)	precio (28)	yo (53)

FIGURA 15. Repertorio de lemas compartidos por los cinco subcorpus: frecuencia total (I)

CATEGORÍA GRAMATICAL:

- **Discursivo:** otro, también, este, el, del, uno, otro, perfecto
- **Nombre:** AVE*, billete, descuento, hora, ida, información*, noche*, opción*, precio, tren, turista*, viaje, vuelta
- **Adverbio:** aquí, bien, bueno, solo, sí (afirmativo), no, muy, más, mañana, mucho*
- **Adjetivo:** posible, perfecto, primero, individual*
- **Verbo:** estar, haber, comprar, costar, ir, llegar, poder, querer, salir, ser, tener, viajar, incluir*, necesitar*
- **Preposición:** a, de, del, por, para, en, con, hasta*
- **Conjunción:** o, y, si, que, pues, pero, a, porque
- **Pronombre:** yo, tú, usted*, él* [pron. Interrog. → qué, cuánto, cómo]
- **Determinante:** mi*

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: muchas gracias, hasta luego, hola, por favor, gracias, vale

INTERJECCIONES: ah

*Lemas que aparecen en 4 subcorpus de los 5 analizados.

Este repertorio formaría parte del léxico nuclear o central que consideramos en un inventario básico (*core inventory*) para el escenario de petición de información y compra-venta de billetes de tren, es decir, el vocabulario que debería estar presente en todos los niveles de la propuesta: B1, B2 y C2 (nivel nativo).

En el conjunto de estas categorías gramaticales nucleares se hace necesario tener en cuenta algunos elementos centrales (*core*) del rasgo discursivo: *el, uno, perfecto*. Los lemas *el* y *uno* podrían relacionarse con *del* y *otro*, respectivamente. *Perfecto* aparece como respuesta y se etiqueta con la función de *Aceptación*.

Por otro lado, se han incluido en este repertorio de lemas compartidos por los 5 subcorpus algunos elementos lingüísticos que se identifican solo en 4 de ellos, es decir, en todos los grupos excepto en uno de los subcorpus analizados y, por ello, se ha considerado, también, vocabulario nuclear. Veamos, a continuación, estos elementos, centrados en su categoría:

- Nombre: *turista, AVE, información, opción* (los cuatro lemas no aparecen en los hablantes nativos espontáneos), *noche* (no aparece en la muestras de hablantes nativos simulados);
- Adverbio: *mucho* (no aparece en la muestra de aprendices B1 entrenados);

- Verbo: *incluir, necesitar* (ambos lemas no aparecen en la muestra de nativos espontáneos);
- Preposición: *hasta* (no aparece en la muestra de aprendices B1 no entrenados) y
- Determinante: *mi* (no aparece en la muestra de nativos espontáneos).

En relación con (b), la figura 16 recoge el repertorio **léxico exclusivo de los nativos simulados** de la muestra sub6.

a_partir_de (1)	en_metálico (4)	rato (1)
abonar (2)	enchufe (1)	realizar (1)
ambos (1)	entender (1)	recibir (1)
aproximadamente (1)	escoger (1)	recoger (2)
así_como (1)	etcétera (1)	retirar (2)
así_que (1)	feliz (1)	reunión (1)
buenas (1)	forma (1)	ring_ring (2)
climatizar (2)	leer (1)	saludo (1)
cómodo (2)	locución (1)	señorito (1)
compra (1)	mail (1)	sí_que (1)
conectar (1)	merienda (1)	sinceramente (1)
confianza (1)	metálico (1)	suponer (2)
confirmación (1)	mientras_que (1)	taquilla (1)
conforme (1)	mostrador (1)	temprano (1)
continuación (1)	nosotros (1)	trabajar (1)
correcto (1)	obvio (1)	trabajo (2)
cosa (1)	oportuno (2)	uf (1)
dato (5)	ordenador (1)	urgencia (1)
de_nada (1)	pago (1)	usar (1)
deber (1)	para_que (1)	vamos_a_ver (1)
demasiado (1)	parecer (1)	variar (1)
descansar (1)	placer (1)	ventaja (2)
desde_luego (1)	por_encima_de (1)	visa (1)
diario (1)	precisamente (1)	wirefire (1)
disfrutar (1)	prisa (3)	ya_que (2)
disponer (4)	pronto (1)	zona (4)
efectivamente (1)	puesto_que (1)	

FIGURA 16. *Repertorio de lemas exclusivos de los nativos simulados*

CATEGORÍA GRAMATICAL:

- **Discursivo:** Ø
- **Nombre:** compra, confianza, confirmación, continuación, cosa, dato, enchufe, forma, locución, mail, merienda, mostrador, ordenador, pago, prisa, rato, reunión, señorito, taquilla, trabajo, urgencia, ventaja, visa, wifi, zona
- **Adverbio:** aproximadamente, demasiado, efectivamente, en metálico, mientras que, precisamente, pronto, sinceramente, temprano
- **Adjetivo:** conforme, cómodo, correcto, diario, feliz, metálico, obvio, oportuno
- **Verbo:** abonar, climatizar, conectar, descansar, deber, disponer, disfrutar, entender, escoger, leer, parecer, realizar, recibir, recoger, retirar, suponer, trabajar, variar, usar
- **Preposición:** (locución prepositiva) a partir de, por encima de
- **Conjunción y locuciones:** así como, así que, desde luego, para que, puesto que, sí que, ya que
- **Pronombre:** ambos, nosotros
- **Determinante:** Ø

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: buenas, de nada, saludo, es un placer

INTERJECCIONES: etcetera, uf

ONOMATOPEYAS: ring-ring

El vocabulario procedente, propiamente, del subcorpus sub6 nativo bajo simulación es el que contiene más vocabulario que otros subcorpus analizados con muestras de lemas exclusivos, incluso cuenta con más vocabulario propio que los hablantes nativos espontáneos. De igual modo se aprecia que este léxico es más formal, puesto que incluye conversaciones semiespontáneas entre adultos, la mayoría administrativos que trabajan de cara al público.

En relación con (c), la figura 17 está formada por el repertorio **léxico exclusivo de los nativos espontáneos** de la muestra sub6. Recordemos que esta muestra está incluida en el corpus *InfoTren:Persona* (Bonafonte et al. 2000), que fue recogida en 1997, y por cronología no aparecen ciertos términos, como por ejemplo *AVE*, y usan *peseta* como moneda en lugar de *euro*.

aja (1)	hasta_ahora (1)	perdonar (1)
alguno (5)	hombre (1)	pico (1)
año (1)	hum (13)	pregunta (1)

cercanías (1)	igual (1)	preguntar (1)
combinación (1)	información_de_renfe (2)	quilómetro (1)
comprobar (1)	información_renfe (2)	referir (1)
concretamente (1)	Intercity (6)	regreso (1)
costo (1)	internacional (1)	sacar (2)
de_momento (2)	litera (8)	siempre (1)
dirección (2)	mayor (1)	sitio (1)
en_punto (2)	muchísimas_gracias (2)	tasar (2)
enlazar (2)	nada_mas (1)	tramo (1)
exactamente (1)	nene (1)	transbordo (1)
exacto (2)	ninguno (1)	vamos (1)
exigir (1)	nocturno (3)	vía (1)
expreso (3)	o_sea (5)	
folleto (1)	pasaje (1)	

FIGURA 17. Repertorio de lemas exclusivos de los nativos espontáneos

CATEGORÍA GRAMATICAL:

- **Discursivo:** alguno, ninguno, o sea
- **Nombre:** año, cercanías, combinación, costo, dirección, expreso, folleto, Intercity, litera, pasaje, pregunta, pico (y pico), quilómetro, regreso, sitio, tramo, transbordo, vía
- **Adverbio y locución adverbial:** concretamente, de momento, en punto, exacto/exactamente, igual, nada más, siempre
- **Adjetivo:** internacional, mayor, nocturno
- **Verbo:** comprobar, enlazar, exigir, perdonar, referir, sacar, tasar
- **Preposición:** ∅
- **Conjunción:** ∅
- **Pronombre:** ∅
- **Determinante:** ∅

EXPRESIONES FIJAS: vamos

FÓRMULAS DE CORTESÍA: muchísimas gracias, perdonar (perdone), hasta ahora, información de renfe, información renfe

INTERJECCIONES: aja, hombre, hum

Como hemos apuntado la producción de lemas entre hablantes nativos consta, en general, de casi la misma cantidad de lemas (278 simulados frente a 276

espontáneos). En cambio, si se toman como referencia los lemas exclusivos de ambos grupos podemos comprobar que la variedad de lemas es mayor en los hablantes nativos simulados que en los hablantes nativos espontáneos. Probablemente, un factor del que depende la menor variedad de lemas se debe a que los hablantes nativos espontáneos usan mayor cantidad de nombres de ciudades, es decir, orígenes y destinos, y este factor no se tiene en cuenta para el análisis y el recuento de lemas; en cambio, los hablantes nativos simulados desarrollan una actividad dentro de unas características que delimitan sus destinos a cuatro ciudades.

En relación con (d), en la figura 18 se presenta el repertorio **léxico exclusivo de los aprendices B2 entrenados multilingües** de la muestra sub6.

acceso (2)	en_efectivo (1)	pesado (1)
adónde (1)	en_línea (1)	pin (1)
aha (1)	en_seguida (1)	por_cierto (1)
ahora_mismo (1)	en_total (1)	por_último (1)
al_final (1)	entrar (1)	posibilidad (1)
anotar (1)	equipaje (1)	preferencia (1)
asistir (2)	es_decir (1)	presupuesto (2)
buen_día (1)	explicar (1)	propio (2)
cabina_litera (2)	final (1)	proponer (2)
caducar (1)	grande (1)	pueblo (1)
caducidad (4)	hi (1)	qué_guay (3)
calidad (1)	idea (1)	recomendar (1)
cambio (1)	ir_bien (1)	registrar (1)
chao (2)	junto (1)	relación (2)
coche_restaurante (1)	limitar (2)	relajar (2)
conocer (1)	molestar (2)	restaurante_cafetería (1)
consultar (1)	montón (2)	sencillo (1)
continuar (3)	NIE (1)	sentir (1)
dar_igual (1)	nunca (1)	tercero (1)
directamente (2)	obviamente (1)	titular (3)
donde (1)	parar (2)	turístico (3)
económico (2)	partir (2)	un_poco (3)

empezar (1)	pasaporte (1)	vez (3)
en_cuanto_al (1)	permitir (2)	

FIGURA 18. Repertorio de lemas exclusivos de los aprendices B2 entrenados multilingües

<p>CATEGORÍA GRAMATICAL:</p> <ul style="list-style-type: none"> - Discursivo: es decir - Nombre: acceso, cabina-litera, caducidad, calidad, cambio, coche-restaurante, final, equipaje, idea, montón, NIE, pasaporte, PIN, posibilidad, preferencia, presupuesto, pueblo, relación, restaurante-cafetería, titular, vez - Adverbio y locución adverbial: al final, adónde, ahora mismo, directamente, económico, en cuanto al, en línea, en seguida, en total, nunca, obviamente, por cierto, por último, un poco - Adjetivo: grande, pesado, propio, sencillo, tercero, turístico - Verbo y locución: anotar, asistir, caducar, conocer, consultar, continuar, dar igual, empezar, entrar, explicar, ir bien, limitar, molestar, parar, partir, permitir, proponer, recomendar, registrar, relajar, sentir, (pagar) en efectivo - Preposición: junto - Conjunción: donde - Pronombre: ∅ - Determinante: ∅ <p>EXPRESIONES FIJAS: ∅</p> <p>FÓRMULAS DE CORTESÍA: buen día, chao, <i>hi</i></p> <p>INTERJECCIONES: aha, qué guay</p>
--

En relación con (e), la figura 19 se ofrece el repertorio **léxico exclusivo de los aprendices B1 entrenados USA.**

alta_velocidad_española (1)	línea (2)
antes (2)	no_directo (2)
confirmar (1)	obligatoriamente (2)
crear (1)	pequeño (1)
cuando (1)	seguridad (2)
de_inmediato (1)	sin_embargo (1)
español (1)	suficiente (1)
espera (1)	tiquet (1)

fin (1)	transición (1)
finalmente (2)	

FIGURA 19. Repertorio de lemas exclusivos de los aprendices B1 entrenados USA

<p>CATEGORÍA GRAMATICAL:</p> <ul style="list-style-type: none"> - Discursivo: sin embargo - Nombre: alta velocidad española, espera, fin, línea, seguridad, tiquet, transición - Adverbio y locución adverbial: antes, de inmediato, finalmente, obligatoriamente - Adjetivo: español, no directo, pequeño, suficiente, - Verbo: confirmar, crear - Preposición: Ø - Conjunción: cuando - Pronombre: Ø - Determinante: Ø <p>EXPRESIONES FIJAS: Ø FÓRMULAS DE CORTESÍA: Ø INTERJECCIONES: Ø</p>
--

En relación con (f), la figura 20 recoge el repertorio **léxico exclusivo de los aprendices B1 no entrenados USA.**

ay (1)	estudiante (1)
base (1)	horrible (1)
comer (1)	prestación (1)
dólar (3)	puah (1)
españa (1)	tanto (1)

FIGURA 20. Repertorio de lemas exclusivos de los aprendices B1 no entrenados USA

<p>CATEGORÍA GRAMATICAL:</p> <ul style="list-style-type: none"> - Discurso: Ø - Nombre: base, dólar, España, estudiante, prestación - Adverbio: Ø - Adjetivo: horrible, tanto

- **Verbo:** comer
- **Preposición:** Ø
- **Conjunción:** Ø
- **Pronombre:** Ø
- **Determinante:** Ø

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: Ø

INTERJECCIONES: ay, puah

Una vez presentado el léxico compartido por los subcorpus y detectado el léxico central o nuclear (*core*) del escenario comunicativo en estudio (figura 15), por un lado, y el vocabulario exclusivo de cada subcorpus (de la figura 16 a la figura 20), por otro. A continuación, tomamos el léxico compartido por todos los grupos del subcorpus sub6 y analizamos la frecuencia en que se utiliza este vocabulario en cada subcorpus en la figura 21 con la finalidad de detectar qué léxico caracteriza cualitativamente a cada subcorpus. Esto es, la presencia de algunas palabras propias de la variedad nativa contribuye a dar más cualidad de nativo a algunos subcorpus de aprendices: por ejemplo, palabras muy específicas como *precio* o *preferente*; la presencia del adverbio *no* que no aparece tanto en los no nativos y el *muchísimas gracias* que solo usan los hablantes nativos espontáneos.

Los datos obtenidos se concretarán en la implementación del escenario en estudio que presentamos en el capítulo 6. Recordamos que esta implementación es relevante para la evaluación de la riqueza léxica y de la progresión esperada por efecto del entrenamiento entre grupos de aprendices.

Lemas_sub6	spontan ¹⁰²	simul	multi	tra-US	no tra-US	Total
a	63	30	101	46	28	268
ah	4	2	28	8	3	45
aquí	1	2	7	3	2	15
barcelona	5	6	7	10	8	36

¹⁰² Los nombres de los diferentes subcorpus aparecen abreviados: *spontan* corresponde al subcorpus *Ferroviele-N_spontaneous*; *simul* al subcorpus *Ferroviele-N_simulated*; *multi* al subcorpus *Ferroviele-L_trained-multilingual*; *tra-US* al subcorpus *Ferroviele-L_trained-USA* y *no tra-US* al subcorpus *Ferroviele-L_no trained-USA*.

bien	6	10	19	13	10	58
billete	7	9	23	15	12	66
bueno	9	2	12	9	8	40
cómo	2	2	2	2	1	9
comprar	1	1	14	3	4	23
con	5	12	28	12	9	66
costar	1	1	8	5	4	19
cuánto	2	3	5	3	3	16
de	46	48	76	44	15	229
del	9	7	6	4	4	30
descuento	1	1	5	6	2	15
el	128	76	143	74	48	469
en	35	25	22	15	12	109
estar	2	6	28	13	4	53
este	12	4	12	2	9	39
gracias	2	5	7	3	7	24
haber	2	5	42	21	17	87
hasta_luego	3	1	2	2	5	13
hola	6	5	12	11	7	41
hora	2	6	16	6	2	32
ida	18	13	34	14	18	97
ir	14	2	26	12	10	64
llegar	9	6	28	5	3	51
mañana	7	7	3	7	6	30
más	5	5	15	11	9	45
muchas_gracias	2	4	6	4	3	19
muy	6	7	9	1	5	28
no	33	14	21	13	18	99
o	8	12	21	13	20	74
otro	5	2	14	4	1	26
para	16	6	30	8	14	74
perfecto	1	5	8	6	1	21
pero	2	5	22	9	7	45
poder	10	12	20	7	3	52
por	21	6	12	6	20	65
por_favor	1	7	7	4	6	25
porque	4	4	8	2	5	23
posible	1	1	1	2	2	7
precio	7	1	12	5	3	28
primero	9	2	1	1	3	16
pues	11	13	4	7	2	37
que	26	7	43	18	3	97
qué	16	9	18	8	11	62
querer	8	10	33	20	25	96
salir	17	8	18	6	6	55
ser	27	21	52	35	25	160

si	5	4	16	3	3	31
sí	34	13	82	17	28	174
sólo	1	8	15	9	16	49
también	3	1	12	3	7	26
tener	11	22	37	15	15	100
tren	7	17	47	25	20	116
tú	3	2	7	3	1	16
uno	31	21	72	41	26	191
vale	17	10	49	14	18	108
viajar	3	10	8	5	3	29
viaje	2	6	8	2	1	19
vuelta	16	7	25	7	6	61
y	54	29	159	44	47	333
yo	10	9	18	11	5	53

FIGURA 21. Repertorio de lemas compartidos por los cinco subcorpus: frecuencias distribuidas por subcorpus (II)

En cuanto a la figura 21, señalamos las palabras más frecuentes y, luego, comentamos las más relevantes por categorías para el dominio o escenario comunicativo, quedando fuera preposiciones, adverbios, determinantes, conjunciones y pronombres. Por ello, nos detenemos en: i) las fórmulas de cortesía; ii) los nombres y iii) los verbos.

En el escenario de estudio las **fórmulas de cortesía** más frecuentes son las siguientes:

- saludos (*hola*)
- agradecimiento (*gracias, muchas gracias, por favor*)
- afirmación o negación (*sí, no, perfecto*)

Las fórmulas de cortesía caracterizan a los hablantes nativos frente a los aprendices, por ejemplo, todos los grupos agradecen con *muchas gracias*, si bien solo los hablantes nativos utilizan *muchísimas gracias*.

Por otro lado, se puede considerar la negación con el adverbio *no* como un rasgo nativo, dado que la negación no se emplea de la misma forma en todas las culturas. Otra justificación para considerar el uso de la negación se debe al efecto simulación, es decir, la negación tiene más presencia en los hablantes nativos espontáneos que en los hablantes nativos en situación de simulación y en aprendices,

porque los hablantes nativos realizan realmente la tarea y los nativos y los aprendices en situación de simulación escenifican.

En cuanto a los **nombres** destacan: *AVE, billete, descuento, hora, ida, información, noche, precio, tarjeta, tiquet, tren, turista, viaje, vuelta*. Este vocabulario es el nuclear, el propio del escenario. Si bien es cierto se han incluido en este repertorio de lemas compartidos por los 5 subcorpus algunos elementos lingüísticos que se identifican solo en 4 de los grupos y, por ello, se han considerado, también, vocabulario nuclear, que son *AVE, información, noche, tiquet, turista*.

Entre dicha relación hay palabras que o son transferencias de sus lenguas maternas como *tiquet* o bien se utilizan solo en la simulación y no en el discurso espontáneo como *información*.

Por último, dentro de los **verbos** sobresalen *costar, haber, ir, llegar, salir, viajar*, los verbos modales *poder* y *querer* versus *preferir*. Se han excluido de la lista de vocabulario nuclear los verbos *ser, tener* y *estar*, porque no podemos discriminar si son auxiliares o uso propio, aunque son frecuentes los tres verbos. En cambio, *hay* es propio de la función *pedir información* y lo hemos considerarlo porque es muy frecuente en los aprendices, caracterizándolos.

Estos verbos modales en el próximo capítulo 5 van a ser tema de estudio junto con las fórmulas de cortesía, bajo el paraguas de la pragmática y la cortesía, con la finalidad de analizar la variación lingüística entre los aprendices y, a su vez, comparar los datos obtenidos con la variación nativa, recogida en este escenario. Esto es, se van a analizar los datos de los hablantes nativos y no nativos para compararlos y, posteriormente, contrastarlos. Este enfoque de estudio recibe el nombre de transcultural (*crosscultural*), porque se realiza un análisis con componentes lingüísticos procedente de dos o más culturas. Asimismo, se va a ofrecer un análisis exhaustivo en el capítulo 5 de las funciones comunicativas que contiene el escenario de información y compra-venta de billetes de tren.

CAPÍTULO 5. ANÁLISIS DE DATOS: FUNCIONES

En esta sección se presenta el análisis de datos relativo a la información pragmático-funcional del corpus oral *FerrovieELE*, de acuerdo con los objetivos de esta investigación, el análisis de la variabilidad nativa y no nativa. Nos acercamos a los datos desde un análisis funcional y dentro de este tipo de análisis, se abordan las producciones de los hablantes nativos y los no nativos no como simples conjuntos de formas gramaticales, sino desde las dos líneas de estudio (Ellis-Baruiken 2005; Ellis 1992):

a) de la función a la forma, se ha intentado dar cuenta de las funciones del lenguaje, por ejemplo, *expresar planes futuros*. Esta forma de análisis es posible gracias a la anotación pragmático-funcional llevada a cabo en el corpus. La descripción de estos resultados se desarrolla en este capítulo;

b) de la forma a la función, se ha tratado de abordar los datos desde las formas lingüísticas en sí, por ejemplo, el uso de *este* en las producciones de los hablantes nativos y los no nativos. Este tipo de análisis se puede aplicar gracias a la anotación morfosintáctica previa. Los resultados se describen en el capítulo 4.

5.1. Sistematización de modelos de interacción a partir del análisis de los datos (*bottom-up*)

Las funciones comunicativas descritas en el capítulo 3 permiten realizar la anotación pragmático-funcional del corpus *FerrovieELE*, diálogo por diálogo. A partir de las nociones y las funciones encontradas, se establecen los turnos (por rol/participante) prototípicos y los patrones funcionales para cada secuencia de *transacción*, esto es: petición de información, compra-venta y reserva. Para la codificación de cada diálogo hay que partir, por tanto, de la modelización del escenario comunicativo de transacción de información y compra-venta (tabla 8) obtenida. Insistimos, a partir del análisis de una muestra representativa del corpus *InfoTren:Persona* (exclusivamente nativo) inicialmente, y enriquecida posteriormente con los datos de nueva obtención, de hablantes nativos y no nativos, en situaciones simuladas, se constituyen los subcorpus de *FerrovieELE*. En esta sección se describe el conjunto de modelos prototípicos de interacción obtenidos del análisis de los datos.

La observación de datos procedentes de todos los corpus mencionados ha permitido establecer tres modelos de interacción para el escenario de transacción contemplado, que se corresponden con: petición de información, compra-venta y reserva; cada uno de dichos modelos se presenta en la tabla 8. Recordamos que se han establecido mediante un análisis *bottom-up*, esto es, estos modelos se han obtenido a partir de una situación espontánea (no simulada) de los propios diálogos, que se han replicado en diversas situaciones de recogida, todas ellas semiexperimentales, con nativos y no nativos, a fin de ampliar el volumen de datos empíricos disponibles sobre la situación y sobre el escenario planteado. Estos modelos, en su detalle, permiten afinar en la descripción o previsión de formas y funciones que aparecen en cada submodelo según el rol del participante y el momento de la interacción (turno de réplica o de respuesta; inicio o cierre; aceptación o confirmación, entre otras). El proceso descrito se recoge esquemáticamente en la figura 22, a continuación.

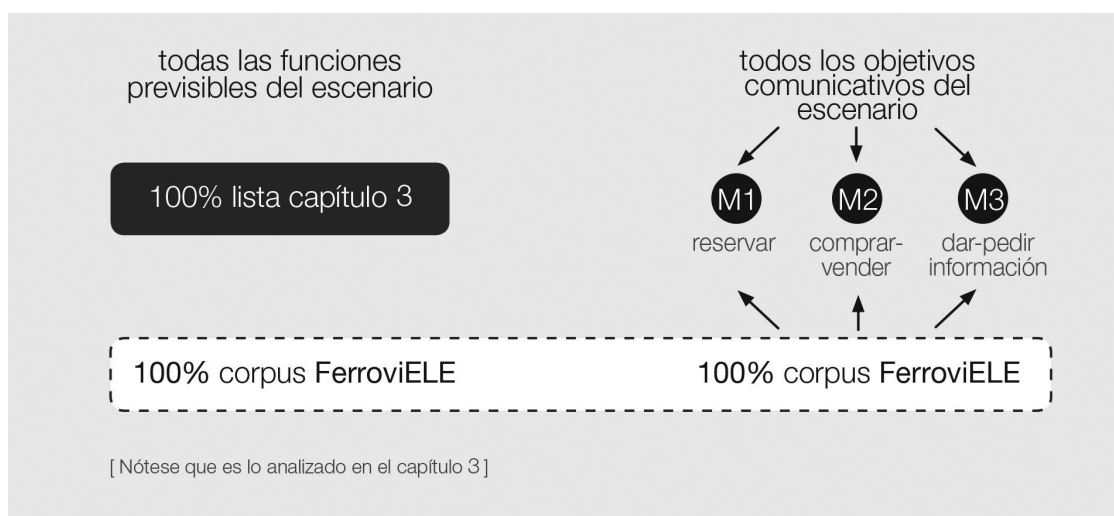


FIGURA 22. Distribución funcional

El modelo prototípico resultante es, por tanto, una síntesis y en él se establecen cuáles son sus partes (secuencias de la interacción) y qué funciones comunicativas son propias de cada una de ellas. Las funciones son sensibles al rol o papel del hablante (cliente u operador) y también lo son al orden de aparición (aislado o como reacción a otro enunciado o información previa), lo que puede implicar una reclasificación de la función —reanálisis de la forma-función—. Por ejemplo, en un saludo, la forma *hola* puede etiquetarse como *inicio de saludo* si,

efectivamente, se produce en primer lugar iniciando la secuencia; pero se etiqueta como *respuesta a saludo* si aparece en segundo lugar. Como en nuestro corpus esa distinción no era relevante, se unificaron bajo la etiqueta *socialising*. En cambio, se mantuvo una diferencia basada en la misma mecánica para: *aceptación/confirmación* (*accepting/confirming*), como se explicó en el capítulo 3.

Como mencionamos arriba, el fundamento teórico que subyace al análisis propuesto es el de función-forma y forma-función (Ellis-Baruiken 2005; Ellis 1992), además de los repertorios mencionados en el capítulo 3, apartado 3.3.2. de esta tesis doctoral. Queremos destacar que es el único caso de estudio, hasta el momento, donde se modeliza tan exhaustivamente la variación de datos de un escenario comunicativo concreto. En el capítulo de conclusiones y líneas futuras se aportará su potencial para aplicaciones posteriores.

5.2. Los tres modelos prototípicos desde el punto de vista funcional

En los modelos recogidos en la tabla 8 (modelos de transacción) se puede observar que cada diálogo se estructura en tres grandes partes diferenciadas: inicio de diálogo, desarrollo de la transacción y cierre de diálogo (columna 1). En cada una de estas partes, se distinguen diferentes funciones (columnas 3, 5 y 7) que pueden variar en función del rol del interlocutor (operador o cliente) y del orden de intervención de cada interlocutor en la transacción (columnas 2, 4 y 6).

	MODELO 1 (M1)		MODELO 2 (M2)		MODELO 3 (M3)	
INICIO DE DIÁLOGO	O	Initiating conversation / Socialising	O	Initiating conversation / Socialising	C	Initiating conversation / Socialising
	C	Initiating conversation / Socialising	C	Initiating conversation / Socialising	O	Initiating conversation / Socialising
DESARROLLO DE TRANSACCIÓN	C	Request a service: making a reservation	C	Request a service: buying a ticket	C	Request a service: information
	O	Giving precise information Enquiring about	O	Giving precise information Enquiring about	O	Giving precise information Enquiring about
	C	Enquiring about • Describing plans • Expressing: feelings/knowledge/preference	C	Enquiring about • Describing plans • Expressing: feelings/knowledge/preference	C	Enquiring about • Describing plans • Expressing: feelings/knowledge/preference
	O	Giving precise information Confirming	O	Giving precise information Confirming	O	Giving precise information Confirming
			C	Accepting	C	Accepting
	C	Accepting	O	• Advice • Suggestions • Obligation and necessity • Checking availability • Counting		
			C	Accepting		
	O	• Advice • Suggestions • Obligation and necessity • Checking availability • Counting	O	Enquiring about information		

	C	Accepting / Making a reservation	C	Giving information / Paying	O	<ul style="list-style-type: none"> • Advice • Suggestions • Obligation and necessity • Checking availability • Counting 	*(<i>muy raramente</i>)
	O	Confirming / Enquiring about information					
	C	Giving information / Paying					
	O	Giving product printing					
			O	Giving product printing			
CIERRE DE DIÁLOGO	O/C	Closing conversation / Socialising	O/C	Closing conversation / Socialising	O/C	Closing conversation / Socialising	
	C/O	Closing conversation / Socialising	O/C	Closing conversation / Socialising	C/O	Closing conversation / Socialising	

TABLA 8. Modelos de transacción: funciones comunicativas

Los tres modelos prototípicos comparten las funciones más sociales de la interacción (inicio-cierre). También, pero sujetas a variación (según competencia/nivel comunicativo de la L2/LE y estilo personal), las confirmaciones, las aceptaciones y la función fáctica. El resto de funciones, como veremos, ya no es compartido y justifica la división en los tres modelos reconocibles con relativa facilidad, a partir de la finalidad comunicativa. Así hacer una reserva es claramente distinto de pedir información. Comprar un billete se acerca a hacer una reserva, pero se diferencia suficientemente de la petición de información.

A su vez, el peso del rol de cliente o de operador también diverge en cada modelo. En el caso de dar información, las secuencias expositivas del operador predominan y las secuencias de preguntas son más cortas. En cambio, en el caso de los modelos de compra-venta y reserva las secuencias de preguntas predominan para el rol de cliente y, en general, a lo largo del diálogo.

En cuanto a la interacción/negociación propiamente dichas, la reserva y compra-venta de billetes son las que más transacción conllevan (modelo 1 y modelo 2, M1 y M2 en adelante). Las secuencias de transacción exigen mayor repertorio (componente cualitativo) y cantidad (componente cuantitativo) de funciones y, por consiguiente, de conocimiento y manejo de formas y estructuras gramaticales y nociones léxicas, esto es, de competencia léxico-gramatical. Esto puede verse, por ejemplo, en el M1 en el rol de operador donde aparecen como funciones comunicativas potencialmente pertinentes o candidatas a usarse: *advising, suggesting, expressing obligation and necessity, checking availability* y *counting*. Un total, pues, de cinco funciones frente a una rara presencia de estas en la secuencia de dar información, modelo 3 (M3 en adelante). A su vez, en el M2, compra-venta de billetes, aparecen las mismas cinco funciones comunicativas del M1, reserva, y no en el M3, petición de información.

La diferencia entre los M1 y M2 es factible porque la reserva de billete incluye una secuencia de transacción de información personal ausente en el M2 de compra-venta directa. Por lo tanto, el factor +/- transaccional, que es gradual, permite diferenciar modelos. También tendrá que permitir diferenciar la competencia de los aprendices, en función de si captan y reflejan en sus producciones la mayor o menor implicación personal y realizan de forma adecuada la transacción.

Asimismo, la tipología de secuencias desde el punto de vista +/- argumentativo (que está relacionada con el conocimiento y puesta en práctica de géneros textuales) permite distinguir también cada modelo, puesto que cada uno de ellos comporta una mayor o menor presencia de argumentación en función del objetivo comunicativo global último. La figura 23, a continuación, presenta lo que acabamos de exponer en síntesis:

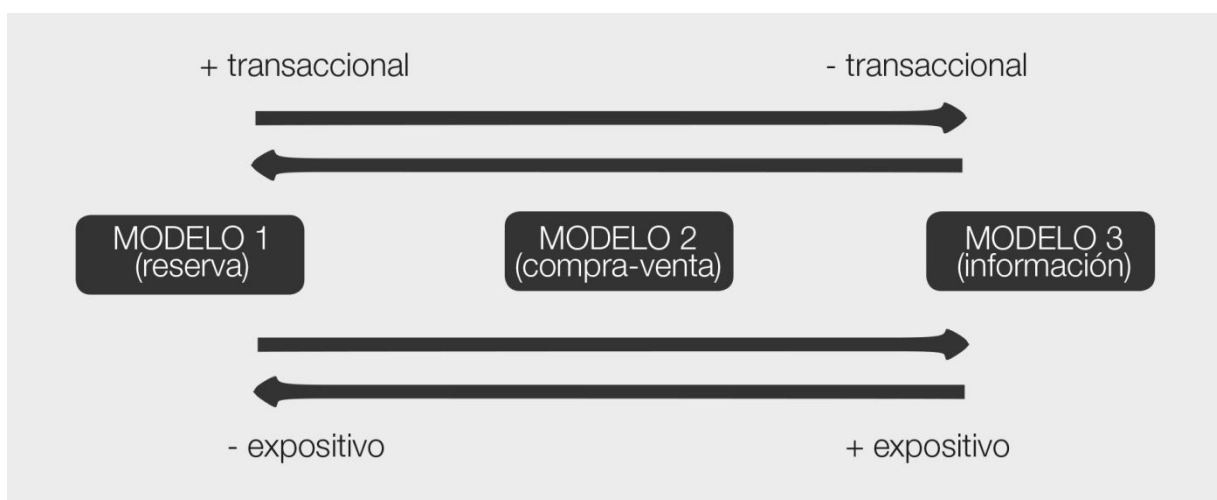


FIGURA 23. Rasgos y secuencias

Una vez detallados los tres modelos prototípicos de la transacción, en la figura 24 se presentan todas las funciones comunicativas de la situación, tanto las más específicas como las menos específicas. Estas se agrupan en los diferentes modelos prototípicos con distintas combinaciones: a) individualmente (M1/M2/M3) b) en grupos (M1+M2) y conjuntamente (M1+M2+M3).

Desde estos diferentes tipos de composición nos interesa resaltar la intersección de los conjuntos, que denominamos matriz nuclear (M1+M2+M3). Así pues, nótese que las coincidencias entre M1+M3 y M2+M3 se reduce a la matriz nuclear, determinante en la situación. Estas son: *accepting, closing conversation, confirming, describing plans, enquiring about, expressing feelings, expressing knowledge, expressing preference, giving precise information, initiating conversation* y *socialising*. Como se puede observar, se trata de las 11 funciones que constituyen el *script* o guion mínimo de una transacción, e incluso negociación, en cualquiera de las tres modalidades (petición de información, compra-venta o reserva). En cambio, las 25 funciones restantes pueden aparecer, o no, y son las que marcan las diferencias o variación entre nativos y no nativos, por un lado, y entre hablantes con mayor o menor fluidez y/o competencia, por otro. Estas 11 funciones son las que

constituyen la matriz expandida. Como veremos en el capítulo 6 donde explicamos la relación de las implicaciones didácticas, los aprendices de B2 y los nativos son los que producen este último tipo de funciones, las 25, predominantemente.

La opcionalidad de la aparición de estas funciones, pues, viene determinada, por factores extrínsecos a la situación (por ejemplo, competencia o estilo personal de interacción y de gestión de la información). Nótese, sin embargo, que, en nuestros datos, los hablantes no nativos más competentes y algunos de los nativos coinciden tanto en el esquematismo, como en la elaboración interactiva. Sin embargo, la variación responde a motivos distintos: en el caso de los no nativos, el esquematismo —el recurso a las once funciones básicas y nada más— suele deberse a la limitación de recursos lingüísticos; en el de los nativos, a la familiaridad con el servicio o a que se imaginan una situación sin opciones (y no abundan en la interacción, pero no por problemas de competencia). Un factor extrínseco que no hay que desdeñar es el factor tarea simulada o tarea entrenada: en ambos casos la percepción que tiene el sujeto de ésta puede incidir en su decisión de ampliar el repertorio de funciones (expandir, confirmar, retomar, completar, concretar, sintetizar, etc.) o no. En los diálogos espontáneos, por su parte, la variabilidad viene dada por el estilo personal de cada sujeto como aprendiz (estilos de aprendizaje) y como hablante (personalidad más o menos extrovertida, comunicativa, entre otros). Esto último, se da tanto en la L1 como en la L2 y en ambos tipos de datos (reales y simulados).

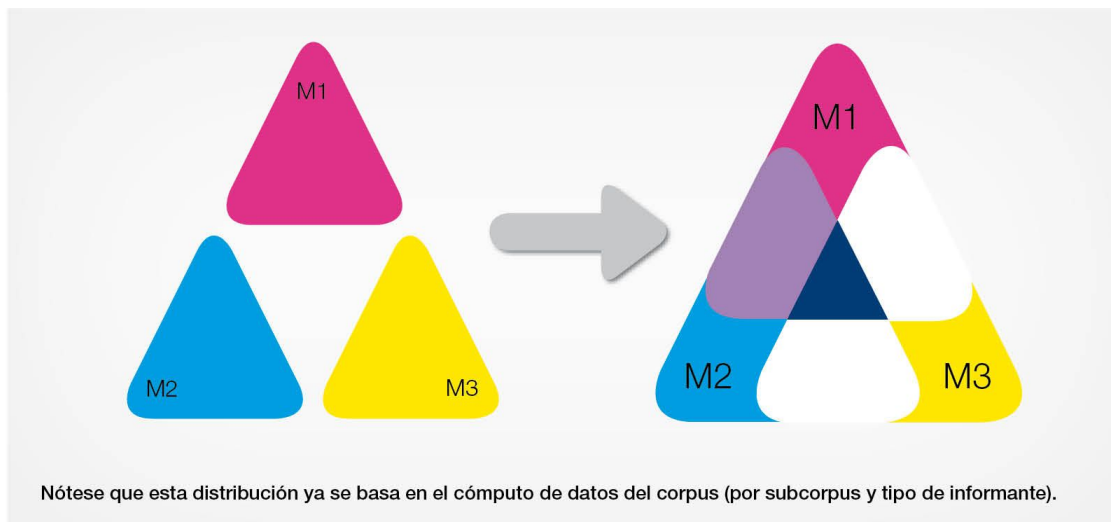
Sin embargo, la intersección de funciones comunicativas en M1+M2 comparte 9 funciones más, además de las nucleares: *advice, checking availability, counting, enquiring about information, giving information, giving product printing, obligation and necessity, paying* y *suggestions*. Esta observación es llamativa porque no ocurre lo mismo con M1+M3 ni con M2+M3, donde no hay intersecciones específicas más allá de la matriz nuclear (como se ve por los espacios sin colorear en la figura 24).

Por último, las funciones específicas de cada modelo son:

–para M1: *making a reservation* y *request a service: making a reservation*;

–para M2: *request a service: buying a ticket*, y

–para M3: *request a service: information*.



- M1
 - Making a reservation
 - Request a service: making a reservation

- M2
 - Request a service: buying a ticket

- M3
 - Request a service: information

- M1 + M2
 - Nótese que allí donde M1 y M2 coinciden es en la transacción de O/C.
 - Advice
 - Checking availability
 - Counting
 - Enquiring about information
 - Giving information
 - Giving product printing
 - Obligation and necessity
 - Paying
 - Suggestions

- M1 + M3 M2 + M3

Nótese que la coincidencia entre M1 + M3 y M2 + M3 se reduce a la matriz.

- M1 + M2 + M3
 - Matriz nuclear
 - Accepting
 - Closing conversation
 - Confirming
 - Describing plans
 - Enquiring about
 - Expressing feelings
 - Expressing knowledge
 - Expressing preference
 - Giving precise information
 - Initiating conversation
 - Socialising

- Matriz expandida
- Agreeing
- Apologising for making a mistake
- Asking for confirmation or denial
- Checking availability
- Check-task
- Correcting
- Counting
- Filler hesitation pause
- Encouraging and inviting another speaker to continue
- Expressing degrees of probability, doubt
- Expressing opinion
- Expressing reaction
- Expressing reason, cause, excuse
- Interrupting

- Inviting agreement
- Making a mistake
- Offer
- Refusing
- Repeating what one has said / other interlocutor has said
- Seeking identification
- Socialising: asking for repetition
- Socialising: asking someone to wait / Socialising: asking someone to wait: empathic offer
- Socialising: asking whether one is heard and understood
- Socialising: supplying a word, repetition or expression
- Spelling
- Synthesising information

FIGURA 24. Funciones comunicativas de la situación de transacción

5.3. Los tres modelos prototípicos desde el punto de vista nocional

De todo el repertorio nocional (léxico) susceptible de aparecer en el escenario seleccionado (cf. *PCIC 2007*), los diálogos reales de los sujetos han permitido obtener una imagen más detallada que descompone las funciones meta casi fotograma a fotograma en la transacción para obtener el léxico necesario para llevar a cabo dichas funciones en este escenario comunicativo. El léxico en nuestro análisis nocional, pues, permite reducir casi a micras el nivel de detalle para cada fase de la transacción y segmentar las funciones (y combinarlas) de una forma más realista de acuerdo con los intereses y necesidades comunicativas tanto del operador como del cliente en cada caso.

La ventaja de esta granularidad es que los elementos se presentan interconectados, más cercanos a la realidad de la comunicación (asociaciones léxicas, ocurrencias y colocaciones) que las tradicionales presentaciones léxicas en listas. Además, el repertorio se observa detallado de manera distinta en función de un rol u otro en la interacción (operador o cliente).

De lo anteriormente dicho se puede prever —aunque verificarlo no es el objetivo de este trabajo— una especie de gramática léxica para cada modelo de transacción y turno (operador o cliente).

A la vista de lo expuesto, se presentan las nociones obtenidas de los corpus de *FerrovíELE* para los tres modelos de transacción. A continuación, en las tablas siguientes se presentan los datos del escenario de transacción para la compra-venta (tabla 10), reserva (tabla 9) o petición de información (tabla 11) de billetes de tren, de la cual se derivan los tres modelos de léxico correspondiente a dicho escenario.

La estructura de las tres tablas corresponde al modelo conjunto ya descrito en la tabla 8, donde también aparecen, a su vez, las tres grandes partes (verticales) en que se organiza cada interacción: inicio de diálogo, desarrollo de la transacción y cierre de diálogo. Cada una de estas partes incluye diferentes funciones que pueden variar, también, según si el rol es el de cliente u operador y según el orden de intervención de cada interlocutor en cada transacción. A su vez, la sucesión de las funciones está supeditada a las nociones (léxicas) con que estén combinadas.

Las nociones (léxicas), en particular, aparecen unidas a la función comunicativa asociada y pueden ser nociones simples (X^1) o nociones combinadas (X^n). En este estudio

(véase los conceptos de función y noción en el capítulo 3), se llama noción a la aparición de una o más palabras referidas a la realidad propia de la situación comunicativa. Ejemplos de noción simple lo constituyen: *hora de salida, tarifa, destino, trayecto, veinticinco euros, tren Ave, Barcelona, Granada*; ejemplos de nociones combinadas serían: *hora de salida y precio y/o tipo de tren; de Barcelona a Granada y tarifa*, etc. Ambas opciones pueden aparecer en una función sola o en una función compleja o combinada como se puede observar en (1) y (2), realizados en los ejemplos seleccionados por hablante con el rol operador (puesto que es el que dispone del conjunto de la información para exponerla toda en el mismo turno):

(1)

– función sola + noción simple:

Eh bueno eh [cuesta [sesenta y cuatro euros]_{NOC=price}] FUNCT=Giving precise information

– función sola + noción combinada:

[[Hay un tren directo]_{NOC=train traject} [esta noche]_{NOC=time} [a las nueve y media]_{NOC=departure time}] FUNCT=Giving precise information

(2)

– función compleja + noción simple:

Pues, [necesito un billete [para hoy.]_{NOC=date}] FUNCT=expressing preference *Eh, [¿puede decirme cuándo puedo [coger el tren?]_{NOC=time}]* FUNCT=enquiring about

– función compleja + noción combinada:

[¡Ah, puedo ayudarte!] FUNCT=confirming *[Hay tres opciones: [hay un tren más rápido se llama Alta Velocidad Española,]_{NOC=train type} [pero es más caro]_{NOC=fare}]* FUNCT=Giving precise information

MODELO 1

INICIO DE DIÁLOGO	O	Initiating conversation / Socialising	
	C	Initiating conversation / Socialising	
DESARROLLO DE TRANSACCIÓN	C	Request a service: making a reservation	X^1 date destination place X^n time prod.train.type
	O	Giving precise information	X^1 arrival time availability of seats availability of tickets buying a ticket departure time fare price seats time train services train stops train trajet train type
	O	Enquiring about	X^1 arrival date arrival time availability of seats credit card CVV number expiration date holder's name number PIN date departure place departure time destination place fare ID document making a reservation passengers payment conditions personal preference price quantity of tickets time train trajet train type
			X^n amount.fare amount.price loc.departure.place loc.destination.place loc.platform prod.train.services prod.train.traject prod.train.type time time.arrival time.date time.departure
	C	Enquiring about	X^1 alternatives arrival date arrival time availability of seats
			<ul style="list-style-type: none"> • Describing plans • Expressing feelings

	C	X^1	buying a ticket departure time fare making a reservation payment conditions preference price seats time train services train trajet train type transport season [tickets travelling with pets	<ul style="list-style-type: none"> Expressing knowledge 	X^1 date departure time fare time train traffic	
			X^n	loc.departure.place prod.train.traffic prod.train.trajet prod.train.type prod.seats time time.arrival time.departure	<ul style="list-style-type: none"> Expressing preference 	X^n loc.departure.place loc.destination.place prod.train.type time.departure
					X^1 arrival date arrival time buying a ticket date departure time destination place fare payment conditions price quantity of tickets time train trajet train type	
	O	Giving precise Information	[las mismas nociones que se detallaron arriba para esta función comunicativa]	Confirming	X^1 amount.fare fac.credit.card loc.departure.place loc.destination.place pers.passengers prod.train.services prod.train.type time.arrival time.date	
					X^n prod.train.traffic prod.train.type time time.date time.departure	
	C	Accepting	X^1	amount.fare fac.credit.card prod.train.trajet prod.train.type prod.printing time time.arrival time.date		

					time.departure
	O	<ul style="list-style-type: none"> • Advice • Suggestions • Obligation and necessity • Checking availability • Counting 			
	C	Accepting			[las mismas nociones se detallaron arriba para esta función comunicativa]
		/			
		Making a reservation			
	O	Confirming			[las mismas nociones se detallaron arriba para esta función comunicativa]
		/			
		Enquiring about information			[las mismas nociones que arriba, para Enquiring about, se detallaron para esta función comunicativa]
	C	Giving information			
			X ¹	credit card	CVV number expiration date holder's name number
				passengers personal time travelling with pets	
			X ⁿ	fac.expiration.date fac.holder.name	
		/			
		Paying			
	O	Giving product printing			
CIERRE DE DIÁLOGO	O/C	Closing conversation / Socialising			
	C/O	Closing conversation / Socialising			

TABLA 9. Modelo de transacción 1: funciones comunicativas y nociones léxicas para hacer una reserva

MODELO 2

INICIO DE DIÁLOGO	O	Initiating conversation / Socialising	
	C	Initiating conversation / Socialising	
DESARROLLO DE TRANSACCIÓN	C	Request a service: buying a ticket	
	O	Giving precise information	<p>arrival time availability of seats availability of tickets buying a ticket departure time fare price seats time train services train stops train traject train type</p> <p>X^1</p> <p>amount.fare amount.price loc.departure.place loc.destination.place loc.platform prod.train.services prod.train.traject prod.train.type time time.arrival time.date time.departure</p> <p>X^n</p>
		Enquiring about	<p>arrival date arrival time availability of seats credit card CVV number expiration date holder's name number PIN date departure place departure time destination place fare ID document making a reservation passengers payment conditions personal preference price quantity of tickets time train traject train type</p> <p>X^1</p> <p>amount.fare fac.CVV.number fac.expiration.date fac.holder.name fac.number loc.destination.place prod.train.traject product.train.type time time.date time.departure</p> <p>X^n</p>
	C	Enquiring about	<p>X^1</p> <p>alternatives arrival date arrival time availability of seats buying a ticket departure time fare making a reservation</p> <p>• Describing plans</p> <p>• Expressing feelings</p> <p>• Expressing Knowledge</p> <p>X^1</p> <p>date departure time fare time</p>

	C		payment conditions preference price seats time train services train trajet train type transport season [tickets travelling with pets	• Expressing knowledge	X^1 train traffic
					X^n loc.departure.place loc.destination.place prod.train.type time.departure
				• Expressing preference	X^1 arrival date arrival time buying a ticket date departure time destination place fare payment conditions price quantity of tickets time train trajet train type
		X^n loc.departure.place prod.train.traffic prod.train.trajet prod.train.type prod.seats time time.arrival time.departure	X^n amount.fare loc.destination.plac prod.train.trajet prod.train.type time time.arrival time.date time.departure		
	O	Giving precise information	[las mismas nociones que se detallaron arriba para esta función comunicativa]	Confirming	X^1 amount.fare fac.credit.card loc.departure.place loc.destination.plac pers.passengers prod.train.services prod.train.type time.arrival time.date
					X^n prod.train.traffic prod.train.type time time.date time.departure
	C	Accepting	X^1 amount.fare fac.credit.card prod.train.trajet prod.train.type		

	C	Accepting	X ¹	prod.printing time time.arrival time.date time.departure
	O	<ul style="list-style-type: none"> • Advice • Suggestions • Obligation and necessity • Checking availability • Counting 		
	C	Accepting		
	O	Enquiring about information		[las mismas nociones que arriba, para Enquiring about, se detallaron para esta función comunicativa]
	C	Giving information	X ¹	credit card CVV number expiration date holder's name number passengers personal time travelling with pets
		/	X ⁿ	fac.expiration.date fac.holder.name
		Paying		
CIERRE DE DIÁLOGO	O	Giving product printing		
	O/C	Closing conversation / Socialising		
	C/O	Closing conversation / Socialising		

TABLA 10. Modelo de transacción 2: funciones comunicativas y nociones léxicas para comprar un billete

MODELO 3

INICIO DE DIÁLOGO	O	Initiating conversation / Socialising		
	C	Initiating conversation / Socialising		
DESARROLLO DE TRANSACCIÓN	C	Request a service: information	X^1	availability of tickets date fare train stops train traject train type travelling with pets
			X^n	amount.fare loc.destination.place prod.train.traject prod.train.type time time.date time.departure
	O	Giving precise information	X^1	arrival time availability of seats availability of tickets buying a ticket departure time fare price seats time train services train stops train traject train type
			X^n	amount.fare amount.price loc.departure.place loc.destination.place loc.platform prod.train.services prod.train.traject prod.train.type time
O	Enquiring about	X^1	arrival date arrival time availability of seats credit card CVV number expiration date holder's name number PIN date departure place departure time destination place fare ID document making a reservation passengers payment conditions personal preference price quantity of tickets time train traject train type	
		X^n	amount.fare fac.CVV.number fac.expiration.date fac.holder.name fac.number loc.destination.place prod.train.traject product.train.type	

	O		time.arrival time.date time.departure		time time.date time.departure	
	C	Enquiring about	X^1 alternatives arrival date arrival time availability of seats buying a ticket departure time fare making a reservation payment conditions preference price seats time train services train trajet train type transport season [tickets travelling with pets		<ul style="list-style-type: none"> Describing plans 	
				<ul style="list-style-type: none"> Expressing feelings 		
				<ul style="list-style-type: none"> Expressing knowledge 	X^1 date departure time fare time train traffic X^n loc.departure.place loc.destination.place prod.train.type time.departure	
					<ul style="list-style-type: none"> Expressing preference 	X^1 arrival date arrival time buying a ticket date departure time destination place fare payment conditions price quantity of tickets time train trajet train type
					X^n amount.fare loc.destination.place prod.train.trajet prod.train.type time time.arrival time.date time.departure	
	O	Giving precise information	[las mismas nociones que se detallaron arriba para esta función comunicativa]		Confirming X^1 amount.fare fac.credit.card loc.departure.place loc.destination.place pers.passengers prod.train.services prod.train.type time.arrival time.date X^n prod.train.traffic prod.train.type time	

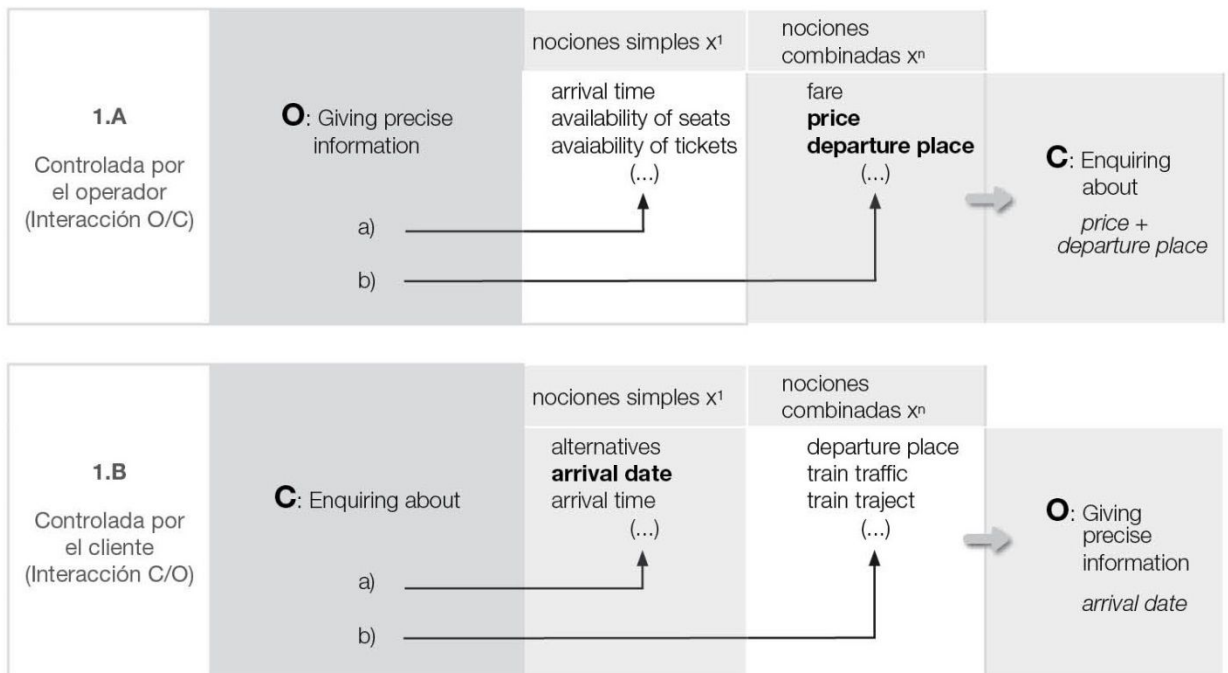
					time.date time.departure
	C	Accepting	X ¹		
	O	<ul style="list-style-type: none"> • Advice • Suggestions • Obligation and necessity • Checking availability • Counting 			
CIERRE DE DIÁLOGO	O/C	Closing conversation / Socialising			
	C/O	Closing conversation / Socialising			

TABLA 11. Modelo de transacción 3: funciones comunicativas y nociones léxicas para pedir información

El orden de la interacción y quién la inicia es relevante. Las secuencias de funciones y nociones en los diálogos, en función de si la interacción la inicia el operador o el cliente, provocan que algunas secuencias asocien una función comunicativa específica con una noción simple o bien la combinen con otra noción, formando una noción compleja, como hemos visto en los ejemplos de arriba (1 y 2). La noción simple o combinada va/n a ser, a su vez, la/s que determinen el léxico empleado en la siguiente secuencia de función más noción en el curso de la interacción, guardando una progresión temática y la coherencia del discurso, regulada por el léxico. En el ejemplo de la figura 25, se observan los exponentes asociados a las funciones comunicativas correspondientes, que, a su vez pueden ser simples o bien estar interconectadas.

Función: Making a reservation

SECUENCIA 1



Siendo 1.A y 1.B opciones que pueden ser alternativas o sucesivas

FIGURA 25. Nociones asociadas a las funciones

Nótese que la función *giving precise information* del ejemplo de la figura 25, controlado por el operador, en el caso a) solo puede tener asociada una noción simple (X¹) como *arrival time*, *availability of seat*, etc. En cambio, en el caso b), esto es, si se aplica a repertorio de nociones combinadas (Xⁿ), como *fare*, *price* o *departure place*, puede incluir más de una (por ejemplo, *price+departure place*). Lo que resulta determinante, pues, es la elección de X¹ o de Xⁿ, o sea, de la cantidad de información (y de la complejidad de esta) que es capaz de gestionar el hablante (dando como resultados las opciones a, b), siendo la razón de la elección tanto la competencia comunicativa como el estilo personal¹⁰³. La respuesta del cliente, en el caso de a) puede ser aceptar y continuar o bien ahondar en los detalles.

En el segundo ejemplo de la figura 25, controlado en esta ocasión por el cliente, en el caso a) solo puede tener asociada una noción simple (por ejemplo, *alternatives*, *arrival*

¹⁰³ No conviene asociar combinación de nociones con nivel de competencia, como veremos más adelante. Los datos mostraron que los nativos (más competentes, en principio) elegían sistemáticamente con más frecuencia funciones con nociones simples (y mayor número de intervenciones), mientras que los no nativos usaban combinaciones y turnos más largos (lo que supone más planificación pero minimizando el número de turnos).

date o *arrival time*), a elegir de la lista. En cambio, en la b) puede combinar más de una noción (por ejemplo, *departure place+train traffic*), también por las mismas razones (competencia o estilo personal de interacción o estilo y capacidad de gestión de información). También el operador, en función de la aceptación, podrá confirmar o bien pasar a otro estadio de la secuencia.

En general, la distribución en el corpus de las funciones y las nociones simples y complejas que se encuentran, muestra que la frecuencia de las funciones con nociones simples (*preguntar el precio de un billete; preguntar la hora de salida; informar de la hora de llegada; etc.*) es más alta —942 en total— que la de las funciones con nociones combinadas (*preguntar por el tipo de tren y el precio; preguntar por el destino y el horario; informar sobre el trayecto y el tipo de tren; etc.*), que suponen 662. En los anexos 5 y 6 se incluyen dos tablas con el recuento detallado de nociones simples y nociones combinadas ordenadas por orden alfabético.

En el próximo apartado detallaremos la diferencia de peso que tiene entre los distintos participantes (nativos y no nativos, espontáneos y simulados) el recurso a nociones simples o combinadas en cada subcorpus asociadas, pues, al nivel de competencia de los aprendices y al factor de espontaneidad o simulación.

En este apartado hemos visto la rentabilidad que supone plantear la presentación del léxico (nociones) por funciones y modelos de interacción, basada en la observación empírica de la gestión de información por hablantes en una situación real. Esto ha permitido presentar el léxico de forma no solo detallada sino también interrelacionada. Sería interesante ver las diferencias entre nuestra propuesta y la de otras más planas, como las de algunos materiales de E/LE o como la del propio *PCIC*. De ello nos ocuparemos en el capítulo 6.

5.4. Análisis del grado de especificidad de las funciones en los distintos subcorpus: funciones de dominio general y de dominio específico (transporte)

Los rasgos sociopragmáticos asociados al uso léxico a través de las funciones pueden detectarse y cuantificarse a través de una codificación específica de funciones de dominio general (o aplicables a cualquier dominio) frente a las específicas del dominio del transporte. Se han etiquetado como FDG las funciones de ámbito o dominio general y

como FDE las propias de este dominio (transporte). Las FDG son funciones simples, sin ninguna noción específica asociada (por ejemplo: *accepting, socialising, confirming*); mientras que las FDE son funciones normalmente asociadas a una noción específica o bien combinadas (*giving precise information: arrival time; enquiring about: departure time and place*).

La proporción de funciones de dominio general (FDG), globalmente, es algo más de una tercera parte que las de dominio específico (FDE) para todos los subcorpus recogidos y analizados. De modo que el porcentaje entre los dos tipos de funciones comunicativas para el corpus *Ferroviele* es:

- FDG → 28,59%
- FDE → 71,40%

A continuación, se describen los resultados obtenidos comparándolos para cada subcorpus. Recordemos que los subcorpus de *Ferroviele* están formados cada uno de ellos por 6 diálogos —llamados muestra sub6—, a fin de extraer los modelos (ver tablas 12 y 14). Además, se ha dispuesto de una muestra posterior más amplia (sub16), que contiene 16 diálogos del corpus multilingüe (B2) entrenado (*Ferroviele-L_trained-multilingual*) y del corpus nativo espontáneo *InfoTren-N_spontaneous* (ver tablas 13 y 15).

En principio, se recogieron 6 diálogos para cada subcorpus y tras examinar los datos se tomó una muestra mayor de los grupos multilingüe y nativos espontáneos, porque pareció relevante asociar el multilingüismo a la competencia experta de los nativos. Si bien simular y entrenar a los no nativos monolingües (ambos B1) parecía dar alguna diferencia, esta se dispara en los no nativos multilingües. Por esta razón se decidió ampliar la muestra para comparaciones específicas entre nativos expertos y no nativos multilingües no expertos y entrenados, comparación que parecía más interesante que la mera distinción entre B1 y B2.

En los subcorpus sub6, los datos muestran (tabla 12) que se utilizan más del doble de FDEs que de FDGs. De este modo, predominarían las funciones de dominio específico. Además, esta característica se confirma uniformemente para todos los subcorpus.

Corpus sub6	FDG	FDE	
No entrenados USA_B1 ¹⁰⁴	21	48	< x2
Entrenados USA_B1	20	50	< x2
Nativos simulados	19	37	< x2
Nativos espontáneos	28	68	< x2
Entrenados multi_B2	29	72	< x2

TABLA 12. Funciones de dominio general y específico para la muestra sub6

Los datos muestran, además, que el entrenamiento en el caso de los B1 monolingües no tiene efecto visible. En cambio, en el caso de los multilingües B2 no nativos, el efecto bien del multilingüismo, bien del entrenamiento (en interacción) hace que se asemejen a los nativos, especialmente a los espontáneos, que, recordemos, son expertos o entrenados (grabaciones secretas en la estación con operadores reales). En cambio, los nativos no expertos y no entrenados, es decir, los nativos simulados, están más cerca de los no nativos B1.

Dado lo sorprendente de esta última constatación, en que se mezcla el factor competencia lingüística (por nivel, por conocimiento de otras lenguas, por entrenamiento) con el de nativez, decidimos ampliar la muestra a 16 diálogos, para ver si persistía la correlación. Como se puede ver en la tabla 13 a continuación, se verifica dicha correlación entre entrenamiento, por un lado, y multilingüismo (que supone una competencia mayor), por otro.

Corpus sub16	FDG	FDE	
Nativos espontáneos	33	95	< casi x3
Entrenados multi_B2	33	87	< casi x3

TABLA 13. Funciones de dominio general y específico para la muestra sub16

Así mismo, si llevamos la comparación del uso de tipo de función (FDG frente a FDE) a la distinción entre roles, esto es, cliente frente a operador en la muestra sub6, se pone de

¹⁰⁴ Los nombres de los diferentes subcorpus en esta tabla 12 como los nombres de los subcorpus de la tabla 13 a la tabla 28 del capítulo 5 aparecen renombrados, en relación con su respectivo subcorpus:

- el grupo *nativos espontáneos* corresponde al subcorpus *FerrovieLE-N_spontaneous*;
- el grupo *nativos simulados* corresponde al subcorpus *FerrovieLE-N_simulated*;
- el grupo *entrenados multi_B2* corresponde al subcorpus *FerrovieLE-L_trained-multilingual*;
- el grupo *entrenados USA_B1* corresponde al subcorpus *FerrovieLE-L_trained-USA* y
- el grupo *no entrenados USA_B1* correspondende al subcorpus *FerrovieLE-L_no trained-USA*.

manifiesto que la distinción entre cliente y operador sigue la misma correlación en todos los grupos, como arriba. En términos cuantitativos, queda como expresa la tabla 14:


Corpus sub6	FDG		FDE	
	cliente	operador	cliente	operador
No entrenados USA_B1	16	13	29	26
Entrenados USA_B1	17	12	33	28
Nativos simulados	14	10	26	21
Nativos espontáneos	21	22	43	45
Entrenados multi_B2	22	22	43	48

TABLA 14. Funciones de dominio general y específico por roles para la muestra sub6

Como se puede ver, se asemejan cuantitativamente los no nativos de nivel B1 con los nativos simulados, por un lado. Por el otro, los no nativos multilingües B2 con los nativos espontáneos. Esto siempre tomando como referencia las muestras de 6 diálogos para cada subcorpus.

Igualmente, al aumentar las dimensiones de la muestra a 16 diálogos, las similitudes entre ambos grupos, los nativos espontáneos y los no nativos entrenados multilingües B2, se mantienen para ambos roles y tipos de función. Se observa que no hay diferencias entre operador y cliente en las FDG, por un lado, ni entre operador y cliente en las FDE, por otro, como se muestra en la tabla 15:

Corpus sub16	FDG		FDE	
	cliente	operador	cliente	operador
Nativos espontáneos sub16	44	49	108	99
Entrenados multi_B2 sub16	48	47	97	103



 > x2

TABLA 15. Funciones de dominio general y específico por roles para la muestra sub16

En resumen, vuelve a coincidir cuantitativamente que el número de FDE, o lo que es lo mismo, de funciones con léxico específico de transporte, son el doble por cada rol respecto a las FDG. Todas aumentan conforme se incrementa el tamaño de la muestra. Por consiguiente, se confirma el patrón de la muestra sub6.

5.5. Efecto del entrenamiento en los subcorpus

Por lo que hace referencia al efecto entrenamiento, recogido en la hipótesis 3, que postula que el entrenamiento tiene efecto en la mejora de la actuación, los datos demuestran que se cumple. En efecto, si se observan con detalle los resultados obtenidos en los subcorpus de la muestra sub6, se constatan diferencias entre ellos (véase tablas 12 y 14). Dichas diferencias se distribuyen entre: a) los grupos de aprendices entrenados antes de la tarea frente a los hablantes nativos en una situación simulada (tabla 16); b) los grupos de nativos en una situación no simulada (espontánea) y los nativos en una situación de simulación que actúan como grupo de control (tabla 17) y c) los grupos de no nativos con y sin entrenamiento (tabla 18). Las diferencias resultantes reflejan, en este caso, la competencia sociopragmática tanto entre no nativos (atribuible a la enseñanza o al multilingüismo) como entre los nativos (espontáneos expertos frente a simulados no expertos). Esta diferencia es particularmente importante, porque refleja la variación, por un lado, intragrupos e intergrupos y, por otro lado, la variación intralingüística e interlingüística¹⁰⁵.

Los datos sociopragmáticos resultantes del análisis se detallan a continuación:

- a) Los hablantes no nativos de nivel B1 y B2, ambos entrenados, usan más FDEs (50 y 72 respectivamente) que los nativos en una situación simulada (37 FDEs). Se podría explicar tales resultados por el efecto entrenamiento en la tarea *role-play* para los no nativos. No obstante, hay que tener en cuenta que los nativos en una situación simulada no fueron entrenados por considerarles competentes en esta situación comunicativa (lo cual deja abierta las características de condiciones de la tarea en ambos grupos, en sentido estricto).

¹⁰⁵ La variación intralingüística se refiere a aquellos fenómenos que se desarrollan a través de la diferenciación de usos y de estructuras de un sistema lingüístico. Mientras que la variación interlingüística supone la utilización de dos o más sistemas lingüísticos y se produce variación por el contacto entre los diversos sistemas lingüísticos.

Corpus sub6	FDG	FDE
Nativos simulados	19	37
Entrenados USA_B1	20	50
Entrenados multi_B2	29	72

TABLA 16. Efecto del entrenamiento (grupos simulados)

b) Los hablantes nativos expertos espontáneos y los nativos no expertos en una situación simulada presentan diferencias claras: los nativos en una situación no simulada utilizan más del doble de FDG y FDE que los nativos en simulación. Vale la pena señalar que los nativos espontáneos ejercen sus roles de cliente y operador sin ficción, mientras que los nativos simulados¹⁰⁶ no.

Corpus sub6	FDG	FDE
Nativos simulados	19	37
Nativos espontáneos	28 > +9	68 > x2

TABLA 17: Diferencias entre nativos espontáneos y nativos simulados

c) Entre los aprendices entrenados y no entrenados, se distingue el efecto entrenamiento de forma visible solo en los aprendices de nivel B2, tanto en las FDGs como en las FDEs, si bien sobre todo es evidente en las FDEs de los B2. En cambio, sorprendentemente, el entrenamiento en el nivel B1 no supone variaciones frente a los estudiantes que simularon sin ningún tipo de ayuda ni de referencia.

Corpus sub6	FDG	FDE
No entrenados USA_B1	21	48
Entrenados USA_B1	20	50
Entrenados multi_B2	29	72

TABLA 18: Diferencias entre no nativos, con o sin entrenamiento

Si se tienen en cuenta los datos recogidos hasta ahora, se hacen evidentes las variaciones entre el grupo de los nativos (con o sin entrenamiento) y el grupo de los

¹⁰⁶ Los participantes en la simulación son profesionales en otras áreas (diseñadores gráficos, secretarías, contables, etc.). No son operadores cualificados, ni estudiantes de lenguas extranjeras.

aprendices entrenados con diferentes niveles de lengua. Así pues, se confirma, además, la hipótesis 4 que postulaba variación sociopragmática. Esto es, se pone de manifiesto que el empleo de las FDGs y de las FDEs no se utilizan de la misma forma en todas las culturas.

Asimismo, y para la distinción por roles, la tabla 19 muestra los resultados del efecto del entrenamiento para cada uno de los grupos de la muestra sub6. Como se puede ver en la tabla 19, los nativos en una situación de simulación usan menos funciones que los nativos en una situación espontánea, casi la mitad de funciones tanto de FDG como de FDE. Igualmente, los no nativos B1 (con o sin entrenamiento) también usan menos funciones que los no nativos entrenados multilingües B2 y que los nativos espontáneos. Por último, se observa que los aprendices entrenados multilingües se asemejan más al grupo nativo espontáneo.

Corpus sub6	FDG			FDE		
	función	cliente	operador	función	cliente	operador
Nativos espontáneos	28	21	22	68	43	45
Nativos simulados	19	14	10	37	26	21
No entrenados USA_B1	21	16	13	48	29	26
Entrenados USA_B1	20	17	12	50	33	28
Entrenados multi_B2	29	22	22	72	43	48

TABLA 19. Efecto del entrenamiento con la tarea role-play

En resumen, los resultados muestran que las diferencias debidas a la dicotomía espontáneo/entrenado para las FDGs y las FDEs son más fuertes cuantitativamente entre nativos en una situación no simulada y no nativos B2 que entre no nativos B1, en ambos roles y en ambas condiciones (con o sin entrenamiento). Entrenar a los no nativos de B1 no desempeña un papel significativo en ninguno de los dos roles, ni para las FDGs ni las FDEs. En cambio, esa dicotomía produce resultados visibles a partir del nivel B2.

En la tabla 20 se presentan los resultados obtenidos con la muestra ampliada de 16 diálogos (sub16), que corrobora las similitudes detectadas en la muestra sub6.

Corpus sub16	FDG	FDE
Nativos espontáneos sub16	61	163
Entrenados multi_B2 sub16	62	159

TABLA 20. Efecto del entrenamiento entre las FDG y las FDE en los subcorpus nativos espontáneos y no nativos B2 para la muestra sub16

En síntesis, las funciones de dominio general y de dominio específico también aumentan al agrandar el tamaño de la muestra. Es interesante revisar con detalle el apartado de análisis morfológico (capítulo 4) para poner en relación qué elementos diferencian a un grupo y a otro cualitativamente. En otras palabras, señalar qué componentes identifican a un hablante nativo frente a un aprendiz de nivel B2, más allá de las cifras que hemos ofrecido aquí.

5.6. Caracterización sociopragmática de los subcorpus a partir de las funciones comunicativas que implican uso de la cortesía: un ejemplo de aplicación

Las funciones FDG y FDE que hemos visto en la sección anterior, como tipos permiten una caracterización del escenario comunicativo de las transacciones en el ámbito del transporte. Pero hasta ahora, se había dejado fuera de las comparaciones entre los distintos subcorpus el componente general que sería común a otras situaciones y escenarios de interacción, como por ejemplo, los aspectos de contacto interpersonal relativos a la cortesía. Alguien podría objetar que la cortesía es relevante en las transacciones y que también diferencia grados de competencia sociolingüística o sociopragmática y que no deberían haber quedado fuera.

El objetivo de este apartado es el análisis de la cortesía a través de la observación de las formas lingüísticas características de determinados actos de habla muy ligados a la cortesía y cuya variación permite identificar a los hablantes nativos y no nativos, por un lado; y el grado de formalidad, adecuación a la situación en mayor o menor grado, por otro. Asimismo, veremos la consistencia o no de esta formalidad entre interlocutores en función del rol de operador o de cliente.

Para llevar a cabo este análisis partimos de los supuestos de los trabajos iniciados por Brown y Levinson (1978, 1987) y de los desarrollos posteriores de tipo interlingüístico o translingüístico de Blum-Kulka et al. (1984, 1989); Bardovi-Harlig y Hartford (1990, 1997);

Kasper (1990, 1992). En particular, se utilizan como patrón para la investigación dos de las nueve categorías de peticiones (*request categories*) procedentes de la propuesta de Blum-Kulka (1987) (véase tabla 21).

5.6.1. Los actos de habla y la cortesía

La pragmática, según Levinson (1983), es una disciplina que estudia la manera en que la lengua funciona en los procesos de comunicación. De los muchos intereses de la pragmática, hay dos aspectos importantes relacionados con el aprendizaje de una LE: los *actos de habla* y la *cortesía*.

Un aspecto difícil en el caso de la producción de hablantes no nativos es la expresión (en el sentido de codificación lingüística) de la distancia social entre los participantes en un intercambio comunicativo, como recoge, por ejemplo en el ámbito translingüístico, el trabajo fundamental de Brown y Levinson (1978, 1987), *Politeness. Some universals in Language Use*.

Según Leech (1983) los factores sociales en la pragmática —como, la edad, el grado de distancia social y el grado de poder entre los participantes— influyen en la selección de estrategias de cortesía, al mismo tiempo que no son las mismas según cada cultura. Un error en la aplicación de las estrategias (por desconocimiento del código, por ejemplo), incide en la imagen (*face*) que podemos tener de un hablante no nativo.

El concepto de imagen negativa, en sí, surge de la noción central de *imagen pública* (*face*), tomada de Goffman (1967) —imagen o prestigio social que cada persona desea mantener (Escandell 1996: 148)—. Desde esta idea, Brown y Levinson (1978, 1987) proponen dos tipos de imagen social: 1) la *imagen negativa* y 2) la *imagen positiva*. Se entiende por *imagen negativa* la necesidad de tener libertad, de controlar el espacio propio y no tener imposiciones de los otros y por *imagen positiva* la necesidad de ser aprobado por los demás.

Estos conceptos se consideran universales en el uso del lenguaje, si bien según la cultura su manifestación o expresión verbal se realizará de manera distinta, generando un código de cortesía, a partir de interpretaciones de la imagen positiva y negativa. Según Kasper (1990: 166) “*there is a variability in types of politeness as an expression for cultural ethos by the different weight cultures place on negative and positive face*”.

El factor sociocultural en el que se desarrolla el acto de habla es, pues, determinante. Delimita unos estándares comunicativos concretos en cada comunidad de habla que hay que tener en cuenta. Emplear comportamientos pertenecientes a otras culturas puede generar situaciones descorteses en la cultura de llegada, o de la L2 que se use en cada caso.

Por esta razón, diversos estudios discuten la universalidad de los actos de habla y señalan la importancia de profundizar en las diferencias interculturales¹⁰⁷. Desde la línea de estudios de pragmática intercultural (Blum-Kulka et al. (1989); Bardovi-Harlig y Hartford (1990, 1997); Kasper (1992)) se ha desarrollado un análisis contrastivo de las peticiones y las disculpas en distintos idiomas, por ejemplo, en el marco del proyecto CCSARP: *The cross-cultural speech act realization project* (Blum Kulka et al. 1989), sin duda uno de los más conocidos. En dicho proyecto se establecen unos parámetros de comparación para analizar y comparar las estructuras internas de las peticiones en distintas lenguas. La metodología de trabajo recurre a cuestionarios o tests, fundamentalmente, escritos como por ejemplo la traducción, el rellenado de huecos, la opción múltiple. Esta metodología se ha aplicado a más de una docena de lenguas, entre ellas el español (Alonso Pérez-Ávila s.f.). En cuanto a los parámetros propuestos por el proyecto CCSARP para identificar las estrategias comunes a diferentes lenguas se destacan:

- a) la variación *translingüística*;
- b) la variación sociopragmática y
- c) la variación intercultural.

La variación translingüística se refiere a las similitudes y las diferencias en los actos de habla en diferentes lenguas. La variación sociopragmática trata de cómo influyen los factores sociales en una misma comunidad de habla. La variación intercultural estudia la comparación de las realizaciones de actos de habla entre hablantes nativos y no nativos en una lengua concreta y en las mismas condiciones sociales y los recursos utilizados en cada una de ellas.

¹⁰⁷ Brown y Levinson defienden una postura “universalista” de las estrategias de cortesía en los actos de habla que amenazan la imagen social (*face threatening act*). Este planteamiento se puede ejemplificar en los mecanismos lingüísticos que se emplean en el caso de las peticiones por medio de una pregunta. En otra posición parece situarse el estudio aplicado a culturas no occidentales de Ide (1989) o las investigaciones de Wierzbicka (1985 y 1991), que discute que se presenten de la misma forma los actos de habla de todas las culturas.

La aplicación de los parámetros mencionados ayuda a describir las estrategias de cortesía en diferentes lenguas. Por ejemplo, en las mencionadas peticiones de objetos y servicios, la presencia (o no) en los enunciados de actos de habla indirectos (frente a los directos), el uso de atenuaciones o la presencia de estrategias de acercamiento al interlocutor, usadas para conseguir el propósito comunicativo del intercambio sin poner en peligro la autoimagen social (*face*¹⁰⁸), marca las diferencias entre situaciones, interlocutores e incluso lenguas intervinientes en el intercambio comunicativo.

Hacer una petición, pues, supone una invasión de la zona personal y de la imagen social del otro que limita, condiciona o puede impedir su libertad de acción (Brown y Levinson 1978, 1987). Así pues, las peticiones, por definición, son una amenaza de la imagen negativa del oyente. Por esta razón, es recomendable *pedir* usando las fórmulas que transmiten la cortesía adecuada en el sentido de que las estrategias lingüísticas utilizadas minimicen la lectura del enunciado como un acto impositivo, o los efectos no deseados, puesto que en sí mismas las peticiones constituyen una amenaza a la imagen negativa del oyente (Blum-kulka y Olshtain 1984). Estos recursos corteses garantizan que se pueda mantener la interacción con el interlocutor y su colaboración sin poner en peligro el éxito comunicativo y el de la petición. Por lo tanto, viendo los riesgos que conllevan las peticiones, son un ámbito de estudio importante en L2 por cuanto que si se realizan inadecuadamente pueden convertirse en un *acto de amenaza a la imagen* (*face-threatening act*) y poner en peligro la comunicación entre un hablante nativo y un no nativo.

Con estas premisas abordamos el análisis de los datos del corpus *FerrovieELE*. En él se tiene en cuenta lo planteado en el mencionado estudio de Blum Kulka et al. (1989), así como los de otros autores que tratan el componente pragmático en la enseñanza de lenguas desde el filtro cultural de la propia (Ballesteros Martín (2001); García (1989); Márquez Reiter (2000); Pinto (2002); Siebold (2006, 2007)). Por todo lo expuesto hasta ahora, hemos aplicado desde el principio, en el propio diseño de las pruebas de elicitación

¹⁰⁸ Para Goffman el concepto de imagen social (*face*), básico en la relación social, es aquella que cumple con lo que la sociedad considera positivo. Ver: Goffman (1967), Brown y Levinson (1978, 1987) y Kerbratt-Orecchioni (2004).

de datos de este corpus, la perspectiva pragmático-funcional¹⁰⁹. Hemos distinguido los roles, hemos hecho una previsión de necesidades para cada interlocutor y rol, hemos atendido a la espontaneidad o no de la situación, y a la nativéz o no del hablante.

En efecto, los subcorpus se han anotado a partir de un enfoque nociofuncional y, posteriormente, se ha aplicado un análisis de datos desde la función a la forma y de la forma a la función. Este análisis nociofuncional permite visualizar y localizar las diferencias en el uso de las estrategias de cortesía entre los hablantes nativos y no nativos para cada rol; entre los no nativos, también, ver diferencias en el uso de las estrategias de cortesía según el nivel de lengua y según las condiciones de la simulación (con o sin entrenamiento). Asimismo, el análisis permite poner de manifiesto la variación entre los propios nativos, lo que constituye no una perspectiva de variación intercultural e interlingüística sino de variación intracultural e intralingüística.

Expresado de otra forma, como se observa en la figura 26, el ámbito de la pragmática translingüística, intercultural representado por Bardovi-Harlig (1999), Blum-Kulka (1987) y Kasper (1992), entre otros, formaría un bloque que podríamos rotular como *macropragmática*, para el que contamos en nuestro estudio con los subcorpus de *FerrovíELE* no nativo. Este bloque intersecciona con el de micropragmática o sociopragmática propiamente, que recoge la pragmática interpersonal específica del español.

En nuestro corpus, concretado en un escenario comunicativo, el de la transacción de información o servicios en el ámbito del transporte, este espacio micropragmático, se ciñe además a la pragmática situacional para saber cómo funciona esta situación concreta extralingüística¹¹⁰ y a la pragmática interpersonal para saber cómo funciona la interacción entre los interlocutores previsibles, los roles operador y cliente.

¹⁰⁹ En el concepto de situación comunicativa en didáctica de una LE aparece el influjo pragmático. Así pues, los manuales para la enseñanza y aprendizaje de una LE se estructuran a partir de ciertas nociones pragmáticas, como son las de acto de habla.

De los actos de habla y de las teorías sobre las funciones del lenguaje surgieron los repertorios de funciones comunicativas para el aprendizaje y enseñanza de una LE. La noción de función comunicativa hace referencia a las intenciones de los interlocutores en el uso del lenguaje en diferentes situaciones de relación social. Son funciones comunicativas, por ejemplo, *informar, pedir, presentarse, quejarse*, entre otras.

¹¹⁰ Se puede estudiar una situación concreta en función de los contextos situacionales, que ofrecen significados y reglas concretas para la interacción entre los interlocutores.

Estudios sobre el español desde la perspectiva intralingüística hay varios: Haverkate (1994), Márquez Reiter (2002), Placencia y Bravo (2000), Escandell (1996), Barron (2005), entre otros. Estos constituyen el bloque de *micropragmática* del español. Sin embargo, no existen casi estudios comparativos entre español y un conjunto no específico de lenguas (multilingüe) ni que combinen lo micro con lo macro (o la variación entre nativos y no nativos).

Para dar cuenta de esto último, se han constituido los subcorpus *InfoTren-N(ativos)_spontaneous* y *FerrovieELE-N(ativos)_simulated*, ambos como grupo de control o referencia de español (peninsular).

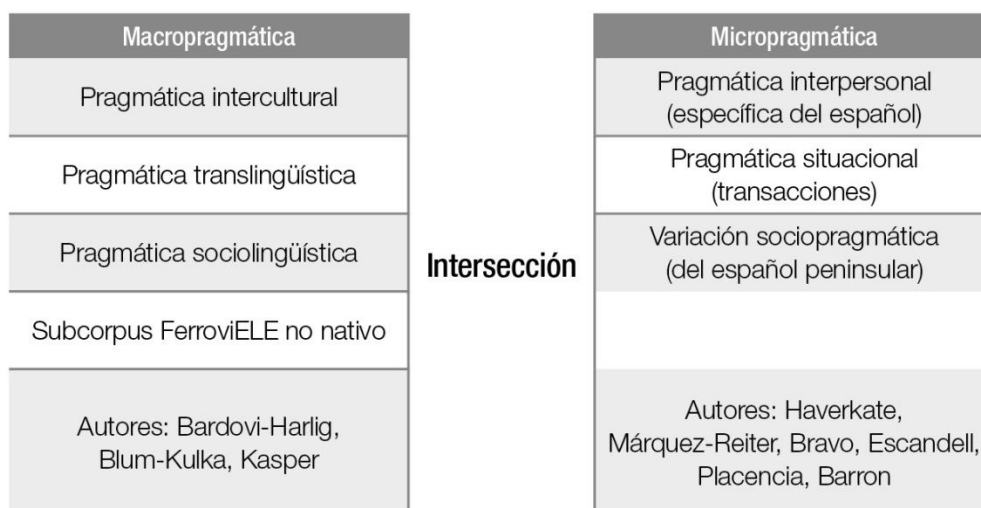


FIGURA 26. *Enfoque del estudio: desde una perspectiva macropragmática a una perspectiva micropragmática*

En la zona de intersección, por tanto, se encontraría la cortesía específica de esta situación o escenario comunicativo para el español, cuya descripción a partir del modelo derivado de los subcorpus nativos permitiría valorar la cortesía no nativa respecto a: su adecuación al rol, su adecuación a la función, su adecuación gramatical y su adecuación léxica.

5.6.2. Los indicadores de cortesía en las peticiones: Hipótesis

Blum-Kulka (1987) y los investigadores que han seguido esta línea de pragmática intercultural e interlingüística establecieron una plantilla de 9 categorías (tabla 21) para los actos de petición (*request categories*). De estas 9 categorías descriptivas se han

seleccionado dos, la 5 *Wants statement* y la 7 *Query preparatory*, porque son las que se encontraban realizadas en el corpus *FerrovieELE* de forma más productiva, para analizar la cortesía.

Descriptive category	Examples
1. Mood derivable	Clean up the kitchen. / Move your car.
2. Performative	I'm asking you to move your car.
3. Hedged performative	I would like to ask you to move your car.
4. Obligation statement	You'll have to move your car.
5. WANTS STATEMENT	I would like you to clean the kitchen. / I want you to move your car.
6. Suggestory formulae	How about cleaning up? / Why don't you come and clean up the mess you made last night?
7. QUERY PREPARATORY	Could you clean up the mess in kitchen? / Would you mind moving your car?
8. Strong Hints (A)	You've left the kitchen in a right mess.
9. Mild Hints (B)	We don't want any crowding (as a request to move the car).

TABLA 21. Blum-Kulka (1987:133) Table 1: Example of the nine Request Categories used in experiments 1-4

La hipótesis de partida es que los indicadores de cortesía seleccionados *Wants statement* y *Query preparatory*, tendrán una presencia distinta en cada subcorpus con diferencias formales en la realización: a) del acto nuclear (*head act*) en situaciones de petición¹¹¹ y b) de los *Wish/Possibly queries (W/P strategies)* (modales Quería.../Podría...), que muestren la asimetría entre los hablantes nativos y no nativos bajo las etiquetas funcionales, de acuerdo con lo establecido por Blum-Kulka et al. (1989).

Las funciones comunicativas en el corpus *FerrovieELE* se representan de dos formas: como se apuntó más arriba, como funciones de dominio general (FDG) o específico (FDE). De estos dos tipos de funciones, se extrae el acto nuclear (*head act*) de la situación de petición (*query, query preparatory*) por su utilidad para describir las diferencias

¹¹¹ Los *head acts* o acto nuclear es la unidad mínima que puede realizar el tipo de acto deseado. Blum Kulka en el análisis de las estructuras internas de las peticiones separa, por un lado, el acto nuclear y, por otro, los otros elementos que constituyen la petición, como son alertadores o apelativos (elementos de apertura que atrae la atención) y los movimientos de apoyo (elementos que modifican el acto de habla). Su trabajo sirve de referencia para esta investigación.

pragmalingüísticas (intra e interlingüísticas) en el manejo de la distancia (más formal frente a menos formal) en la cortesía, diferenciando la variación nativa y no nativa en el uso de las estrategias en los distintos subcorpus.

Además de las situaciones de petición, en este estudio también se analizan otras diferencias específicas entre las formas de cortesía propias de este escenario concreto. A continuación, se presenta el repertorio previsto para este escenario:

- a) *fórmulas de participación social (social contact formulae)*;
- b) *peticiones de servicio* por parte del cliente y *peticiones de aclaración* sobre el producto por parte del operador; así como *confirmación* por el operador o *aceptación* por el cliente y *pago*; todos ellos repartidos entre los dos participantes (operador y cliente)¹¹². Formalmente se tratan bajo la modalidad y la cortesía;
- c) *control de la comunicación y de la información*, que se manifiesta a partir de los rasgos espontáneos de gestión de la comunicación. En otras palabras, el trabajo meta-tarea que supone el conocimiento de ciertos elementos de cómo gestionar la tarea —planificación, secuenciación, segmentación, evaluación—. Este componente no forma parte de los contenidos lingüísticos sistematizables propiamente y, por lo tanto, no es planificable ni se puede entrenar. De ahí, que cuando se ponen en marcha estas estrategias en la comunicación hablamos de *contenido no planificable de la tarea*. A este grupo de estrategias pertenecen las repeticiones para comprobar que el interlocutor entendió, marcadas con la etiqueta *checking*, y las recapitulaciones, señaladas con la etiqueta *synthesizing*. También se incluyen las vacilaciones (rellenos) y las expresiones con función fática, aunque no se trate exactamente de lo mismo (véase Enríquez, Díaz y Taulé 2015). Repetir lo dicho por el interlocutor es en algunos contextos un ejemplo de cortesía, porque muestra atención e interés por lo dicho por el otro en algunas culturas, pero no en otras. De ahí que la función fática sea objeto de variación intercultural.

¹¹² Estos roles si en la interacción comparten la misma posición en la escala social, sus relaciones serán simétricas. En cambio, si ocupan posiciones diferentes, sus relaciones serán asimétricas.

Para dar cuenta, por tanto, de estas características de estudio se han recogido muestras en las que aparecen hablantes no nativos frente a hablantes nativos y situación real frente a simulada, para abordar las diferencias en la cortesía nativa y no nativa, por un lado, y real o simulada, por otro.

5.6.3. Indicadores de cortesía: Resultados

Una vez detallados los parámetros de estudio de la situación información y compra-venta de billetes de tren en el ámbito de la pragmática de la cortesía, se ha procedido a describir qué mecanismos expresivos se han encontrado entre los que son indicadores de la cortesía para cada categoría de estudio descrito arriba. Es decir, qué casos son los más específicos para expresar cortesía en los parámetros seleccionados de esta situación comunicativa.

5.6.3.1. Fórmulas que reflejan el contacto social

El **bloque a** ó la primera categoría de estudio aborda **la participación social** (*social contact formulae*), que se refiere a los actos de cortesía estandarizados socioculturalmente. Las fórmulas que se utilizan para abrir y cerrar una relación o interacción social. En concreto, los elementos de inicio funcionan como atenuantes de la petición que a continuación se plantea.

En el escenario de estudio las fórmulas que se usan son las siguientes:

- saludos (*hola, buenos días, buenas tardes, ¿qué tal?, hola señora*);
- establecer contacto (*información Renfe, Renfe*) y
- las fórmulas rituales en la despedida (*hasta luego, (muchas) gracias, hasta pronto, que tengas un buen viaje, buen viaje, viaje seguro*)¹¹³.

Estos actos rituales claramente establecidos no varían por tratarse siempre del mismo encuentro, tipo de roles y situación comunicativa. El interlocutor sigue lo establecido por la sociedad. Así pues, se tiene control de la interacción y, en principio, es un (acto) ritual planificable. Debido a estas características, no nos hemos ocupado aquí de

¹¹³ Se incluyen tanto las nativas como las no nativas, por ejemplo, *que tengas un viaje seguro* es no nativa.

estas estrategias de contacto social por no discriminar ni aportar datos relevantes de cortesía al tratarse de fórmulas y de léxico estandarizado, no sujeto a variación en los distintos subcorpus.

En cambio, referentes a los puntos **b) las peticiones de servicio** y **c) el control de la comunicación y de la información** sí tienen mecanismos lingüísticos sujetos a variación que marcan la cortesía y la procedencia de los hablantes (nativos y no nativos). Los rasgos para caracterizar estos dos bloques y los recuentos se realizan sobre la muestra sub6.

5.6.3.2. Peticiones de servicio: cortesía a través de la modalidad y de la indirección

En cuanto al **bloque b**, esto es, **peticiones del servicio** de información y compra-venta de billetes de tren, recogen también diversos elementos que pueden clasificarse como componentes planificables y que rigen el desarrollo de la interacción. Sin embargo, estos actos de habla están condicionados por la cortesía estratégica¹¹⁴, sujeta a variación, puesto que el interlocutor escoge qué mecanismo lingüístico considera más adecuado en el contexto para lograr su objetivo comunicativo y la colaboración del interlocutor sin dañar sus respectivas imágenes sociales.

Recordemos que esta situación, la de petición de información y compra-venta de billetes de tren, implica que entre operador y cliente exista una jerarquía interactiva, que justifica el juego de preguntas entre operador y cliente (como se vio en el apartado 5.3), y es preceptivo para el segundo interlocutor responder para resultar colaborador y respetar el principio de cooperación.

La limitación en este escenario es ceñirse al objetivo prioritario de realizar la transacción. Esto es lo importante: transmitir de manera eficaz la información de este ámbito y obviar otros intercambios sobre otros temas. La prioridad es mantener la distancia social¹¹⁵.

¹¹⁴ En la conversación, los interlocutores se pueden acercar de manera cortés como estrategia para lograr un fin argumentativo (la colaboración del interlocutor, disposición a escuchar, a permitir ataques a su imagen, entre otros) y no solo por ser cortés (seguir una norma de conducta social o cultural, que es la cortesía ritual).

¹¹⁵ En una prueba piloto previa (no incluida en este estudio), alumnos de A2 maximizaban el contacto en la interacción y se preguntaban por aspectos personales. Esto en una situación real es motivo de ruptura por invasión en un contexto público de la esfera privada.

Para llevar a cabo el análisis de las peticiones, se ha centrado el estudio en las formas de las denominadas *Wish/Possibly queries (W/P strategies)* que comportan el uso de modales y son fácilmente localizables, como por ejemplo preguntas del tipo: *quería información.../podría darme...* Son estrategias de cortesía con una intención atenuadora, puesto que la petición o acto de habla directivo se considera un acto invasivo en la zona del otro y los interlocutores no tienen una relación de confianza sino jerárquica. Este rasgo es perceptible en la mayoría de las lenguas que intervienen en este estudio.

Otros mecanismos lingüísticos que funcionan como atenuantes son fórmulas del tipo ([*Me gustaría* (léxico+morfológico)+ INF. (sintáctico)]), en principio, calco de la fórmula de sus lenguas maternas y que no se tiene en cuenta en este estudio. Tampoco se ha tenido en cuenta en este estudio la forma *por favor* porque casi no aparece en la situación de estudio.

Los resultados del análisis de los elementos seleccionados como indicadores de variación se presentan a continuación. Recordamos que se tratan de b.1.) la modalidad expresada léxicamente (modalidad léxica); b.2.) la inflexión verbal como modalidad y b.3.) el uso de las fórmulas indirectas como indicadores formales, todos ellos de la cortesía.

b.1) La modalidad léxica

Para obtener una fotografía de la modalidad léxica (tabla 22) se han extraído todas las ocurrencias de los verbos modales que aparecen en todos los subcorpus de la muestra sub6. Ejemplos de estas formas se muestran en (3), (4) y (5):

(3)

C: *Quiero comprar un billete para ir a... desde Barcelona a Granada este sábado* [FerroviELE-L_trained-multiling_13]

(4)

C: *¿Me podía decir el costo del pasaje a Montpellier?* [InfoTren-N_spontaneous_38]

(5)

O: *¿Preferereres directos o no?* [FerroviELE-L_no trained_No entrenados USA_B1_1]

Verbos modales	No entrenados USA_B1	Entrenados USA_B1	Entrenados multi_B2	Nativos simulados	Nativos espontáneos	Total
Querer	24	20	31	10	8	93
Poder	3	7	18	12	10	50
Preferir	7	3	7	6	0	23
Necesitar	2	4	2	8	0	16
Gustar	0	4	7	0	0	11
Desear	0	0	1	5	0	6
Interesar	0	0	1	0	1	2
Total	36	38	67	41	19	201

TABLA 22. Modalidad léxica para la muestra sub6

Como se puede observar en la tabla 22, solo 3 de los verbos modales se usaron ampliamente: *querer*, *poder* y *preferir*. El primero, *querer*, ha sido el más utilizado, sobre todo, por los no nativos el doble y triple de veces que los hablantes nativos. En relación con *poder*, en cambio, es más usado por los nativos y los no nativos multilingües B2. *Preferir* solo aparece en los subcorpus de los nativos y aprendices en una situación simulada, curiosamente ninguna ocurrencia en los nativos espontáneos. En esta tabla no se ha tenido en cuenta la distinción de rol.

b.2) La inflexión verbal como modalidad

En relación a la cortesía expresada a través de la inflexión verbal, en español hay 3 tiempos diferentes para vehicular la modalidad: el condicional, el imperfecto de indicativo y el imperfecto de subjuntivo. También se utilizan la inflexión personal (*tú tienes* frente a *usted tiene*) y el número (*tiene/tienen*). En la tabla 23 se presentan las formas de inflexión verbal y personal, añadiendo el dato de la distinción de rol para describir la actuación de cliente y operador.

Verbos modales	Inflexión verbal	Inflexión personal	Rol	No entrenados USA_B1	Entrenados USA_B1	Entrenados multi_B2	Nativos simulados	Nativos espontáneos	Total
QUERER	Quiero	1°	C	13	7	15	4	2	41
	Quieres	2°	O	7	4	4	0	0	15
	Quiere	3° (ud.)	O	1	5	9	2	2	19
PODER	Puedo	1°	C	1	4	7	2	1	15
	Puedes	2°	O	0	0	6	0	3	9
	Puede	3° (ud.)	O	0	3	4	3	3	13
PREFERIR	Prefiero	1°	C	4	2	6	0	0	12
	Prefieres	2°	O	1	1	0	1	0	3
	Prefiere	3° (ud.)	O	2	0	1	5	0	8

TABLA 23. *Inflexión verbal como modalidad por roles para la muestra sub6 (I)*

Como se ve en la tabla 23, el cliente emplea la primera persona del presente de indicativo (*quiero, puedo, prefiero*) en la función de *asking for...* o *request a service*, mientras que el operador utiliza la segunda y tercera persona del presente de indicativo (*quieres, quiere, puedes, puede*) para *enquiring about* y *offering*. De modo que es el rol de cliente u operador el que determina el uso de una flexión personal u otra, manteniéndose invariable la opción de flexión temporal (presente de indicativo) en todos los subcorpus. Por ejemplo, la primera persona solo la usan para el rol de cliente (que es el que manifiesta sus necesidades y preferencias de servicio), mientras que la segunda y tercera personas (con los pronombres tú y usted) solo se usan para el rol de operador.

Respecto a las diferencias entre hablantes nativos y no nativos, vemos que el uso del verbo modal *querer* en primera persona (*quiero*) decrece ligeramente con el entrenamiento (cuando lo hay, es decir, en el caso de los entrenados USA_B1). Sin embargo, en los hablantes nativos el empleo de esta forma verbal es siempre menor. Dado que la situación está claramente prefijada y hay una situación jerarquizada que no hay que negociar, no parecen considerar necesario usar recursos corteses en este contexto, especialmente cuando no se trata de una situación simulada sino vivida directamente. En cuanto al uso de la segunda persona (formal e informal) en el modal *querer*, llama la atención los 2 ceros en los nativos para la segunda persona (tú) de singular (*quieres*) del operador y la baja presencia de la forma con usted. Esto es atribuible a que en español se

asocia a la invitación y por lo tanto los nativos la evitan en este escenario (por ejemplo, *¿Quieres un té?/¿Quieres un billete a Zaragoza?*).

En cuanto al modal *preferir*, usado en primera personal singular del presente (*prefiero*) solo lo emplean los no nativos, incluidos los multilingües B2. Llama la atención que los nativos no lo usen nunca. Esto es explicable porque ante un servicio (necesidad) y una situación jerárquica, una vez más, son más directos¹¹⁶.

En relación con la inflexión verbal para el verbo modal *poder*, en la forma de tercera persona cuando se atribuye a la segunda persona formal de cortesía (el *usted* como forma de tratamiento), vemos que solo los B1 no entrenados no la utilizan en absoluto. En el resto de subcorpus de la tabla 23 dicha forma tiene una presencia similar: entre 3 y 4 ocurrencias. La forma de segunda persona informal (con *tú*) no se utiliza en tres de los subcorpus: los dos de B1 y los nativos simulados. Los aprendices de B2 producen el doble que los nativos espontáneos; pero los nativos simulados, en cambio, tampoco lo producen.

Dentro del mismo verbo modal *poder*, las formas *puede/-s* solo aparece en el rol de operador, como se ha dicho arriba, y se trata siempre de ofrecimientos. Lo que se nota es que tanto los multilingües como los nativos espontáneos usan casi indistintamente el *tú* y el *usted* (en un número de ocurrencias parecido en cada caso). En cambio, en los otros subcorpus o no usan este verbo modal (el grupo B1 no entrenado) o solo usan la forma más formal o cortés (el *usted*), posiblemente como fórmula léxica.

En cuanto al modal *poder* en primera persona *puedo* está presente en todos los subcorpus, pero con más peso en los no nativos entrenados.

Por último, en relación con la tabla 23, queremos destacar que lo que resulta caracterizador es el uso de la primera persona por los no nativos con todos los modales. Este rasgo no nativo es atribuible al peso de la estrategia *Wish/Possibly queries (W/P strategies)* de la cortesía interlingüística, que se transfieren tanto en lo sintáctico como en lo léxico desde las lenguas maternas de los aprendices o bien por influencia del inglés (como lengua intermedia conocida por todos) y que afectan a los tres ámbitos léxicos

¹¹⁶ Hemos observado que en situaciones de compra-venta donde la elección del producto implica desplazamiento físico del vendedor por el escenario (a buscar un producto, folleto, etc.) o bien implica alguna dificultad añadida (de tipo físico: subirse a una escalera o alcanzar en un estante, etc.) sí aparece una forma modal más cortés para minimizar ese esfuerzo “adicional” o molestia; pero no en una situación de compra-venta de billetes o ante un ordenador no lo consideran necesario.

estudiados en este apartado de modalidad léxica y flexiva del español para esta situación y escenario concretos. A saber:

- *querer*: *I would like* (C) / *Would you like* (O);
- *poder*: *Possibly/can I* (C);
- *preferir*: *I'd rather* (C).

Otros datos que distinguen los grupos de participantes entre sí son el uso de todos los tiempos verbales previstos para la cortesía en una situación jerárquica como la de este escenario: condicionales, imperfecto de indicativo, imperfecto de subjuntivo, además del presente (que es el tiempo que no esperamos como ejemplo de cortesía). Estas previsiones quedan reflejadas en los datos como se puede apreciar en la tabla 24, que es una tabla de modalidad (II) ampliada, que incluye los mismos tres verbos modales (léxico) de la tabla 23, pero amplía el repertorio flexivo con las formas de condicional y los modos indicativo y subjuntivo.

Verbos modales	Inflexión verbal	Inflexión personal	No entrenados USA_B1	Entrenados USA_B1	Entrenados multi_B2	Nativos simulados	Nativos espontáneos	Total
QUERER	Quiero	1º	13	7	15	4	2	41
	Quería	1º	3	3	1	3	2	12
	Querría	1º	0	0	0	1	0	1
	Quisiera	1º	0	1	2	0	1	4
PODER	Puedo	1º	1	4	7	2	1	15
	Podía	1º	0	0	0	0	0	0
	Podría	1º	2	0	0	3	0	5
PREFERIR	Prefiero	1º	4	2	6	0	0	12
	Prefería	1º/3º (ud.)	0	0	0	0	0	0
	Preferiría	1º	0	0	0	0	0	0
	Prefieres	2º	1	1	0	1	0	3
	Prefiere	3º (ud.)	2	0	1	5	0	8

TABLA 24. *Inflexión verbal como modalidad para la muestra sub6 (II)*

Como se ha mencionado anteriormente, la forma de presente es la que se utilizó mayoritariamente por los no nativos para realizar las peticiones, aunque era el tiempo menos esperado desde la teoría de la cortesía y, como hemos visto, no se cumplía para este escenario cara a cara y con esta inmediatez de satisfacción del servicio. En cambio, los

tiempos característicos para indicar cortesía (imperfectos y condicional) aparecen solo levemente (imperfectos de indicativo de *querer* y de subjuntivo de *querer*), mientras que el condicional de todos los modales no se utiliza en absoluto, como tampoco el imperfecto de indicativo de *poder* ni de *preferir*.

Se confirma lo visto en el tabla 23, si bien resulta llamativo nuevamente que usen la forma *quería* pero no la forma *podía*. Las formas esperadas *quisiera* y *podría* (muy corteses) casi no se utilizan en esta situación (y solo la usan los aprendices entrenados, por lo que podría estar asociado a una fórmula).

b.3) Fórmulas indirectas para preguntar

Los actos de habla indirectos, en concreto las preguntas indirectas¹¹⁷, como vimos al principio, son recursos corteses en la comunicación. En la situación objeto de estudio, peticiones de servicio (información y compra-venta), la indirección se centra en el uso de los recursos formales específicos ligados a las fórmulas o estrategias de *Wish/Possibly queries (W/P strategies)*, en español, por ejemplo, frases introducidas por los modales *Quería.../Podría...*, como se ha apuntado más arriba.

En español las interrogativas del tipo *¿Puede darme un billete para Granada?* se consideran un tipo de fórmula indirecta convencional, que se emplea para expresar una petición cortés. Estas estructuras no se han codificado explícitamente en el corpus *FerrovíELE*, porque también en las lenguas de los informantes no nativos, mayoritariamente de lenguas anglogermánicas, aparecen como fórmulas de cortesía en las peticiones. Por eso, las hemos recuperado a través de la anotación morfológica.

En este apartado, nos hemos fijado en aquellas fórmulas indirectas para preguntar que resultan más complejas para la producción de los no nativos por no tratarse de fórmulas convencionales. Por ejemplo (6):

¹¹⁷ Se entiende, por un lado, por pregunta indirecta aquellas estructuras que utilizan “una fuerza ilocutiva que el significado literal no lo tiene” (Escandell 1995: 50); por otro lado, los actos indirectos pueden ser convencionales, que son aquellos que responden a condiciones formales y de uso marcados por cada sociedad.

(6)

Sé que hay un Intercity a las diez y media de la mañana Barcelona-Madrid [InfoTren-N_spontaneous_2]

En este ejemplo aparece una estructura marcada, pues se incluye un verbo epistémico (*saber*), referido al que habla, que expone una información cuya confirmación espera del interlocutor. Es decir, hay una petición implícita.

Esta estrategia de indirección se usa poco en relación con las indirectas convencionales, si bien cabe destacar que los nativos espontáneos hacen un uso muy significativo de ellas, comparados con el resto de grupos. Véase a propósito de ello la tabla 25. Obsérvese así mismo que los nativos simulados no producen ninguna.

Corpus sub6	fórmulas indirectas		Total
	operador	cliente	
No entrenados USA_B1	2	1	3
Entrenados USA_B1	0	0	0
Nativos simulados	0	0	0
Nativos espontáneos	4	4	8
Entrenados multi_B2	3	0	3
			14

TABLA 25. Fórmulas indirectas para preguntar por rol para la muestra sub6

A la vista de los datos, parece más relevante su uso en el caso del operador que en el del cliente, seguramente como rasgo común de formalidad o cortesía.

En general, coincide con lo observado en los otros recursos de cortesía donde operador, que es el que *ofrece el servicio* usa más formas corteses (verbos modales y la forma de tratamiento usted) que el cliente.

Otro mecanismo para pedir información de forma indirecta es el uso de adverbios interrogativos finales (*¿no?/¿verdad?*) al final de una oración afirmativa. Se trata de un tipo de preguntas en español que se utiliza con otra finalidad que la interrogativa propiamente; en concreto, se usan para confirmar o rechazar una información (etiquetadas en el corpus como *inviting agreement*). Estas preguntas sirven para *confirmar* en el rol de operador y para *aceptar* en el rol de cliente. Un ejemplo (7) de estas lo constituye el siguiente:

(7)

C: *Y la CV eh es en el parte de delante de la tarjeta, ¿no?*

[FerroviELE-L_trained-multiling_4]

5.6.3.3. La cooperación como indicador de cortesía

Tras el análisis de las preguntas para pedir un servicio, a continuación, se trata el **bloque c, el control de la comunicación y de la información**. Este bloque agrupa una serie de procedimientos lingüísticos que se utilizan para gestionar la información. Son las estrategias que no se pueden planificar ni están ritualizadas y, al mismo tiempo, responden a elementos fuera del alcance de la cortesía propiamente dicha. Estos son: c.1) las repeticiones; c.2) resumir o reformular la información recibida y c.3) las vacilaciones y las pausas.

c.1. Las repeticiones

La repetición es un recurso pragmático-discursivo que ayuda al interlocutor a tener control tanto de la información como del proceso de negociación, a la vez que constituye un elemento que asegura la resolución de la tarea. Además de la finalidad práctica (perlocutiva), tiene un efecto empático en el transcurso de la comunicación, más allá de la función fática de los monosílabos e interjecciones, puesto que implica que el que repite información demuestra interés, cortesía, colaboración.

Las repeticiones, codificadas como *check* en el corpus se dividen, por un lado, en repeticiones a corta distancia (*check_short distance*) que funcionan al mismo tiempo como confirmaciones para el rol de operador y como aceptaciones para el rol de cliente; por otro, en repeticiones a larga distancia (*check_long distance*) que remiten a enunciados que están a mayor distancia en turnos de interacción.

Un ejemplo (8) de repetición a corta distancia lo constituye el siguiente:

(8)

C: *No, solo quiero ir sentado.*

O: *Ah sí, **sentado** y...*

[FerroviELE-L_no trained_3]

Si tuviéramos que ilustrar una repetición a larga distancia, formalmente no diferiría del anterior excepto en que estarían alejados ambos enunciados en el texto. En medio se insertaría o intercalaría una (o varias) secuencias con más información y (*sentado, acostado*) que se recogería o se retomaría al final de forma global o por separado.

En la tabla 26 se los resultados de los tipos de repetición obtenidas:

Corpus sub6	repeticiones (<checking>)	
	checking short distance	Checking long distance
No entrenados USA_B1	7	0
Entrenados USA_B1	0	0
Nativos simulados	6	0
Nativos espontáneos	47	8
Entrenados multi_B2	15	2

x 6

TABLA 26. Repeticiones a media distancia y larga distancia para la muestra sub6

Lo primero que llama la atención es que no haya ninguna ocurrencia de *repetición a larga distancia* en tres grupos: B1 (entrenados y sin entrenar) y nativos simulados. Hay que señalar que los no nativos de B2 solo producen dos casos. Por tanto, parece que la repetición a larga distancia es un rasgo característico de una situación espontánea y nativa.

Sin embargo, en el caso de las *repeticiones a corta distancia (check_short distance)* lo primero que llama la atención es que el único grupo que no utiliza este mecanismo es el B1 entrenado. Esto es explicable porque es una estrategia que hemos considerado no entrenable (es propia de la comunicación y de la gestión de información, no característica de una situación concreta). Para los no nativos de nivel B1 (con y sin entrenamiento) este recurso de gestión de información excede su competencia.

La segunda observación llamativa es la proporción de repeticiones de corta distancia entre los nativos espontáneos (entre 3 y 6 veces mayor que el resto de los grupos, tanto nativos como no nativos).

Finalmente, como también hemos indicado en la tabla 26, destacamos el mayor uso de repeticiones a corta distancia frente a las de larga distancia entre todos los grupos, siendo especialmente llamativa la que se da entre los dos últimos grupos de la tabla: los nativos espontáneos y los no nativos B2 multilingües (x6).

En el caso de los nativos simulados el bajo número de repeticiones a corta distancia, podría corresponder al efecto simulación. En efecto, la repetición va ligada a la comprobación de que el interlocutor entendió la petición inicial. Para expresar acuerdo los nativos simulados utilizan recursos lingüísticos de afirmación del tipo: *efectivamente/muy bien, entiendo/por supuesto que sí...* sin emplear la repetición a corta distancia; mientras que en los nativos espontáneos las repeticiones se usan realmente como los auténticos recursos de gestión, puesto que la compra se lleva a cabo de verdad y son más literales en el uso de la información recibida y negociada. Las repeticiones, sobre todo las de larga distancia, son parte de la *negociación* y un recurso para indicar que se está de acuerdo (ambas cosas). Eso explica que se multipliquen por 6 su presencia incluso cuando la lengua no es un problema entre los nativos.

En cuanto los no nativos multilingües es menos claro. En parte es negociación y en parte aclaración o petición de aclaración.

En relación con los nativos que simulan, estos van resolviendo lineal o secuencialmente la tarea, mientras que los nativos espontáneos negocian realmente que el producto se ajuste a sus necesidades. Eso hace que en los simulados no aparezca la repetición a larga distancia, recurso propio de la negociación extensa, sino la negociación simulada turno a turno (no retrospectiva). En cuanto al grupo de aprendices no nativos, solo los de nivel B2 entran, muy minoritariamente, en este tipo de negociación en el transcurso de su simulación.

Todo ello pone de manifiesto, por un lado, la dificultad de esta estrategia, y por otro, su falta de visibilidad en los inputs de clase, materiales didácticos, entre otros recursos.

c.2. Resumir o reformular la información

Relacionado con la estrategia de gestión de información en el transcurso de la interacción, un recurso diferenciador de competencia (meta)lingüística y estratégica es la capacidad de resumir o reformular la información. En nuestro trabajo hemos asociado a esta capacidad una etiqueta funcional o función comunicativa, *resumir*, codificada en el corpus como *synthesising information*. Consiste en exponer a modo de conclusión el

servicio que ha escogido el cliente tras la negociación en el discurso. Es una estrategia que requiere el control de la información y de la organización del significado procedente del interlocutor. Para ello el cliente debe saber gestionar el léxico del ámbito tratado, las anáforas, las elisiones, los sinónimos, los mecanismos de borrado o tematización, y las funciones comunicativas, entre otros recursos. Por lo tanto, se trata de una estrategia muy compleja y muy útil para discriminar entre hablantes nativos y no nativos. Como se ve en la tabla 27, en la muestra sub6 solo aparecen ocurrencias de este mecanismo en el subgrupo de hablantes nativos espontáneos.

Dado que este resultado nos pareció altamente discriminatorio para distinguir los nativos espontáneos (natividad) de los no nativos más competentes, decidimos ampliar la muestra (cuantitativamente) a 16 sujetos. Si observamos la tabla 28 con la muestra sub16, se puede observar que entonces sí hay ocurrencias en B2, confirmándose la hipótesis 5 de que es un rasgo discriminatorio de nivel.

Corpus sub6	No entrenados USA_B1	Entrenados USA_B1	Entrenados multi_B2	Nativos simulados	Nativos espontáneos
	0	0	0	0	2

TABLA 27. Resumir o reformular la información para la muestra sub6

Corpus sub16	Entrenados multi_B2	Nativos espontáneos
	2	4

TABLA 28. Resumir o reformular la información para la muestra sub16

No obstante, el número de ocurrencias es muy bajo (2 y 4, respectivamente).

Como ocurría con las repeticiones, el grupo B1, con o sin entrenamiento, no usa esta estrategia/función comunicativa. En realidad, no es un resultado inesperado, puesto que de lo que no disponen los alumnos en B1 es de una competencia suficiente como para usar una estrategia de meta-representación de la tarea, ni fluidez en el manejo de estrategias de control de la comunicación de nivel por encima de la oración ni de los enunciados breves (resumen de idea); no pueden encabalar ni agrupar contenidos de más de un turno retrospectivo, ni sintetizarlos.

c.3. Las vacilaciones y las pausas

Como mecanismo de colaboración en el curso de la interacción las pausas rellenas y las vacilaciones sirven como indicadores de que se toma o mantiene el turno en la conversación, puesto que los silencios, en algunas culturas, es signo de falta de implicación, colaboración o, simplemente, incómodo.

En la realización de una tarea comunicativa aparecen, también, las vacilaciones como indicadores de dificultades de procesamiento de los aprendices y de los déficits en su interlengua (o de su competencia lingüística en sentido amplio) para llevar a cabo la tarea. En el caso de los nativos esta estrategia sigue otro patrón de uso y manifiesta menos frecuencia de aparición, así como una distribución diferente en la interacción.

Las vacilaciones y las pausas en *FerrovíELE*, en concreto para los subcorpus sub6 se analizaron en el artículo de Enríquez, Díaz y Taulé (2015)¹¹⁸. En los resultados se describieron una serie de diferencias en el empleo de pausas y vacilaciones entre nativos y no nativos, en función de que si las realizaban los nativos indicaban que había dificultad en la toma de decisiones. En cambio, en los diálogos de los no nativos señalaban dificultad en la toma de decisiones sobre la sintaxis, la inflexión verbal y en la gestión de la información. Por tanto, dificultades sobre el código o la lengua y sobre la tarea. Asimismo, el estudio muestra a partir de las vacilaciones y de las pausas que las estructuras interrogativas tienen un alto grado de dificultad en producción oral tanto para los no nativos como para los nativos.

Hasta aquí hemos realizado el análisis de las funciones y las nociones aparecidas en los diálogos de *FerrovíELE* y, en especial, se ha estudiado una selección de funciones que implica el uso de la cortesía en la producción nativa y no nativa en la situación de transacción de información y compra-venta de billetes de tren. En el próximo capítulo se recogen los datos de análisis obtenidos en este capítulo y se presenta una propuesta de inventario básico con ellos.

¹¹⁸ <<http://www.sciencedirect.com/science/article/pii/S1877042815013348>>

CAPÍTULO 6. APLICACIONES E IMPLICACIONES DIDÁCTICAS

Una vez se han abordado los datos del corpus oral *Ferroviele* con un análisis morfológico y funcional-pragmático, en el capítulo 6 se intenta dar cuenta de los resultados con una propuesta sobre los contenidos del escenario comunicativo *Transacción de información y compra-venta de billetes de tren*.

Previamente, se realiza una revisión exhaustiva sobre el tratamiento de las directrices que proporciona el *MCER* (2001) (6.1) y, especialmente, cómo se han concretado esas directrices para el español como LE (6.2): primero, en el *PCIC* (2006) (6.2.1) y, segundo, en los manuales didácticos (6.2.2).

Las directrices del *MCER*, también, se han concretado para el inglés como LE desde el enfoque por escenarios de *EAQUALS* (2010). A continuación, se presenta esta propuesta por escenarios, y también se detallan algunos descriptores del *MCER*, ligados al escenario en estudio (6.3). Por último, se implementa el programa por escenarios de *EAQUALS*, además de los descriptores del *MCER* y los datos del análisis empírico de esta investigación, con la finalidad de ofrecer la propuesta de aplicación de esta tesis (6.4).

6.1 Convergencia y divergencia en el *MCER*

En los primeros capítulos de esta tesis se ha presentado la relación de la propuesta del trabajo por escenarios comunicativos con el estándar proporcionado por el *MCER* (2001), por un lado, y la relación de los niveles de lengua previstos como aplicación de sus descriptores, directa o indirectamente, con el *PCIC* (2006), por otro lado (véase capítulo 1).

Se planteó que el *MCER* ofrece una descripción de las competencias lingüísticas sin vincularlas a ninguna lengua en particular —es neutra lingüísticamente—, recogidas en 13 escalas del capítulo 5 del *MCER*. Esta falta de concreción es voluntaria y tiene como finalidad proporcionar una herramienta que sirva para dar pie a que cada lengua concrete los objetivos comunicativos establecidos con sus correspondientes exponentes lingüísticos.

Asimismo, según Hulstijn (2014:13), el *MCER* establece un “*tertio comparationis*” que permite comparar las competencias para cada una de las lenguas europeas. Conseguir

un asesoramiento y evaluación comparables está detrás de la propuesta del *MCER* de forma más o menos evidente. No obstante, las escalas de competencia recogidas en el documento de referencia requieren de una adaptación para cada lengua en particular. Es decir, una traducción aplicada para cada una de ellas, que no es objetivo del *MCER*. Y es ahí donde empiezan la convergencia de intereses y la divergencia de criterios, objetivos y medios. Señala Hulstjin que la concreción en cada lengua es un proceso que está en marcha y que, hasta la fecha, el resultado diverge por un conjunto de razones no triviales (entre ellas las económicas).

Por ello, se hace necesario que todos los profesionales que participan en el área (responsables de políticas lingüísticas, autores de *syllabus* y materiales, profesores, examinadores y aprendices) contribuyan con su experiencia docente, investigadora, sus datos de aula, exámenes, etc. a ir concretando lo que hay detrás de etiquetas como definir qué se entiende por “la competencia léxica de un aprendiz de nivel B1” o la “competencia gramatical de un aprendiz de nivel B2” o, más concretamente, cuál es el nivel de precisión, control de vocabulario, el control fonológico de cada nivel, para cada una de las lenguas. Estas etiquetas nos exigen un trabajo descriptivo de base empírica, fundamentado en corpus de aprendices y hablantes nativos, que nos permita rellenar lingüísticamente (con funciones, léxico, estructuras sintácticas básicas, etc.) las bandas de descriptores.

Desde 2001, por tanto, ha habido diversos proyectos europeos cuyo objetivo viene siendo “traducir” o aplicar los seis niveles del *MCER* que reflejan los ideales de competencia para que sean reconocibles entre instituciones certificadoras, instituciones académicas, editoriales, entre otras, de forma fiable para las diferentes lenguas europeas. La unidad de política lingüística del Consejo de Europa ha constituido un sitio *web* donde pueden localizarse las descripciones de niveles de referencia: *Reference Levels Descriptors* en inglés¹¹⁹. Estos descriptores, hasta la fecha, están disponibles para diez lenguas europeas, entre ellas el español. La propia organización publicó unas directrices para elaborar estos descriptores (la segunda versión circuló en noviembre de 2005) y declaraba explícitamente que la producción de esas descripciones debe ir avalada oficialmente por una institución gubernamental, de tipo político o científico, como por ejemplo, el

¹¹⁹ <http://www.coe.int/t/dg4/linguistic/DNR_EN.asp>, dirección web de acceso libre.

Ministerio de Educación, un instituto especializado, entre otros, que pueda garantizar su legitimidad y difusión en el contexto educativo. En el caso del español, la institución fue el Instituto Cervantes.

Decoo (2011) valora los resultados de la aplicación del *MCER* usando la mera comparación entre 11 inventarios de diferentes lenguas para el nivel B1. Las diferencias, cuantitativamente, son considerables y oscilan, por ejemplo, para el léxico, entre una horquilla de 1.500 y 7.000 ítems léxicos. Esto llevó a Decoo (2011: 100) a concluir que se había interpretado incorrectamente el objetivo del *MCER*, puesto que el resultado parecía volver a los inventarios de los niveles umbrales de los años 80 y 90, con algunas modificaciones.

En estos momentos, parece coherente cuestionarse que tras más de veinticinco años de trabajo para implementar las especificaciones en listas, primero, y sintetizarlas en el *MCER*, después, parecía que se había entrado en vía muerta. La falta de liderazgo y financiación para acometer una obra de detalle de las características que se han intentado esbozar son una razón de peso. La otra es de tipo político, porque también implica asignar una autoridad institucional para controlar los trabajos y para aprobar o no los resultados de los grupos de trabajo, considerarlos de mejor o peor calidad o colocarlos en una posición susceptible de poner en peligro la garantía de un estándar internacional. La convergencia ha generado, en este ámbito, una divergencia de difícil solución centralizadamente.

No obstante, también han aparecido propuestas provocadas por la intervención de otros agentes como la asociación internacional que evalúa y garantiza la calidad en la enseñanza de lenguas extranjeras, EAQUALS con el British Council. Nos hemos referido ya a EAQUALS (véase capítulo 1). EAQUALS publica *A Core Inventory for General English* (North, B. et al. 2010), un ejemplo de una buena práctica documentada a partir de la aplicación del concepto de escenarios comunicativos —fiel al *MCER*—, y como estructura paraguas para describir la producción de lengua (lo esperado, lo aportado, como *input* oral, escrito, entre otros) con vistas a un asesoramiento y evaluación posterior homologable. A pesar del potencial de la propuesta, su impacto ha sido muy desigual. Las razones de este impacto reducido hay que buscarlas, por un lado, en la falta de investigaciones en L1 para establecer qué habilidades intervienen en cada tarea y para tomarlas como referencia al elaborar las escalas de manera empírica; por el otro, en la falta de corpus de aprendices

para los distintos niveles, en producción oral y escrita, recogidos con la finalidad de ser susceptibles de análisis, descripción y evaluación de las situaciones comunicativas, mencionadas en el capítulo 4 del *MCER*. El uso de corpus es una recomendación que se viene haciendo por diferentes autoridades (Robinson y Ellis 2008; Römer 2008), puesto que es un procedimiento que permite trabajar no solo con la palabra como unidad, sino también con sus combinaciones, N-gramas o multipalabras.

En este sentido, esta tesis y las investigaciones que vienen conduciendo a ella suponen asumir el desafío para el español como LE. Como el objetivo es ambicioso, el trabajo por escenarios comunicativos y la metodología *bottom-up* ayudan a acotarlo.

6.2. La propuesta del *MCER* en el *PCIC* y en los manuales didácticos inspirados en él

6.2.1. El *PCIC*

Para el español, el *PCIC* (2006) ofrece listas de los diferentes componentes para la enseñanza y aprendizaje del español como LE —gramatical, pragmático, discursivo, nocional, cultural y de aprendizaje—, agrupados por nivel de lengua —A1-A2/B1-B2/C1-C2— pero sin relacionarlos por escenarios comunicativos. Es decir, la propuesta consiste en ofrecer los distintos inventarios en forma de listas planas o bidireccionales. A buen seguro la cronología tiene que ver con ello, puesto que la propuesta de EAQUALS es de 2010, algo posterior al *PCIC*.

En este apartado, en el que vamos a llevar a cabo nuestra propuesta de aplicación de los resultados de nuestra investigación ofreciendo un inventario básico y una core-grammar, al estilo de la propuesta mencionada de EAQUALS, nos interesa ver los puntos de contacto (o no) con el *PCIC* y los manuales que lo aplican. Concretamente, nos fijaremos en los componentes del *PCIC* relativos a las funciones comunicativas y a las nociones léxicas, en primer lugar. Asimismo, consideraremos la interpretación que del *PCIC* hacen algunos manuales didácticos de ELE de reciente publicación. En concreto hemos seleccionado tres de los más recientes y más utilizados, según una encuesta a docentes de E/LE¹²⁰.

¹²⁰ Los tres manuales seleccionados son: *Prisma (B2)2004*; *Destino Erasmus 2 (B1-B2) 2008*; *Aula 4 (B1.2)2014*. El primero se usa en toda España por una red de centros muy potente y bien implantada, lo que

En primer lugar, describimos la organización del inventario de las funciones¹²¹ del *PCIC* para todos los niveles (no se incluyen sus exponentes), puesto que la organización es la misma para todos los niveles. En segundo lugar, presentamos la lista de nociones específicas para los niveles B1 y B2 y para el ámbito de *Viajes, alojamiento y transporte*, porque corresponden al escenario comunicativo objeto de nuestro estudio. Y, por último, analizamos los índices que aparecen en las primeras páginas de los tres manuales de E/LE que hemos seleccionado como muestra.

La lista de *Funciones y de Tácticas y estrategias de aprendizaje* del *PCIC* están incluidas dentro del *Componente pragmático* del propio *PCIC* donde se explica que los conocimientos gramaticales y léxico-funcionales van ligados a las situaciones comunicativas y, por tanto, a la pragmática (características de la acción comunicativa y de la interacción social). Además, se asocian las funciones comunicativas con las estrategias pragmáticas. Como hemos dicho, en el *PCIC* los componentes funcionales se presentan en formato de lista plana y el docente o los creadores de materiales son los que deben hacer, a su criterio y desde un modelo *top-down*¹²², la selección de funciones comunicativas necesarias en la situación comunicativa que estén trabajando.

Para ver el inventario de funciones relacionadas con nuestro escenario para todos los niveles del *PCIC*, véase el anexo 9, figura 40. En él, se incluyen solo aquellas funciones estrechamente relacionadas con el escenario de nuestro estudio. Para dicho esquema de funciones, el documento prevé, conforme avanzan los niveles, la progresión del alumno en su competencia lingüística y, por ello, se incrementan tanto el número de funciones como de exponentes y su grado de dificultad. Según el *PCIC* (2007: 177) “desde el nivel A1 hasta el nivel B1, los recursos lingüísticos que se presentan se hallan de algún modo «tasados»”. No obstante, no se especifica el procedimiento empleado para dicha tasación de los exponentes que recoge cada función comunicativa. Por otro lado, apunta el documento que, a partir del nivel B2 hasta C2, los aprendices tienen conocimientos más sólidos. Por ello los exponentes que se presentan son abiertos.

lo hace representativo; el segundo se usa en la Universidad de Barcelona; el último está muy extendido tanto en España como en las redes del exterior.

¹²¹ <http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/indice.htm>

¹²² Ofrece muestras de lengua introspectivas e individuales, orales u escritas.

En efecto, en sus inventarios, el *PCIC* señala que la progresión de un nivel a otro viene determinada por el uso de un mayor número de funciones y por una mayor complejidad de los exponentes. Este aumento de complejidad va acompañado de una mayor complejidad de uso comunicativo y, por tanto, de recursos pragmáticos, si bien estos recursos pragmáticos se recogen en otro inventario localizado en otra sección del *PCIC*, desconectadamente. Veamos en el ejemplo (1) cómo se refleja la progresión de complejidad y de nivel ilustrado con la función *pedir información*. Los exponentes del nivel A se refieren a la categoría *persona* (esto es: pedir información personal o sobre una persona); en el resto de niveles, B y C, no se restringen a ninguna categoría (por lo que los ejemplos son de cualquier ámbito):

(1) «Pedir información»

A1: ¿Quién es? [señalando con un gesto] → INFORMACIÓN PERSONAL

A2: ¿Con quién vives?, ¿Para quién es?

B1: Por favor, ¿puedes decirme dónde está la estación?

B2: Me gustaría saber dónde puedo conseguirlo.

C1: ¿Tienes idea de si hay una oficina de Correos por aquí?

C2: ¿Dónde diablos has dejado el dichoso equipaje?

A la vista de lo expuesto, se hace necesario revisar el planteamiento que hace el *PCIC* de las funciones y de los exponentes. Así pues, a la hora de hacer un plan docente que recoja estas funciones, exponentes, estrategias, etc. a partir de la situación o escenario comunicativo que se vaya a trabajar, esto corre a cargo del docente a partir de su experiencia. Lo que usa es un modelo *top-down* de lengua separado en compartimentos estancos. No obstante, a la vista de nuestro estudio, parece coherente cuestionarse qué funciones y exponentes específicos marcan la diferencia entre un nivel B1 y un nivel B2 y cuáles son transversales y deben incluirse no solo en B1 ni B2 sino también en C1 o C2... e incluso en A2 (en producción).

Por su parte, las nociones específicas o el léxico se organiza de la misma forma que el componente funcional: se ofrece el vocabulario en listas planas y en otro apartado, separado del resto de componentes. A continuación, en la figura 27 y, en el anexo 9, en la figura 41, se

exponen las listas de nociones realizadas por el PCIC¹²³ (2007) para los niveles B1-B2 del ámbito de *Viajes, alojamiento y transporte*, que supone el tema 14, para el escenario comunicativo de petición de información y compra-venta de billetes de tren, de los 20 temas totales que incluye. A modo de ejemplo, únicamente se han incorporado aquellas nociones que están relacionadas con el escenario de nuestro estudio.

14.1. Viajes [v. Saberes y comportamientos socioculturales 1.13.1.]	
B1	B2
<p>14.1.1. Objetos y documentos relacionados con los viajes</p> <ul style="list-style-type: none"> • DNI (Documento Nacional de Identidad), visado • cheques de viaje • equipaje • bolsa ~ de aseo/de viaje • exceso de equipaje • mapa de carreteras, plano turístico, callejero • anular/cambiar ~ un billete • facturar/perder ~ el equipaje/las maletas <p>14.1.2. Tipos de viajes</p> <ul style="list-style-type: none"> • hacer ~ un crucero/un safari viaje ~ de trabajo/de estudios/de negocios/de larga duración • viajar ~ en familia/en grupo • irse/marcharse/salir ~ de puente/de fin de semana • hacer/contratar ~ un viaje organizado • hacer turismo • subir(se) a/bajar(se) de ~ un tren/un autobús/un avión <p>(...)</p>	<p>14.1.1. Objetos y documentos relacionados con los viajes</p> <ul style="list-style-type: none"> • registrar/inspeccionar ~ las maletas/el equipaje, extraviar las maletas <p>14.1.2. Tipos de viajes</p> <ul style="list-style-type: none"> • viaje ~ de novios/de placer/de recreo/oficial • turismo ~ rural/de aventura/cultural/de montaña • ruta, trayecto, itinerario • recorrer/atravesar/cruzar ~ el océano/el país • viajar ~ en clase turista/en clase preferente/en primera (clase) • dar la vuelta al mundo • pasar el verano • montar/organizar ~ un viaje <p>(...)</p>

FIGURA 27. *Inventario de las nociones específicas para el ámbito de Viajes, alojamiento y transporte (I) del PCIC (2007)*

¹²³http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/niveles/09_nociones_especificas_inventario_b1-b2.htm#p143t

Según la propuesta del *PCIC*, el léxico recogido en el inventario del documento de referencia debe ser atendido como una lista abierta según las características de cada situación de enseñanza. De modo que el docente o el creador de materiales será el encargado de ampliar los exponentes léxicos de cada área temática en función de las situaciones comunicativas y las necesidades de los estudiantes. Otra de las recomendaciones del documento es adaptar el léxico en la distribución por nivel. Luego, los exponentes que se ofrecen en el *PCIC* sirven como orientación al usuario.

Por ello, para dar cuenta del léxico que se emplea en el escenario de nuestro estudio, petición de información y compra-veta de billetes de tren, el vocabulario del *PCIC* es, en principio, demasiado general al ofrecer muestras del ámbito de transporte a modo de referencia. Para nuestro escenario se podría tomar de la figura 27: *anular/cambiar un billete; viajar en familia/en grupo; subir(se) a/bajar(se) de un tren; viajar en clase turista/en clase preferente/en primera (clase)* y de la figura 41, en el anexo 9: *vía, andén, ir con retraso, asiento, fila, pasillo, hacer transbordo, RENFE, compartimento*, entre otros. Si bien es cierto se hace difícil discernir qué léxico pertenecería al nivel B1 y cuál al nivel B2.

6.2.2. Los manuales didácticos

Por otro lado, se hace necesario revisar el tratamiento de las funciones y de las nociones léxicas en diferentes manuales que desarrollan un *enfoque nociofuncional o comunicativo*. Los manuales seleccionados se enmarcan tras la publicación del *MCER* (2001) y el *PCIC* (2007). Estos manuales escogidos recogen las lecciones de los índices donde se incluyen los contenidos del ámbito de viajes: *Aula 4* (figura 28) que corresponde al nivel B1.2 del *MCER*; *PRISMA. Avanza* que corresponde al nivel B2 del *MCER* (figura 29) y *Destino Erasmus 2* que corresponde al nivel B1-B2 del *MCER* (figura 30).

3

PÁG. 34

UNIDAD 3 / EL TURISTA ACCIDENTAL

CONTAR ANÉCDOTAS REALES O INVENTADAS

RECURSOS COMUNICATIVOS

- recursos para contar anécdotas
- recursos para mostrar interés al escuchar un relato
- hablar de causas y consecuencias

RECURSOS GRAMATICALES

- algunos conectores para hablar de causas y consecuencias: **como, porque, así que, de modo que**, etc.
- el pretérito pluscuamperfecto de indicativo
- combinar los tiempos del pasado en un relato (pretérito perfecto, pretérito indefinido, pretérito imperfecto, pretérito pluscuamperfecto)

RECURSOS LÉXICOS

- viajes
- tipos de turismo

FIGURA 28. Aula 4. Nueva edición (B1.2). Editorial Difusión (2014)

Unidad 12

Funciones comunicativas

- Dar instrucciones para llevar a cabo una reclamación
- Describir objetos detallando sus características para diferenciarlos de los demás
- Reclamar, expresar indignación, rechazar una propuesta amablemente, expresar finalidad

Contenidos gramaticales

- Oraciones finales
- Transmisión del estilo indirecto al directo

Contenidos léxicos

- Léxico relacionado con los viajes, aeropuerto y equipajes

Contenidos culturales

- Chile
- Literatura: Antonio Muñoz Molina, Luis Sepúlveda
- La polémica entre los términos "castellano" y "español"

FIGURA 29. PRISMA. Avanza (B2). Editorial Edinumen (2004)

UNIDAD 4 RECORRIDOS

46

Textos y pretextos

46

Funciones:
Expresar preferencias por un destino turístico.
Hablar de hábitos y gustos relacionados con los viajes.
Reaccionar ante una opinión.

Léxico:
Medios de transporte.
Turismo.

Textos:
Guías de viaje.
Reportaje radiofónico.

Formas y funciones

50

Funciones:
Ubicar objetos en el espacio.
Describir recorridos.
Expresar recomendaciones.

Contenidos gramaticales:
Preposiciones y locuciones prepositivas con valor de movimiento.
Imperativo afirmativo y negativo.
Uso de los pronombres personales con el imperativo.
Lo mejor es / te recomiendo + que + SUBJUNTIVO
Poder + INFINITIVO

Comunicación escrita

54

Tipología textual:
Descripción de un itinerario.

Mecanismos discursivos:
Desfocalización referencial.
Marcadores espaciales.

FIGURA 30. Destino Erasmus 2. Niveles intermedio y avanzado (B1-B2). Editorial SGEL (2008)

Cada una de estas unidades pertenecientes a diferentes manuales varían a la hora de presentar el índice de los contenidos. Si bien, en principio, la mayoría de estos índices incluyen las macrofunciones que se van a trabajar sin apuntar las funciones

específicas o microfunciones (*Aula 4* y *PRISMA*). En cambio, *Destino Erasmus 2* entremezcla a la hora de hacer explícitas las macrofunciones (describir recorridos) y las microfunciones (expresar recomendaciones).

Otra observación que puede hacerse a estos índices es que, en general, reducen las funciones y los contenidos de aprendizaje que presentan para los diferentes niveles. Al respecto, veamos cómo desarrollan estos manuales sus correspondientes índices:

- a) *Aula 4* para un nivel B1.2 plantea *relatar una anécdota*, si bien previamente el docente debería tener en cuenta la posibilidad de refrescar otra macrofunción, *relatar un viaje*. Ello serviría para relacionar una anécdota del ámbito de viajes con una narración de viajes.
- b) *PRISMA* para un nivel B2 propone unas funciones adecuadas para trabajar la petición de servicios. No obstante, en cuanto a los contenidos gramaticales se infiere que podrían añadir alguno más, por ejemplo, las preguntas indirectas, las preguntas retrospectivas, las reformulaciones, todo ello en relación con los aspectos sociopragmáticos que son necesarios a la hora de comunicar de manera eficaz.
- c) *Destino Erasmus 2* para los niveles B1 y B2 no distingue qué contenidos pertenecen a cada nivel; los presentan de manera general para ambos niveles. Por ello, resulta complicado entender si este índice, que pertenece a la unidad 4, se dirige a un nivel B1 todo él o se tiene que dividir una parte de los contenidos para un B1 y otra para un B2.

El léxico en estos manuales aparece integrado en las macrofunciones, como por ejemplo, en *Destino Erasmus 2*, *hablar de hábitos y gustos relacionados con los viajes*. Mientras que en el índice de otro libro de texto aparece en contenidos separados: por un lado, los recursos comunicativos y, por otro, los recursos léxicos. Este es el caso de *Aula 4* y, además, se supone que la macrofunción “contar anécdotas” está vinculada al ámbito léxico explicitado, los viajes. No obstante, dentro de la unidad se trabajan las anécdotas cubriendo otros ámbitos léxicos.

De todo lo anterior se infiere que no es suficiente con las muestras de descriptores de referencia o de situaciones modelos como pone de manifiesto Hulstjin (2014).

6.3. Los escenarios como salida a la divergencia

Hemos visto en el apartado 6.1, como Decoo (2011) ponía de manifiesto la existencia de diferencias en la aplicación de las directrices del *MCER* a la hora de concretarse en las distintas lenguas y niveles. Esas diferencias le parecían conducir a una divergencia difícil de resolver. La propuesta de trabajo por escenarios puede constituir una salida para volver a converger en los objetivos del *MCER* (2001).

Una vez revisados el documento de referencia del *PCIC* (2006) y algunos manuales para la enseñanza del E/LE, en este apartado se describe más detalladamente la propuesta de EAQUALS (2010) en que hemos basado nuestro estudio, a partir de la implementación de un escenario comunicativo. De ahí obtendremos la propuesta *bottom-up* de contenidos para B1 y B2 (más la *core grammar*) para el español, que estará en sintonía con la que recoge el documento de EAQUALS (2010) para el inglés.

A continuación, presentaremos, en primer lugar, los descriptores de ALTE para ver su relación con el escenario y con la actividad propuesta para la extracción del corpus. Luego, veremos la concreción en niveles a partir de las bandas (cap. 3 del *MCER*). Posteriormente, incluiremos la cortesía como aspecto estratégico (cap. 5 del *MCER*). Por último, vamos a EAQUALS para seguir linealmente su propuesta (2010: 13) y sus directrices para buenas prácticas (2010: 18-19).

Primero, en cuanto la variabilidad esperada, entre los estudiantes no nativos de acuerdo con la previsión de competencia según su nivel de conocimiento de la L2 que el propio *MCER* hace para cada nivel y ALTE (*The Association of Language Testers in Europe*)¹²⁴, sintetizada y especificada para nuestro escenario comunicativo, quedaría recogido en las escalas de los «*Can Do statement*» (Puedo hacer) de *The ALTE Can Do Project, English versión 1992-2002*, p.39. Esta escala (véase la figura 31) corresponde al área sociedad y turismo y se ha seleccionado entre las diferentes escalas de capacidad lingüística que se ofrecen en diferentes áreas (sociedad y turismo, trabajo y estudio):

¹²⁴ <<http://www.alte.org>>

Category A: Social and Tourist

Concern: Day-to-day survival 1
 Activity: Shopping
 Environment: Self-service shops, counter service shops, market place
 Language Skill: Listening/Speaking

- CEF A1** (ALTE Breakthrough) CAN go to a department store or other shop where goods are on display and ask for what (s)he wants.
- CEF A2** (ALTE 1) CAN ask for what is required, if it is something which the shopkeeper can readily understand. CAN exchange basic information, related to place in the queue, etc., with other customers. CAN, where appropriate, bargain in the market place to a minimal extent, with the help of body language (fingers, nod/shake of head, etc.).
- CEF B1** (ALTE 2) CAN go to a counter service shop and ask for most of what (s)he wants. CAN understand where the shopkeeper explains the difference between two or more products all serving the same basic purpose. CAN bargain in the market place where what is purchased is a relatively straightforward item and where the transaction is restricted to the exchange of the item for cash.
- CEF B2** (ALTE 3) CAN ask effectively for refund or exchange of faulty or unwanted goods. CAN bargain for what (s)he wants and reach an agreement.
- CEF C1** (ALTE 4) CAN deal with complex or sensitive transactions, for example the export of an antique.
- CEF C2** (ALTE 5) N/A

FIGURA 31. ALTE Can Do statements

Segundo, en relación con los descriptores ilustrativos que aparecen en el *MCER*, en el capítulo 3, epígrafe 3.5, cuadro 2, presentamos la *escala de autoevaluación*. Estos descriptores reflejan la actuación que se debería poder ejecutar en cada nivel de lengua y centrados por destrezas, a partir de las diferentes actividades comunicativas. Al mismo tiempo, a través de ellas se infieren aspectos de las competencias o del uso de estrategias, de acuerdo con una finalidad de autoevaluación.

Para nuestro estudio, a partir del conjunto de descriptores de dicha escala de autoevaluación seleccionamos las bandas para los niveles B1 y B2 así como para la destreza oral en interacción cara a cara, como se observa en la tabla 29, correspondiente al escenario de petición y compra-venta de billetes de tren.

	B1	B2
Interacción oral	<p>—Sé desenvolverme en casi todas las situaciones que se me presentan cuando viajo donde se habla esa lengua.</p> <p>—Puedo participar espontáneamente en una conversación que trate temas cotidianos de interés personal o que sean pertinentes para la vida diaria (por ejemplo, familia, aficiones, trabajo, viajes y acontecimientos actuales).</p>	<p>—Puedo participar en una conversación con cierta fluidez y espontaneidad, lo que posibilita la comunicación normal con hablantes nativos.</p> <p>—Puedo tomar parte activa en debates desarrollados en situaciones cotidianas explicando y defendiendo mis puntos de vista.</p>

TABLA 29. Cuadro 2. Niveles comunes de referencia: cuadro de autoevaluación. *MCER*, capítulo 3, epígrafe 3.5

Tercero, por lo que se refiere a los mecanismos de cortesía son tratados en el *MCER*, epígrafe 5.2.2.3, dentro del capítulo 5, sección 5.2.2. *La competencia sociolingüística*. En el epígrafe se presentan las normas de cortesía como elementos que “varían de una cultura a otra y son una fuente habitual de malentendidos”. De igual modo se distinguen, por un lado, la cortesía positiva (*expresar gratitud, mostrar interés*, entre otros) y, por otro, la cortesía negativa (órdenes directas, preguntas cortas de confirmación, entre otros), además del uso apropiado de *por favor, gracias* y se identifica el incumplimiento de las normas de cortesía (brusquedad, queja fuerte, impaciencia, entre otros).

Por otro lado, el documento de referencia presenta una única escala de descriptores para la competencia sociolingüística, llamado *Adecuación sociolingüística* y ligado a los diferentes niveles de lengua. Aquí incluimos las bandas para los niveles B1 y B2, en la tabla 30, que son los niveles tratados en nuestra investigación y, a su vez, señalamos en negrita cuando hablan de las normas de cortesía.

B2	—Se expresa con convicción, claridad y cortesía en un registro formal o informal que sea adecuado a la situación y a la persona o personas implicadas.
	—Sigue con cierto esfuerzo el ritmo de los debates e interviene en ellos aunque se hable con rapidez y de forma coloquial.
	—Se relaciona con hablantes nativos sin divertirles o molestarles involuntariamente, y sin exigir de ellos un comportamiento distinto al que tendrían con un hablante nativo. —Se expresa apropiadamente en situaciones diversas y evita errores importantes de formulación.
B1	—Sabe llevar a cabo una gran diversidad de funciones lingüísticas utilizando los exponentes más habituales de esas funciones en un registro neutro.
	—Es consciente de las normas de cortesía más importantes y actúa adecuadamente.
	—Es consciente de las diferencias más significativas que existen entre las costumbres, los usos, las actitudes, los valores y las creencias que prevalecen en la comunidad en cuestión y en la suya propia, y sabe identificar tales diferencias.

TABLA 30. *Adecuación sociolingüística. MCER, capítulo 5, epígrafe 5.2.2*

Las normas de cortesía pueden relacionarse, igualmente, con el conocimiento sociocultural y con los aspectos estratégicos para llevar a cabo la tarea, donde participan elementos funcionales (gramática y léxico) y discursivos (estrategias de interacción).

Por último, recogemos los pasos que sigue la propuesta de EAQUALS (North et al. 2010) para el inglés como LE, debido a los beneficios que podemos obtener por su propuesta de trabajo a partir de escenarios. Esta proporciona un inventario básico o nuclear (*core inventory*). Es decir, los investigadores de EAQUALS identifican los contenidos lingüísticos más relevantes (rasgos lingüísticos, pragmáticos y estratégicos) que integran un escenario comunicativo para los diferentes niveles del aprendizaje, centrados en la enseñanza del inglés como LE.

La propuesta parte de las previas directrices del *MCER* y ofrece las características lingüísticas más comunes a modo de referencia. No obstante, los autores de la propuesta sugieren que podrían incluirse otros contenidos diferentes que se considerasen centrales o nucleares (*core*), tras un análisis de necesidades e intereses de los aprendices. En la figura 32 adjuntamos el inventario básico (*core inventory*), que propone EAQUALS donde indican funciones, gramática, marcadores del discurso, vocabulario y áreas temáticas. En dicha figura, como se puede observar, hemos destacado los niveles de lengua B1 y B2; hemos hecho menos visible el resto de niveles, sin recortar la imagen, para que el lector pueda hacerse una idea global del total del inventario básico, propuesto por EAQUALS.

	A1	A2	B1	B2	C1
Functions	Directions Describing habits and routines Giving personal information Greetings Telling the time Understanding and using numbers Understanding and using prices	Describing habits and routines Describing past experiences Describing people Describing places Describing things Obligation and necessity Requests Suggestions	Checking understanding Describing experiences and events Describing feelings and emotion Describing places Expressing opinions; language of agreeing and disagreeing Initiating and closing conversation Managing interaction (interrupting, changing topic, resuming or continuing)	Critiquing and reviewing Describing experiences Describing feelings and emotions Describing hopes and plans Developing an argument Encouraging and inviting another speaker to continue, come in Expressing abstract ideas Expressing agreement and disagreement Expressing opinions Expressing reaction, e.g. indifference Interacting informally, reacting, expressing interest, sympathy, surprise, etc. Opinion, justification Speculating Taking the initiative in interaction Synthesising, evaluating, glossing info	Conceding a point Critiquing and reviewing constructively Defending a point of view persuasively Developing an argument systematically Emphasising a point, feeling, issue Expressing attitudes and feelings precisely Expressing certainty, probability, doubt Expressing opinions tentatively, hedging Expressing reaction, e.g. indifference Expressing shades of opinion and certainty Responding to counterarguments Speculating and hypothesising about causes, consequences, etc. Synthesising, evaluating and glossing information
Grammar	Adjectives: common and demonstrative Adverbs of frequency Comparatives and superlatives Going to How much/how many and very Common uncountable nouns I'd like Imperatives (+/-) Intensifiers - very basic Modals: can/can't/could/couldn't Past simple of "to be" Past Simple Possessive adjectives Possessive s Prepositions of place Prepositions of time, including in/on/at Present continuous Present simple Pronouns: simple, personal Questions There is/are To be, including question+negatives Verb + ing: like/hate/love	Adjectives - comparative, - use of than and definite article Adjectives - superlative - use of definite article Adverbial phrases of time, place and frequency - including word order Adverbs of frequency Articles - with countable and uncountable nouns Countables and Uncountables: much/many Future Time (will and going to) Gerunds Going to Imperatives Modals - can/could Modals - have to Modals - should Past continuous Past simple Phrasal verbs - common Possessives - use of 's, 's' Prepositional phrases (place, time and movement) Prepositions of time: on/in/at Present continuous Present continuous for future Present perfect Questions Verb + ing/infinitive: like/want-would like Wh-questions in past Zero and 1st conditional	Adverbs Broader range of intensifiers such as too, enough Comparatives and superlatives Complex question tags Conditionals, 2nd and 3rd Connecting words expressing cause and effect, contrast, etc. Future continuous Modals - must/can't deduction Modals - might, may, will, probably Modals - should have/might have/etc Modals: must/have to Past continuous Past perfect Past simple Past tense responses Phrasal verbs, extended Present perfect continuous Present perfect/past simple Reported speech (range of tenses) Simple passive Wh- questions in the past Will and going to, for prediction	Adjectives and adverbs Future continuous Future perfect Future perfect continuous Mixed conditionals Modals - can't have, needn't have Modals of deduction and speculation Narrative tenses Passives Past perfect Past perfect continuous Phrasal verbs, extended Relative clauses Reported speech Will and going to, for prediction Wish Would expressing habits in the past	Futures (revision) Inversion with negative adverbials Mixed conditionals in past, present and future Modals in the past Narrative tenses for experience, incl. passive Passive forms, all Phrasal verbs, especially splitting Wish/if only regrets
Discourse Markers	Connecting words, and, but, because	Linkers: sequential - past time	Connecting words expressing cause and effect, contrast, etc Linkers: sequential past time	Connecting words expressing cause and effect, contrast, etc. Discourse markers to structure formal speech Linkers: although, in spite of, despite Linkers: sequential - past time - subsequently	Linking devices, logical markers Markers to structure and signpost formal and informal speech and writing
Vocabulary	Food and drink Nationalities and countries Personal information Things in the town, shops and shopping Verbs - basic	Adjectives: personality, description, feelings Food and drink Things in the town, shops and shopping Travel and services	Collocation Colloquial language Things in the town, shops and shopping Travel and services	Collocation Colloquial language	Approximating (vague language) Collocation Colloquial language Differentiated use of vocabulary Eliminating false friends Formal and informal registers Idiomatic expressions

FIGURA 32. Core Inventory para el inglés, British Council y EAQUALS (North, B. et al. 2010: 10-11)

En efecto, el *inventario básico* está destinado a ser utilizado como una propuesta de referencia y se ofrece un resumen de los contenidos lingüísticos más destacados para la enseñanza y aprendizaje en cada nivel de lengua. En el caso que se quisiera tener más detalle sobre algún elemento o rasgo concreto, las directrices de EAQUALS derivan al lector a las concreciones de los niveles A2 (*Waystage*), B1 (*Threshold*) y B2 (*Vantage Level*) (van Ek y Trim 2001).

Por otro lado, los investigadores de EAQUALS perfilan el significado del término *core* en su programa y para ellos “*core is also, by definition, inadequate to meet the needs of any particular group of learners*” (North et al. 2010: 20). Esto es, se infiere que la propuesta debería servir como referencia —se vuelve a insistir en dicha idea—, y subrayan que habría que ir más allá y analizar las necesidades de los estudiantes con el objetivo de cubrirlas y de

este modo, construir los contenidos de enseñanza y aprendizaje. Para ello, proporcionan una guía de qué *se puede* y *no se puede hacer* (North et al. 2010: 20).

Respecto a lo que *se puede hacer*:

- identificar las necesidades de los estudiantes relacionadas con situaciones de la vida real y derivar estas necesidades a los contenidos de enseñanza desde escenarios de situaciones reales;
- proponer bandas propias del *MCER* basadas en escenarios;
- complementar algunos puntos del inventario propuesto por EAQUALS con información que sabes que los aprendices necesitan para cubrir un escenario comunicativo;
- adaptar algunos exponentes concretos al escenario que se trabaje (nombre de lugares, actividades, etc.) y
- siempre que sea posible, dar a los aprendices los exponentes lingüísticos a partir de un escenario comunicativo.

Respecto a lo que *no se puede hacer*:

- enseñar contenidos poco productivos del lenguaje solo porque se incluyen en un *syllabus*;
- revestir las actividades pedagógicas como escenarios reales;
- pensar que solo tienes que enseñar lo que está en el inventario básico o que debes enseñar todos los contenidos que están en este;
- añadir listas de exponentes sin sentido crítico en tu *syllabus* y
- hacer que los estudiantes aprendan simples listas de exponentes.

Las buenas prácticas de EAQUALS se desarrollan en los apéndices de la propuesta que se presenta para el inglés como LE. Estos apéndices caracterizadores son:

Apéndice B: modelo de escenarios y parrillas «*On holiday*»;

Apéndice C: modelo de tipos de texto para *Spoken sources*;

Apéndice D: modelo de mapeo (*mapping*) de contenidos lingüísticos funcionales-gramaticales por niveles, centrados en los niveles A1, B1, B2 y C1;

Apéndice E: exponentes lingüísticos concretados en niveles de lengua (A2, B1, B2...).

A continuación, recogemos cada uno de ellos:

Apéndice B: modelo de escenarios y parrillas «*On holiday*»

Se proporciona un ejemplo de escenario, “*On holiday*” (de vacaciones), en la figura 33 con su material correspondiente. Los ejes de descripción de un escenario para EAQUALS

y el British Council es, por un lado, el esquema general y, por otro, la implementación del escenario. El esquema general lo componen i) “*Can Do statement*” (Puedo hacer); ii) los criterios de fluidez: precisión, coherencia e interacción y iii) las competencias lingüísticas. La implementación del escenario está formada por los exponentes que aparecen en este escenario con sus respectivas funciones y las propuestas de actividades para el profesor.

En el próximo apartado (6.4) para la confección del inventario básico de la transacción de información y compra-venta de billetes de tren vamos a desarrollar, en principio, el esquema general. En este estudio no se establece la sección de la implementación y recordamos que el inventario no se adscribe a ninguna metodología de enseñanza.

Appendix B: SCENARIO: ON HOLIDAY

A1

DOMAIN	CONTEXT	TASKS	ACTIVITIES	TEXTS
Personal Public	Places: airports, hotels, shops, restaurants, public transport, entertainment. People: officials, hotel/shop and other personnel, other travellers/holidaymakers	Getting information to book a trip Travelling Contracting and using services (e.g. hire and pick up car; check in at hotel) Finding your way Socialising	Listening to announcements and instructions Reading instructions/for information and orientation Spoken Production Spoken Interaction Written Production	Sustained monologue; description Transactions, conversations Forms, postcards Notices, instructions, brochures

LEVEL	A1	COMPETENCES
CAN-DO'S*	Can understand questions and instructions addressed carefully and slowly to him/her, and follow short, simple directions. Can understand numbers, prices and times. Can ask for and give personal information (address, telephone number, nationality, age, family and hobbies). Can handle numbers, quantities, cost and time. Can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.	STRATEGIC Establish contact with people using simple words, phrases and gestures Ask someone to repeat what they said more slowly Invite help with gesture when you don't know the word PRAGMATIC Functional Understanding and using numbers Understanding and using prices Greetings Giving personal information Understanding simple directions and instructions Discourse Connecting words (and, but, because) LINGUISTIC Grammatical Simple verb forms Questions Present simple I'd like Prepositions and prepositional phrases Lexical Nationalities and countries Personal information Things in the town, shops and shopping
CRITERIA*		
INTERACTION	Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	
COHERENCE	Can link words or groups of words with very basic linear connectors such as "and" or "then".	
RANGE	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	

EXPONENT	COMPETENCE
There are 500 people in our village How much does the room cost? 45 Euros per night	Understanding and using numbers Understanding and using prices
Hi John, how are you today? My name is Carlos I am from the north of China I live in Beirut.	Greetings Giving personal information
The hotel is on the left Go to the end of the street and turn right Where is the supermarket? It's straight ahead She lives in Switzerland and goes skiing a lot I don't like Indian food but I like Chinese	Understanding simple directions and instructions Prepositions and prepositional phrases Connecting words (and, but, because)
Are you French? No I'm not Is she from Egypt? What is your name? How much does it cost? Pedro is Spanish but he works in France She's married and has three children I am 26 years old, single and I work in a bank No I'm not tired France is a wonderful country I am a psychology student at the moment	Questions Nationalities and countries Personal information Simple verb forms Present simple
Where is the supermarket/train station? How much does this cost? Does this bus go to the town centre? I want to buy a phone I'd like a cup of coffee/I'd like to go home	Things in the town, shops and shopping Travel and services vocabulary I'd like
	Establish contact with people using simple words, phrases and gestures Ask someone to repeat what they said more slowly Invite help with gesture when you don't know the word

*From the CEFR.

This is a sequence of assessed tasks related to a sequence of tasks associated with being on holiday. The precise order of activities is left to the teacher or learner.

Susan Sheehan

FIGURA 33. Appendix B: Scenario: on holiday, British Council y EAQUALS (North et al. 2010: 26)

Apéndice C: modelo de tipos de texto para Spoken sources:

EAQUALS ofrece un esquema a modo de mapeo (*mapping*) de tipos de texto, escritos u orales. En la figura 34 se presenta una parte específica del contenido total de tipos de texto orales y, en concreto, la que hace referencia a nuestro escenario.

	A1	A2	B1	B2	C1
Spoken Sources					
interlocutor					
third party interaction		topic	main points	modified	complex
discussions and debates			everyday	modified	complex
technical discussions					complex

FIGURA 34. Appendix C: Mapping Text Types, British Council y EAQUALS (North et al. 2010: 36-37)

Apéndice D: modelo de mapeo (*mapping*) de contenidos lingüísticos funcionales-gramaticales por niveles, centrados en los niveles A1, B1, B2 y C1:

Se incluye una parte del mapeo (*mapping*) del contenido lingüístico, como se puede observar en la figura 35: funciones, nociones, funciones del discurso y gramática.

Appendix D: Mapping Language Content

Produced from main project sources (descriptor analysis, analysis of CEFR-based syllabi, coursebooks, examination content)

Functions/Notions	A1	A2	B1	B2	C1
Numbers					
Prices					
Telling the time					
Directions					
Greetings					
Giving personal information					
Describing habits and routines					
Describing people					
Describing things					
Requests					
Suggestions					
Advice					
Invitations					
Offers					
Arrangements/ -ing to meet people					
Obligation and necessity					
Describing places					
Describing past experiences and storytelling					
Describing feelings, emotion, attitudes					precisely
Describing hopes and plans					
Giving precise information					
Expressing abstract ideas					
Expressing certainty, probability, doubt					
Generalising and qualifying					
Synthesising, evaluating, glossing info					
Speculating, and hypothesising					hypothesising
Expressing opinions					teritatively
Expressing shades of opinion					
Expressing agreement/disagreement					
Expressing reaction, e.g. indifference					
Critiquing and reviewing					
Developing an argument					systematically
Conceding a point					
Emphasising a point, feeling, issue					
Defending a point of view					persuasively
Responding to counter-arguments					

Discourse Functions	A1	A2	B1	B2	C1
Initiating and closing conversation -					
Checking understanding ->					
Managing interaction (interrupting, + changing topic, resuming or continuing)					
Taking the initiative in interaction					
Encouraging and inviting another speaker to continue, come in					
Interacting informally, reacting, expressing interest, sympathy, surprise					

Modals: Can	A1	A2	B1	B2	C1
Can/can't					
Can/could					
Modals: Possibility	A1	A2	B1	B2	C1
Might, may					
Possibly, probably, perhaps					
Might, may, will, probably					
Must/can't (deduction)					
Modals: Obligation and Necessity	A1	A2	B1	B2	C1
Must/mustn't					
Have to					
Must/have to					
Should					
Ought to					
Need to/needn't					
Modals: Past	A1	A2	B1	B2	C1
Should have/might have, etc.					
Can't have, needn't have					

FIGURA 35. Appendix D: Scenario: Mapping Language Content, British Council y EAQUALS (North et al. 2010:

Apéndice E: exponentes lingüísticos concretados en niveles de lengua (A2, B1, B2...):

En este apéndice se recoge un listado de exponentes para dotar de contenido lingüístico al escenario seleccionado. En la figura 36 hemos incluido algunos de los exponentes lingüísticos junto a sus funciones y nociones, correspondientes al nivel B1, a modo de ejemplo.

B1			
2 Functions/notions 19 Describing places <ul style="list-style-type: none">■ Cairo is the capital city of Egypt. It is on the banks of the River Nile. It has a population of more than ten million people. Cairo has a rich history. The famous pyramids and the sphinx are located just outside the city. 20 Describing past experiences and storytelling <ul style="list-style-type: none">■ Last year we went to Thailand for our holidays. We visited many interesting places. I went scuba-diving while my boyfriend went on an elephant ride. We also tried lots of different kinds of food. We had a great time. 21 Describing feelings, emotions, attitudes <ul style="list-style-type: none">■ He felt a little nervous about the exam.■ I'm fed up with this British food.■ We should all use public transport as much as possible.■ I'm sorry to hear that. 29 Expressing opinions <ul style="list-style-type: none">■ I think England will win the World Cup.■ I don't think he is old enough to get married.■ In my opinion, it's too expensive. 31 Expressing agreement/disagreement <ul style="list-style-type: none">■ I think so too.■ You're right■ Exactly!■ Yes, I agree.■ I think you are absolutely right.■ So do/am I.■ Neither do I.	<ul style="list-style-type: none">■ Well, actually ...■ Well, not really.■ I'm sorry but I think you're wrong.■ I see what you mean but ... 33 Talking about films and books <ul style="list-style-type: none">■ <i>Meryl Streep was brilliant.</i>■ <i>It was difficult to follow.</i>■ <i>It was set in Chicago.</i>■ <i>It was about a woman who went around the world.</i>■ <i>If you like action movies you will like this one.</i>■ <i>It had a happy ending.</i>■ <i>I think you should read this book.</i>■ <i>I liked this book because ...</i> 39 Discourse functions 40 Initiating and closing conversation Initiating <ul style="list-style-type: none">■ Hi! My name's Paula. What's your name?■ Excuse me, can I talk to you for a minute?■ Excuse me, please. Have you got a minute?■ Excuse me, please. I wonder if you could help me.■ Let me introduce myself.■ Guess what?■ You will never believe what I saw yesterday. Closing <ul style="list-style-type: none">■ It's been nice talking to you. Bye.■ I'm sorry. I've got to go now.■ Must go – see you later.■ See you later. Take care.	41 Checking understanding From speaker's point of view <ul style="list-style-type: none">■ Is that clear?■ Do you follow me?■ Do you know what I mean?■ Do you understand? From listener's point of view <ul style="list-style-type: none">■ I'm sorry, did you say "..."?■ Is this what you are saying? ...■ I'm not sure I understand. Are you saying that ...?■ Do you mean ...? 42 Managing interaction (interrupting, changing topic, resuming or continuing) Interrupting <ul style="list-style-type: none">■ Sorry to interrupt you but ...■ I have a question.■ Could I interrupt here?■ Do you mind if I say something?■ Could I just say something?■ Sorry, I just wanted to say ... Changing the topic <ul style="list-style-type: none">■ Anyway, ...■ By the way, there's something else I wanted to tell you. Resuming a conversation <ul style="list-style-type: none">■ Anyway, ...■ Anyway, what was I saying?■ What were we talking about?■ To get back to what I was saying ... Continuing <ul style="list-style-type: none">■ Anyway, ...■ So, as I was saying ...■ Okay, ...	46 Discourse Markers 48 Linkers: sequential – past time (later) <ul style="list-style-type: none">■ He finished the email and then went out for a while.■ Later, he looked at it again, to see if he had missed anything important.■ After that, he changed the text a little.■ Finally he spell-checked it and sent it. 49 Connecting words expressing cause and effect, contrast, etc. <ul style="list-style-type: none">■ On the other hand, we could stay at home and watch television.■ However, this depends on the number of people you've invited.■ Therefore, it is cheaper to take the bus. 52 Markers to structure informal spoken discourse <ul style="list-style-type: none">■ <i>Right.</i>■ <i>Really?</i>■ <i>Well, anyway ...</i>■ <i>Oh I know.</i>■ <i>Yes, I suppose so.</i>■ <i>I know how you feel.</i>■ <i>You know, I don't like her either.</i> 59 Questions 61 Wh-questions in the past <ul style="list-style-type: none">■ Who did you see at the party?■ How long have you been studying English?■ Why did you get the tattoo?■ What happened then?■ What have you been doing since you left school? 62 Complex question tags <ul style="list-style-type: none">■ He hasn't come home yet, has he?■ He built the house himself, didn't he?

FIGURA 36. Appendix E: Exponents for Language Content (B1), British Council y EAQUALS (North et al. 2010: 52-53)

6.4. La propuesta de aplicación de esta tesis

De acuerdo con lo expuesto en los descriptores de ALTE, ofreciendo que *pueden hacer* lingüísticamente los estudiantes en el área social del transporte, además, con lo planteado en los descriptores del *MCER* para describir la interacción oral y la cortesía en las diferentes etapas de aprendizaje y, por último, con lo descrito en la propuesta por escenarios de EAQUALS, a continuación, presentamos nuestra propuesta de

implementación para el escenario petición de información y compra-venta de billetes de tren para E/LE.

Para ello, aportamos los datos analizados en los capítulos 4 y 5 con el objetivo de desarrollar nuestra propuesta, que sigue la estructura de clasificación y ordenación de los apéndices de EAQUALS, descritos en el apartado anterior.

Vamos a implementar el Apéndice B y el Apéndice D de EAQUALS, uniendo las estructuras de ambos, con la finalidad de ejemplificar el escenario de estudio de manera detallada. Asimismo, se implementa el Apéndice C y el Apéndice E, todo de forma unificada.

Veamos, a continuación, la propuesta del escenario transacción de información y compra-venta de billetes de tren para E/LE para los niveles B1 y B2.

DOMINIO	CONTEXTO	TAREAS	ACTIVIDADES	TEXTOS
-Público	-Lugar: estación de tren -Institución: Renfe	-Pedir información -Reservar o comprar	-Producción oral -Interacción oral	-Conversaciones transacciones (B1-B2) -Información horarios de trenes (folleto o documento pdf) -Dinero -Tarjeta de crédito
-Personal	-Personas: operador/ cliente	-Utilizar transporte		-Conversaciones informales sobre hechos y experiencias (anécdotas) (B2)

TABLA 31. *Parámetros generales del escenario transacción de información y compra-venta de billetes de tren para E/LE (B1-B2)*

En efecto, los parámetros anteriores se concretan en los descriptores «*Can Do statement*» (Puedo hacer), centrados en el nivel B1 y del nivel B2 (tabla 32). A este respecto, se ha tenido en cuenta para construir esta escala la presentada por ALTE (figura 31. *ALTE Can Do statements*):

B1	Puede entender si el operador explica las características y las diferencias entre dos tipos de trenes. Puede negociar un mínimo el cliente a nivel oral la información sobre horarios, tipos de billetes, trayectos, etc. para realizar la transacción que consiste en intercambiar un producto por dinero o pedir información.
-----------	---

B2	Puede gestionar el operador la información sobre precios, trayectos, alternativas de tipos de trenes, etc. Puede pedir efectivamente el cliente e incluso discutir si es necesario las ventajas y los inconvenientes. Puede negociar el cliente para conseguir lo que quiere y llegar a un acuerdo.
-----------	---

TABLA 32. *Can Do statements* (Puedo hacer) en el escenario *transacción de información y compra-venta de billetes de tren para E/LE (B1-B2)*

COMPETENCIA

1. ESTRATÉGICA

Para la ejecución del escenario, previamente, el estudiante *planifica e identifica* el vocabulario del ámbito, relacionado con aspectos socioculturales (sistema ferroviario español).

En este escenario de estudio los interlocutores de los dos niveles (B1-B2) *cooperan* ante las preguntas y las respuestas que dependen del rol cliente u operador. *Per se* son colaborativos ambos roles y en consecuencia son corteses. Así pues, si hay alguna cuestión que no entienden hay peticiones de aclaración.

Por otro lado, si el estudiante tiene un nivel B1 emplea mayor grado de *estrategias de compensación*¹²⁵ y a medida que progresa en su grado de dominio, es decir, un nivel B2, estas disminuyen.

2. PRAGMÁTICA

2.1. Funcional

Las combinaciones más frecuentes de funciones para cualquier transacción son las *funciones de la matriz nuclear* que se emplean, básicamente, en un nivel B1. A continuación, para la transacción específica de petición de información y compra-venta de billetes de tren resultan realizables las *funciones específicas* a partir de un nivel B1.

Asimismo, se pueden desarrollar otras funciones, las de la *matriz expandida*, que indican un mayor manejo de la lengua y que se aproximan al uso que hacen los hablantes

¹²⁵ Las estrategias de compensación consisten en utilizar otras maneras de expresión para lograr el objetivo comunicativo por parte del estudiante, por ejemplo, parafrasear, explicar mediante ejemplos, usar un término de otro ámbito léxico para que el receptor lo puede entender, entre otros.

nativos. También, algunas de ellas pueden venir dadas por la interacción y el estilo personal. Por ello, se suelen esperar a partir de un nivel B2. Todas ellas se recogen, a continuación, en la tabla 33:

	B1	B2
Funciones de la matriz nuclear		
Accepting		
Closing conversation		
Confirming		
Describing plans		
Enquiring about		
Expressing feelings/ knowledge/ preference		
Giving precise information		
Initiating conversation		
Request		
Socialising		
Funciones específicas de la situación de transacción petición y compra-ventade billetes		
Request a service: information/ making a reservation/ buying a ticket		
Advice		
Checking availability		
Counting		
Enquiring about information		
Giving information		
Giving product printing		
Obligation and necessity		
Paying		
Suggestions		
Funciones de la matriz expandida		
Agreeing		
Apologising for making a mistake		
Asking for confirmation or denial		
Check-task		
Correcting		
Counting		
Filler hesitation pause		
Encouraging and inviting another speaker to continue		
Expressing degrees of probability, doubt		
Expressing opinion		
Expressing reaction		

Expressing reason, cause, excuse		
Interrupting		
Inviting agreement		
Making a mistake		
Offer		
Refusing		
Repeating what one has said/other interlocutor has said		
Seeking identification		
Socialising: asking for repetition		
Socialising: asking someone to wait/ Socialising: asking someone to wait: emphatic offer		
Socialising: asking whether one is heard and understood		
Socialising: supplying a word, repetition or expression		
Spelling		
Synthetising information		

TABLA 33. *Funciones comunicativas en el escenario transacción de información y compra-venta de billetes de tren para E/LE por niveles (B1-B2)*

En relación con algunas funciones de la matriz nuclear y de la matriz expandida, en concreto las señaladas en negrita, son funciones comunicativas que, a su vez, se emplean como funciones para gestionar la información. Estas se detallan en el siguiente apartado junto a otros elementos.

2.2. Discursiva

En este escenario se requiere el uso de diálogos donde intervienen dos interlocutores: un operador y un cliente. Sin embargo, la dificultad del conocimiento sociocultural del ámbito de trenes y su respectivo léxico hace complicada la tarea para el rol del operador. Por ello, el alumno es capaz de simular al cliente en un nivel B1 y tiene dificultades para realizar el rol del operador. En cambio, en un nivel B2 los aprendices son capaces de desenvolverse con soltura en ambos roles.

Otras características generales que debe tener el discurso dialógico son:

- i. alternancia de turno;
- ii. turnos cortos por la situación de transacción, y
- iii. aprovechar los momentos de transición para tomar el turno.

Respecto al punto i, la alternancia de turno es propio de ambos niveles: B1 y B2.

En cuanto el punto ii, la intervención del interlocutor es más extensa en el nivel B1, mientras que en el nivel B2 los turnos empiezan a ser más breves y, al mismo tiempo, aparecen más transiciones en el robo de turno (rasgo iii).

Asimismo, se deberían tener en cuenta qué funciones del discurso deben ser capaces de usar en los niveles B1 y B2. En estas, además, se incluye la variedad nativa, que equivaldría al nivel C2. Dentro de estas funciones del discurso, se señalan las siguientes:

- las fórmulas corteses para abrir y cerrar el diálogo (aspectos sociopragmáticos) *buenos días, hola señora, que tengas un buen viaje, buen viaje* aparecen, en general, en un nivel B1 si bien a partir de una nivel B2 sus exponentes son más variados y hay mayor presencia en los turnos de interacción;
- la función *animar e invitar al interlocutor a continuar (encouraging and inviting another speaker to continue)* debe enseñarse a partir del nivel B2;
- las repeticiones a corta distancia (*checking short distance*) y a larga distancia (*checking long distance*): primero, las repeticiones a corta distancia son un recurso que excede la competencia del nivel B1 si bien no para el nivel B2; segundo, las repeticiones a larga distancia son un rasgo de una situación espontánea y nativa, y
- la función *resumir o reformular la información a modo de conclusión (synthesising information)* que es una estrategia compleja, realizable a partir del nivel B2.

Hay otros elementos que ayudan a gestionar la tarea. Entre ellos tenemos las anáforas y las fechas y horarios. Primero, para las anáforas se consideran los determinantes y los pronombres (*este tren, el tren*). Segundo, para las fechas y horarios se tiene en cuenta el vocabulario específico de la tarea y si se utiliza adecuadamente refleja un conocimiento de mundo (fechas, horas, SPrep —*el tren de.../la salida.../el precio de...*— y como complemento —*el de ida...*—). Ambos elementos son apropiados a partir del nivel B2.

A modo de resumen, en la tabla 34 se recogen los diversos recursos previstos en el componente discursivo que deben ser capaces de hacer los niveles B1-B2 y, además, en este componente se incluye el nivel C2 que representa la variante nativa:

	B1	B2	C2
Interlocutores			
Cliente			
Operador			
Características generales			
Alternancia de turno			
Turnos cortos por la situación de transacción			
Aprovechar los momentos de transición para tomar el turno			
Funciones del discurso			
Iniciar y cerrar el diálogo: formulas corteses	-	+	
Animar e invitar al interlocutor a continuar			
Repeticiones a corta distancia			
Repeticiones a larga distancia			
Resumir a modo de conclusión			
Otros elementos para gestionar			
Anáforas			
Fechas y horas			

TABLA 34. Recursos discursivos en el escenario transacción de información y compra-venta de billetes de tren para E/LE por niveles de lengua (B1-B2)

3. LINGÜÍSTICA

3.1. Gramática

Los aprendices son capaces de usar el contenido gramatical que, a continuación, se ofrece, listado por:

- los elementos lingüísticos que comparten B1, B2 y C2 (variante nativa) (a) y
- el contenido gramatical previsto para cada nivel: B1 (b), B2 (c) y C2 (variante nativa) (d).

a) Niveles B1, B2 y C2 (variante nativa):

CATEGORÍA GRAMATICAL:

- **Discursivo:** otro, también, este, el, del, uno, otro, perfecto
- **Adverbio:** aquí, bien, bueno, solo, sí (afirmativo), no, muy, más, mañana, mucho*
- **Adjetivo:** posible, perfecto, primero, individual*
- **Verbo:** estar, haber, comprar, costar, ir, llegar, poder, querer, salir, ser, tener, viajar, incluir*, necesitar*
- **Preposición:** a, de, del, por, para, en, con, hasta*
- **Conjunción:** o, y, si, que, pues, pero, a, porque
- **Pronombre:** yo, tú, usted*, él* [pron. Interrog. → qué, cuánto, cómo]
- **Determinante:** mi*

*Lemas que aparecen en 4 subcorpus de los 5 analizados

b) Nivel B1:

CATEGORÍA GRAMATICAL:

- **Discursivo:** sin embargo
- **Adverbio y locución adverbial:** antes, de inmediato, finalmente, obligatoriamente
- **Adjetivo:** español, no directo, pequeño, suficiente,
- **Verbo:** confirmar, crear
- **Preposición:** Ø
- **Conjunción:** cuando
- **Pronombre:** Ø
- **Determinante:** Ø

c) Nivel B2:

CATEGORÍA GRAMATICAL:

- **Discursivo:** es decir
- **Adverbio y locución adverbial:** al final, adónde, ahora mismo, directamente, económico, en cuanto al, en línea, en seguida, en total, nunca, obviamente, por cierto, por último, un poco
- **Adjetivo:** grande, pesado, propio, sencillo, tercero, turístico
- **Verbo y locución:** anotar, asistir, caducar, conocer, consultar, continuar, dar igual, empezar, entrar, explicar, ir bien, limitar, molestar, parar, partir, permitir, proponer, recomendar, registrar, relajar, sentir, (pagar) en efectivo
- **Preposición:** junto
- **Conjunción:** donde
- **Pronombre:** Ø
- **Determinante:** Ø

d) Nivel C2 (variante nativa):

CATEGORÍA GRAMATICAL:

- **Discursivo:** alguno, ninguno, o sea
- **Adverbio y locución adverbial:** aproximadamente, concretamente, demasiado, de momento, efectivamente, en metálico, en punto, exacto/exactamente, igual, mientras que, nada más, precisamente, pronto, siempre, sinceramente, temprano
- **Adjetivo:** conforme, cómodo, comprobar, correcto, diario, feliz, internacional, mayor, metálico, nocturno, obvio, oportuno
- **Verbo:** abonar, climatizar, conectar, descansar, deber, disponer, disfrutar, enlazar, entender, escoger, exigir, leer, parecer, perdonar, realizar, recibir, recoger, referir, retirar, sacar, suponer, tasar, trabajar, variar, usar
- **Preposición:** (locución prepositiva) a partir de, por encima de
- **Conjunción y locuciones:** así como, así que, desde luego, para que, puesto que, sí que, ya que
- **Pronombre:** ambos, nosotros
- **Determinante:** Ø

Por otro lado, se debería prestar especial atención a los rasgos lingüísticos que detallamos y describimos, a continuación, con el objetivo de acercar la producción de los aprendices a las características de la variante nativa en este escenario comunicativo.

- i. el uso de algunos verbos modales, en concreto, *querer, poder, preferir*;
- ii. las peticiones: las interrogativas indirectas convencionales (*¿Podría informarme...*) vs. las interrogativas directas (*Dame...*). Los grados de cortesía;
- iii. las fórmulas indirectas para preguntar (*Sé que hay...*);
- iv. la petición de información de forma indirecta con el uso de adverbios interrogativos finales (*¿no?/¿verdad?*);
- v. el uso de los pronombres *tú/usted* y
- vi. las relativas (*el tren que va, el que sale*).

En relación con (i), se debería insistir a los estudiantes que los hablantes nativos en este escenario emplean menos el verbo *querer* que los aprendices. Estos últimos utilizan la primera persona del singular (*quiero*) en el rol de cliente; en cambio, los nativos no parecen considerar necesario usar dicha forma, puesto que la situación está prefijada y no hay que negociar. De este modo, los nativos son más directos y no suelen utilizar *quiero*. Lo mismo ocurre con el verbo *preferir*, usado en primera persona del singular (*prefiero*), en principio, por los aprendices. Por otro lado, la segunda persona del singular del *querer* (*quieres*) en el rol de cliente tampoco la emplean los hablantes nativos, porque se vincula a la invitación y la situación no es lo que requiere.

En cuanto al modal *poder*, por una lado, la forma *puedo* la utilizan lindistintamente los nativos y los aprendices en el rol de cliente, y, por otro lado, la forma *puede/s* solo aparece en el rol de operador y se utiliza para ofrecer, también, por hablantes nativos y no nativos.

En relación con (ii), los aprendices utilizan en esta situación mayoritariamente la forma de presente para realizar las peticiones, por tanto debería introducirse otros tiempos verbales para señalar la cortesía en las peticiones: imperfecto de indicativo (nivel B1) y de subjuntivo (nivel B2), además del condicional (nivel B1). En suma, se debería empezar a hacer conscientes sobre los diferentes grados de cortesía en español a partir de un nivel B1.

En relación con (iv), las fórmulas indirectas para preguntar por su complejidad podrían enseñarse de manera eficaz en un nivel B2.

En relación con (v), aunque las formas *tú* y *usted* se introducen en las clases de E/LE a partir del nivel inicial, lo cierto es que los aprendices utilizan la flexión personal (*tú/usted*) de los

diferentes verbos modales de forma lexicalizada. Por ello, se debería hacer más instrucción en los niveles B1 y B2.

En relación con (vi), las estructuras relativas como *el tren que va* no las producen los estudiantes, así que se deberían introducir a partir de un nivel B2 por su complejidad cognitiva.

3.2. Léxica

Se presenta el léxico básico distribuido a partir de los diferentes niveles y relacionado con el ámbito de tren y transporte, con el objetivo de entender el contenido expresado en las funciones comunicativas de este escenario. A continuación, se ofrece:

- el léxico que comparten los niveles B1, B2 y C2 (variante nativa) (a) y
- el vocabulario previsto para cada nivel concreto: B1 (b), B2 (c) y C2 (variante nativa) (d).

a) Niveles B1, B2 y C2 (variante nativa):

NOMBRES: AVE*, billete, descuento, hora, ida, información* (no utilizan los C2), noche*, opción*, precio, tiquet, tren, turista, viaje, vuelta

*Lemas que aparecen en 4 subcorpus de los 5 analizados

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: saludos (hola); agradecimiento (gracias/muchas gracias/por favor); afirmación o negación (sí/no/perfecto)

INTERJECCIONES: ah

b) Nivel B1:

NOMBRES: alta velocidad española, espera, fin, línea, seguridad, *tiquet*, transición

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: Ø

INTERJECCIONES: ay, puah

c) Nivel B2:

NOMBRES: acceso, cabina-litera, caducidad, calidad, cambio, coche-restaurante, final, equipaje, idea, montón, NIE, pasaporte, PIN, posibilidad, preferencia, presupuesto, pueblo, relación, restaurante-cafetería, titular, vez

EXPRESIONES FIJAS: Ø

FÓRMULAS DE CORTESÍA: buen día, chao

INTERJECCIONES: aha, qué guay

d) Nivel C2 (variante nativa):

NOMBRES: año, cercanías, combinación, compra, confianza, confirmación, continuación, cosa, costo, dato, dirección, enchufe, expreso, folleto, forma, Intercity, litera, locución, mail, merienda, mostrador, ordenador, pago, pasaje, pregunta, prisa, pico (y pico), kilómetro, rato, regreso, reunión, sitio, taquilla, trabajo, tramo, transbordo, urgencia, ventaja, vía, visa, wifi, zona

FÓRMULAS DE CORTESÍA: buenas, de nada, es un placer, hasta ahora, información de renfe, información renfe, muchísimas gracias, perdonar (perdone), saludo

INTERJECCIONES: aja, etcétera, hombre, hum, uf

EXPRESIONES FIJAS: vamos, hombre

ONOMATOPEYAS: ring-ring

CAPÍTULO 7. CONCLUSIONES

Al inicio de esta tesis planteamos la necesidad que hay de disponer de modelos empíricos en los que basar descripciones y propuestas en la enseñanza de lenguas extranjeras con orientación comunicativa y pragmático-discursiva (hipótesis 1).

Esta es la línea que aplican los grandes consorcios de enseñanza y acreditación de lenguas extranjeras, como el British Council, por ejemplo, y que nosotros hemos tomado como referencia para la propuesta desarrollada a partir de EAQUALS (capítulo 6)

Por ello consideramos que la aplicación y la explotación de escenarios comunicativos va más allá de las propuestas del *MCER* e incluso del *PCIC*, tanto desde un punto de vista cronológico como empírico.

Con esta implementación de la propuesta de EAQUALS (2010) al español se ha conseguido el doble objetivo que planteamos al inicio de esta tesis: primero, obtener corpus lingüísticos a partir de escenarios, en concreto, transacción de información y compra-venta de billetes de tren para sostener las descripciones en la enseñanza y el aprendizaje del E/LE, y, segundo, obtener los exponentes lingüísticos correspondientes al repertorio funcional vinculado al escenario, graduados en tres niveles: B1, B2 y C2 (los nativos como referencia para C2). Estos dos objetivos constituían las hipótesis 1 y 2 inicialmente planteadas.

Hemos justificado la metodología de extracción de datos adecuados para estos objetivos en el capítulo 3 y, como resultado de su aplicación, se han obtenido, mediante un proceso de transcripción y anotación, los subcorpus de nativos y no nativos de *FerrovíELE*. Estos corpus constituyen, a su vez, una importante aportación al campo de la enseñanza y aprendizaje del E/LE y que servirán directamente para inspirar otros trabajos similares o ampliaciones de este, así como para posteriores desarrollos de aplicaciones didácticas.

Es importante señalar las extensiones de los resultados y aplicaciones posibles — algunas no previstas inicialmente— como por ejemplo:

- 1) Obtención de un repertorio específico de elementos caracterizadores/identificadores de nativos frente a no nativos (capítulos 4 y 5).

- 2) Obtención de repertorios de elementos caracterizadores de la espontaneidad en la interacción dialógica ante una situación comunicativa. Esto se ha verificado de dos formas distintas: a) en las diferencias halladas entre hablantes nativos espontáneos frente a nativos simulados (a partir de subcorpus distintos) y b) en las diferencias entre hablantes no nativos (aprendices) con o sin entrenamiento. (capítulos 4 y 5).
- 3) Obtención de un repertorio de C2, es decir, de datos para los que los nativos sirven como referencia —desde lo experto a lo no experto y desde lo real a lo simulado— con los que asesorar, calibrar y evaluar de forma realista y con base empírica la producción oral de aprendices para cada uno de los dos roles (cliente y operador) de la situación estudiada. Esto resulta particularmente útil para revisar las pruebas de evaluación al uso, las prácticas de aula, los materiales didácticos (*inputs* de modelos, guiones) e incluso los exámenes orales existentes desde una nueva perspectiva.
- 4) Vinculada a la anterior, el estudio ha permitido la obtención de criterios pragmáticos (cortesía, pausas, interrupciones, peticiones directas, estructuras indirectas, negación) y criterios estratégico-cognitivos (reformulaciones, repeticiones) que permiten calibrar (*assess*) y evaluar (*evaluate*) la producción lingüística de los aprendices por su proximidad a la espontaneidad nativa (como es la presencia de rasgos compartidos) o el exceso de dificultad de la tarea (falta de competencia y de recursos estratégicos), que se ponen de manifiesto en el exceso de pausas, vacilaciones, errores por sobrecarga cognitiva.

Como ejemplo de transferencia de resultados, consecuencia de todo lo anterior y, como se desprende de los análisis derivados de los capítulos 4, 5 y 6 creemos que debería haber una consideración de todos estos aspectos en las propuestas didácticas de E/LE. A saber:

- Refinar los repertorios funcionales cuando se proponen escenarios (o parte de ellos) en los materiales para el aula de E/LE;
- Refinar las descripciones de contenidos no solo lingüísticos sino pragmáticos y socioculturales propios de los escenarios y los roles, desde los primeros niveles, A2, B1.

- Refinar los requisitos y exigencias para la evaluación de las producciones orales de los aprendices, evitando exigir en el aula “borrados” de aspectos propios de la oralidad en la producción espontánea, como las vacilaciones o improvisaciones presentes en los nativos (simulados o no). Es decir, no exigir a los aprendices ni esperar como profesores más de lo que hacemos los nativos, estando atentos a los parámetros descritos en el capítulo 6.

Por último, hemos concretado los hallazgos de la investigación empírica desarrollada en la tesis en una propuesta detallada que se ha presentado en el capítulo 6. En ella se aplica por primera vez el programa y las buenas prácticas recomendadas por EAQUALS (2010) para el español. Se ha cubierto, con los datos de *FerrovíELE*, no solo la provisión de formas y estructuras, es decir, lemas y funciones en listas, sino que se han presentado de acuerdo con los formatos y categorías de EAQUALS en sus apéndices. Esto es:

- a) Apéndice A. *Core-inventory* de funciones, gramática, marcadores y vocabulario circunscritos al tema/dominio del transporte en tren, en concreto, en situaciones de petición de información y compra-venta de billetes de tren.
- b) Apéndice B. Ejemplificación de un escenario, en concreto, de una interacción específica: la situación de ventanilla con su material correspondiente.
- c) Apéndice C. Correspondencia (o mapeo) de tipos de texto para la situación o el escenario, que se ha concretado en la interacción oral en el transcurso de la compra-venta y petición de información.
- d) Apéndice D. Correspondencia (o mapeo) de contenido lingüístico: análisis de descriptores y *syllabus* para contenidos didácticos y de exámenes.
- e) Apéndice E. Listado de exponentes para dotar de contenido lingüístico (en español) al escenario seleccionado.

Todo ello se halla detallado en la propuesta final del capítulo 6.

En síntesis, a lo largo de esta tesis hemos sostenido argumentativa y empíricamente la importancia de disponer de corpus para el diseño de material didáctico en E/LE, así como

para la evaluación de la competencia de los aprendices. Mediante la aplicación de un modelo *bottom-up* a nuestro estudio, hemos obtenido una fotografía viva de la competencia en desarrollo desde el nivel B1 hasta B2 (y C2 para los nativos), así como también una clasificación de recursos propios de cada nivel de competencia para una determinada situación y rol (cliente y operador en la petición de información y compra-venta de billetes de tren) siempre en interacción oral. Gracias a haber delimitado tan estrecha o restrictivamente el objeto de estudio hemos podido poner de manifiesto el potencial de esta forma de analizar la lengua y de trasladarla a la didáctica y la evaluación. Con ello contribuimos a reforzar, empíricamente, los postulados de autores como Thornbury (2011), Larsen-Freeman (2001), Römer (2004, 2008), North et al. (2010), Ellis (1992), muy en boga en los foros de metodología del inglés como LE (TOEFL, TESOL, entre otros) y de evaluación y de acreditaciones lingüísticas (North et al. 2010; Alderson 2005, bandas ALTE). Esto es, proporcionamos también base empírica con la que nutrir los descriptores de comportamientos esperables (reales), de estrategias implementables (reales y por nivel), de recursos lingüísticos (reales y por nivel) para una situación de un escenario muy frecuente en los exámenes (*tests*) y materiales publicados, que eran la asignatura pendiente después del *MCER* para cada una de las lenguas de la Comunidad Europea (aunque no solo). Esperamos haber contribuido modestamente a ello.

7.1. Contribuciones de esta tesis

En este estudio se han realizado las siguientes aportaciones

1. Propuesta de metodología basada en corpus para la descripción de escenarios comunicativos en L2, implementada y ejemplificada para el español en el escenario concreto de interacción oral dialógica en el transcurso de una transacción de información y compra-venta de billetes de tren, destinada a la descripción y evaluación de producciones no nativas y de muestras nativas espontáneas y no espontáneas. Esta metodología es ampliable a otros escenarios comunicativos del E/LE como a otras lenguas extranjeras.
2. Creación de un nuevo recurso lingüístico, el corpus oral de aprendizaje *FerrovieLE* (nativo y no nativo), transcrito ortográficamente y anotado morfológica y pragmático-funcional con su correspondiente guía de anotación, reutilizable.

3. Propuesta de modelización del escenario elegido con tres submodelos: información, compra-venta y reserva, en función del propósito comunicativo predominante en la interacción. Reelaboración de los listados nociofuncionales disponibles ajustados a B1 y B2, a partir de datos empíricos.
4. Propuesta de caracterización de la producción no nativa por niveles a partir de dicha descripción de corte empírico.
5. Descripción de la variación nativa en función del grado de experimentalidad de la situación (datos espontáneos frente a simulados), que permiten caracterizar el llamado nivel C2 del *MCER*.
6. Propuesta de inventario básico y de *core grammar* de acuerdo con EAQUALS (2011) para todas y cada una de las especificaciones de su recomendación de buenas prácticas (apéndices A-E del documento de EAQUALS) circunscrito a la interacción oral dialogada del escenario seleccionado.
7. Revisión detallada y actualizada de los corpus disponibles para español LE y su metodología, tanto de los diseñados para recogida en contexto de aprendizaje o de los utilizados como recurso didáctico.

7.2. Líneas futuras

En cuanto a las líneas futuras, como se desprende de lo que acabamos de decir, está:

- Ampliar el estudio enfocando en otros componentes del escenario y en otros escenarios, lo que implicará contemplar nuevas funciones, léxico y aspectos sociopragmáticos, además de estratégicos, de la comunicación y que, sin duda, permitirán poner a prueba y ampliar el potencial de la metodología y herramientas empleadas.
- Ampliar el estudio a otros niveles de competencia de los aprendices (A2, C1).
- Aumentar el número de sujetos de la muestra, las L1s por nivel y clasificar los fenómenos por L1s para su uso en recursos didácticos, formación de profesores, etc.
- Profundizar en el análisis funcional y discursivo por niveles y L1 a fin de poder activar la reflexión metalingüística y sobre el uso allá donde no llegan las gramáticas

de uso que no son sensibles a este tipo de análisis discursivo basado en corpus (colocaciones en diálogos, sensibles al turno, al rol, etc.).

REFERENCIAS BIBLIOGRÁFICAS

- ALBELDA, M. (2010): "Rentabilidad de los corpus discursivos en la didáctica de lenguas extranjeras" en *Actas del XXI Congreso Internacional de ASELE. Del texto a la lengua. La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, Salamanca. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0083.pdf>, 05/15].
- ALBELDA, M. y FERNÁNDEZ, M.J. (2006): "La enseñanza de los registros en ELE. Una aplicación al español coloquial" en *MarcoELE*, núm. 3. [En: <<http://marcoele.com/la-ensenanza-de-los-registros-linguisticos-en-ele-una-aplicacion-a-la-conversacion-coloquial/>>, 05/15].
- ALBELDA, M. y FERNÁNDEZ, M.J. (2008): *La enseñanza de la conversación coloquial*, Madrid: Arco/Libros.
- ALDERSON, J. C. (2005): *Diagnosing foreign language proficiency: The interface between learning and assessment*, New York, NY: Continuum.
- ALONSO PÉREZ-ÁVILA, E. (2007): "El corpus lingüístico en la didáctica del léxico del español como LE" en *Boletín de Asele*, núm. 37, pp. 11-32.
- ALONSO PÉREZ-ÁVILA, E. (s.f.): "Petición en español. Aproximación a la pragmática de interlengua". [En: <<http://www.cisi.unito.it/artifara/rivista5/testi/peticiones.asp>>, 07/15].
- ALONSO RAMOS, M.; WANNER, L.; VINCZE, O.; CASAMAYOR G.; VÁZQUEZ, N.; MOSQUEIRA, E. y PRIETO, S. (2010): "Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora" en *7th International Conference on Language Resources and Evaluation (LREC)*, Malta: La Valetta, pp. 3209-3214.
- ALTE (1992-2002): *The ALTE Can Do Project*, English versión, p.39 [En: <<http://www.cambridgeenglish.org/images/28906-alte-can-do-document.pdf>>, 09/2015].
- ANDREU, M.; ASTOR, A.; BOQUERA, M.; MACDONALD, P.; MONTERO, B. y PÉREZ, C. (2010): "Analysing EFL learner output in the MiLC project: An error it's*, but which tag?" en M. C. Campoy; B. Belles-Fortunó y M.L. Gea-Valor (eds.), *Corpus-Based Approaches to English Language Teaching*, London: Continuum.
- AUSTIN, J. (1962): *Cómo hacer cosas con las palabras: Palabras y acciones (How to Do Things with Words)*, 4ª ed., Barcelona: Paidós, 2004.
- AZORÍN, D. y JIMÉNEZ RUÍZ, J.L. (1997): *Corpus oral de la variedad juvenil universitaria del español hablado en Alicante*, Alicante: Instituto de Cultura Juan Gil-Albert.
- AZORÍN, D. (coord.) (2002): *Alicante corpus oral del español*, Alicante: Universidad de Alicante, edición electrónica en CD-ROM.

- BAILINI, S. (2013): "SCIL: A Spanish Corpus of Italian Learners" en *Procedia-Social and Behavioral Sciences. Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013)*, núm. 95, pp. 542-549. [En: <<http://www.sciencedirect.com/science/article/pii/S1877042813042006>>, 04/15].
- BALLESTEROS MARTÍN, F. (2001): "La cortesía española frente a la cortesía inglesa. Estudio pragmlingüístico de las exhortaciones impositivas" en *Estudios Ingleses de la Universidad Complutense*, núm. 9, pp. 171-207.
- BARALO, M. (1998): *Teorías de adquisición de lenguas extranjeras y su aplicación a la enseñanza del español*, Fundación Antonio de Nebrija, Madrid: Colección EXEELE.
- BARDOVI-HARLIG, K., y HARTFORD, B. S. (1990): "Congruence in native and nonnative conversations: Status balance in the academic advising session" en *Language Learning*, núm. 40, pp. 467-501.
- BARDOVI-HARLIG, K. y HARTFORD, B. (eds.) (1997): *Beyond Methods. Components of Second Language Teacher Education*, Boston: McGraw-Hill.
- BARDOVI-HARLIG, K. (1999): "The interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics" en *Language learning*, núm. 49, pp. 677-713.
- BARRAS, C.; GEO, E.; WU, Z. y LIBERMAN, M. (2001): "Transcriber: Development and Use of a Tool for Assisting Speech Corpora Production", *Speech Communication*, núm. 33.
- BARRON, A. (2005): "Variational pragmatics in the foreign language classroom" en *System* 33(3), pp. 519-536.
- BARROS, P. (ed.) (2012): *COGILA: Español oral conversacional: corpus y guía didáctica*, Grupo de Investigación de Lingüística Aplicada, Universidad de Granada, Granada: EUG.
- BARROS, M^a J.; MORALES, J. y BARROS, P. (2015): "Enseñanza de español real: corpus y guía didáctica" en *Cuadernos Canela*, núm. 26, pp. 93-106. [En: <<http://www.canela.org.es/cuadernoscanela/>>, 07/15].
- BELLO, A. (1981): *Gramática de la lengua castellana destinada al uso de los americanos*, edición crítica de Ramón Trujillo, Santa Cruz de Tenerife: Instituto Universitario de Lingüística Andrés Bello, Litografía A. Romero, S.A.
- BELLVER, V. y SAN ANDRÉS, M^a C. (2010): *¡Viva el vocabulario! Nivel intermedio B1-B2*, Madrid: enClave-ELE, pp. 68-69.

- BERNARDINI, S. (2004): "Corpora in the classroom. An overview and some reflections on future developments" en J. M. Sinclair(ed.), *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp. 15-36.
- BLUM-KULKA, S. y OLSHTAIN (1984): "Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP)" en *Applied Linguistics*, 5(3), pp. 196-213.
- BLUM-KULKA, S. (1987): "Indirectness and Politeness in Requests: Same or Different?" en *Journal of Pragmatics* 11, North-Holland: Elsevier Science Publishers B.V., pp. 131-146.
- BLUM-KULKA, S.; HOUSE, J. y KASPER, G. (eds.) (1989): *Cross-Cultural Pragmatics: Requests and Apologies*, Norwood (NJ): Ablex.
- BONAFONTE, A. y LLEIDA, E. (1999): *Corpus InfoTren: Persona. Corpus oral de llamadas a un servicio de información sobre trenes de viajeros*, Universitat Politècnica de Catalunya y Universidad de Zaragoza.
- BONAFONTE, A.; AIBAR, P.; CASTELL, N.; LLEIDA, E.; MARINÑO, J.B.; SANCHIS, E.; y TORRES, M.I. (2000): "Desarrollo de un sistema de diálogo oral en dominios restringidos" en T. López (ed.), *I Jornadas en Tecnología del Habla*, Sevilla.
- BOTTA, M. (2002): "Krashen y su teoría del filtro afectivo: ansiedad, autoestima y motivación en la adquisición de segunda lengua" en *Cuadernos Cervantes*, núm. 41, año VIII, Madrid.
- BRAVO, D. y BRIZ, A. (eds.) (2004): *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, Barcelona: Ariel Lingüística.
- BRIZ, A. (2002): *El español coloquial en la clase de E/LE. Un recorrido a través de los textos*, Madrid: SGEL.
- BRIZ, A. (2008): "Aportaciones del análisis del discurso oral" en J. Sánchez Lobato y I. Santos Gargallo (eds.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid: SGEL, pp. 219-242.
- BRIZ, A.; HIDALGO, A.; PADILLA, X.; PONS, S.; RUIZ, L.; SANMARTÍN, J.; ALBELDA, M. (2002): "La transcripción de la lengua hablada: El sistema del Grupo Val.Es.Co." en *Español Actual: Revista de español vivo*, núm. 77-78, pp. 57-86.
- BRIZ, A. y ALBELDA, M. (2009): "Estado actual de los corpus de lengua española hablada y escrita: I+D" en *El español en el mundo. Anuario del Instituto Cervantes 2009*. [En: <http://cvc.cervantes.es/lengua/anuario/anuario_09/briz_albeida/p01.htm>, 04/15].
- BROWN, P y LEVINSON, S. (1978, 1987): *Politeness: Some universals in Language Use*, Cambridge: CUP.

- BULLOCK, B. E. y TORIBIO, A. J. (2013): *The Spanish in Texas Corpus Project*. COERLL, The University of Texas: Austin. [En: <<http://spanishintexas.org>>, 07/15].
- BURNARD, L y SPERBERG-McQUEEN, C.M. (1995): *TEI Lite: Una introducción / Text Encoding for Interchange*, documento nº: TEI U 5, junio. Traducido por M. Sánchez Quero, Biblioteca Virtual Miguel de Cervantes, Universidad de Alicante, 2001. [En: <http://www.tei-c.org/Vault/P4/Lite/teiu5_sp.html>, 05/15].
- BUYSE, K. y GONZÁLEZ, E. (2012): “El corpus de aprendices Aprescrilov y su utilidad para la didáctica de ELE en la Bélgica multilingüe” en *XXIII Congreso Internacional de la ASELE. Plurilingüismo y enseñanza de ELE en contextos multiculturales*, Girona. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/23/23_0025.pdf>, 04/15].
- CABEDO, A. Y PONS, S. (eds.): *Corpus Val.Es.Co 2.0*. [En:<<http://valesco.es>>, 04/15].
- CAMPILLOS, L. (2007): “Adaptación del corpus C-ORAL-ROM a la enseñanza de español para extranjeros” en *redEle*, núm. 7, segundo semestre. [En: <<http://www.mecd.gob.es/redele/Biblioteca-Virtual/2007/memoriaMaster/2-Trimestre/CAMPILLOS-L.html>>, 05/2015].
- CAMPILLOS, L. (2014a): *Oral expression in Spanish as a foreign language: interlanguage and corpus-based error analysis*. PhD dissertation: Universidad Autónoma de Madrid en *Procesamiento del Lenguaje Natural*, núm. 52, marzo de 2014, pp. 91-92. [En: <<http://www.llf.uam.es/leonardo/pdf/SEPLN-2014-resumentesis.pdf>>, 05/15].
- CAMPILLOS, L. (2014b): “Análisis de errores pragmático-discursivos en un corpus oral de español con lengua extranjera” en *Círculo de lingüística aplicada a la comunicación*, núm. 58, pp. 23-59. [En: <<http://revistas.ucm.es/index.php/CLAC/article/viewFile/45469/42753>>,05/15].
- CAMPILLOS, L.; GONZALO, P. y MORENO, A. (2007): “El corpus C-ORAL-ROM en la enseñanza de ELE” en *Actas del XVII congreso internacional de ASELE. Las destrezas orales en la enseñanza del español*, Logroño. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/17/17_1115.pdf>, 05/15].
- CAMPILLOS LLANOS, L.; GOZALO GÓMEZ, P.; GUIRAO, J. M^a. y MORENO SANDOVAL, A. (2010): *Español oral en contexto*, Madrid: UAM ediciones.
- CARBONERO, P., CORTÉS, L. y BAÑÓN, A. (2005): “Corpus para el estudio de las hablas andaluzas I. El corpus del habla de Sevilla y el corpus del habla de Almería”, *Oralia*, núm. 8, pp. 161-187.
- CARRANZA, M (2011): “La corrección automática de la pronunciación (EPAO). Diseño de un corpus oral de muestras de lengua de estudiantes japoneses” en *XII Congreso Internacional de la ASELE. La red y sus aplicaciones en la enseñanza-aprendizaje del*

- español como lengua extranjera*, Valladolid. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/asele_xxii.htm>, 05/15].
- CARRANZA, M. (2013): “Errores y dificultades específicas en la adquisición de la pronunciación del español LE por hablantes de japonés y propuestas de corrección” en *Asociación GIDE* <<http://gidetokio.curhost.com>>. [En: <<http://gidetokio.curhost.com/archivos/201210MarioCarranza.pdf>>, 04/15].
- CESTERO, A. M^a; PENADÉS, I.; BLANCO, A.; CAMARGO, L. y SIMÓN, J. (2002): “Corpus para el análisis de errores de aprendices de E/LE (CORANE)” en Gimeno, A. M. (ed.), *Actas del XII Congreso Internacional de ASELE. Tecnologías de la información y de las comunicaciones en la enseñanza de E/LE*, Valencia, pp. 527-534. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/12/12_0527.pdf>, 04/15].
- CHIAPELLO, S.; GONZÁLEZ, C. y MURA, A. (2014): “Teletándem: actividad interactiva del discente de I/LE y su percepción mediante una encuesta de autoevaluación” en *XII Jornadas de Redes de Investigación en Docencia Universitaria. El reconocimiento docente: innovar e investigar con criterios de calidad*, Alicante: Universidad de Alicante. [En: <<http://rua.ua.es/dspace/handle/10045/42163?locale=en>>, 04/15].
- CHOMSKY, N. (1965): *Syntactic Structures*, Mouton classic (2002): Berlin/New York.
- CLInt: *Clinical Interview*, Barcelona: Universitat de Barcelona. [En: <<http://clic.ub.edu/corpus/clint>>, 04/15].
- CORDULINO DA SILVA, M. (2010): “Práctica de enseñanza del español coloquial en aula E/LE para brasileños” en *Revista Foro de Profesores de E/LE*, vol. 6. [En: <<http://foroele.es/revista/index.php/foroele/article/view/172>>; 05/15].
- CORPAS, J.; GARMENDIA, A. y SORIANO, C. (2014): *Aula 4. Nueva edición*, Barcelona: Difusión.
- COUNCIL OF EUROPE (2001): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- CRESTI, E. y MONEGLIA, M. (2005): *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: John Benjamins.
- CRUZ, M. (2012): *Lingüística de corpus y enseñanza del español como 2/L*, Madrid: Arco/Libros.
- DAVIES, M. (2002): “Un corpus anotado de 1000.000.000 palabras del español histórico moderno” en *SEPLN 2002* (Sociedad Española para el Procesamiento del Lenguaje Natural), pp. 21-27.

- DECOO, W. (2011): *Systemization in Foreign Language Teaching: monitoring content progression*, New York: Routledge.
- DÍAZ, L. (2007): *Interlengua española. Estudio de casos* (el corpus se adjunta en CD). Barcelona: Printulibro.
- DÍAZ, L.; MARTÍNEZ, R. y REDÓ, J.A. (2011): *Guía de contenidos lingüísticos por niveles del español. Según el Marco Común Europeo de Referencia para ELE*, Colección recursos, núm. 124, Barcelona: Octaedro.
- DÍEZ-ITZA, E.; SNOW, C. y MACWHINNEY, B. (1999): "La metodología Retamhe y el Proyecto Childes: breviario para la codificación y análisis del lenguaje infantil" en *Psicothema. Revista Anual de Psicología*, vol. III, núm. 3, pp. 517-530. [En <<http://www.unioviedo.es/reunido/index.php/PST/article/view/7537/7401>>, 05/15].
- DÖRNYEI, Z. (2010): *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Second Language Acquisition Research Series. Monographs on Research Methodology (2nd ed.), London: Routledge.
- EAGLES Guidelines. *Expert Advisory Group on Language Engineering Standards* (1996). [En: <<http://www.ilc.cnr.it/EAGLES96/browse.html>>; 09/2014].
- ELLIS, R. (1992): *Second language acquisition and language pedagogy*, Clevedon: Multilingual Matters.
- ELLIS, R. Y BARKHUIZEN, G. (2005): "6 Functional analysis" en *Analysing Learner Language*, Oxford: Oxford University Press, pp. 111-138.
- ENRÍQUEZ, N.; DÍAZ, L. y TAULÉ, M. (2015): "Mental process in the oral production of non-native Spanish speakers: pauses and self-correction" en *XXXII Congreso internacional AESLA. Lingüística aplicada: industrias de la lengua y cambio social. Procedia-Social and Behavioral Sciences*, vol. 173, pp. 24-30, Elsevier. [En: <<http://www.sciencedirect.com/science/article/pii/S1877042815013348>>, 08/15].
- EQUIPO DE LA UNIVERSIDAD DE ALCALÁ (2001): *Sueña 3*. Libro del alumno. Nivel avanzado, Madrid: Anaya.
- EQUIPO PRISMA (2004): *Método de español para extranjeros PRISMA. Avanza (B2)*. Prisma del alumno, Madrid: Edinumen.
- ESCANDELL-VIDAL, V., (1995): "Cortesía, fórmulas convencionales y estrategias indirectas" en *Revista Español de Lingüística*, 25, vol. 1, pp. 31-66 [En: <http://portal.uned.es/pls/portal/docs/PAGE/UNED_MAIN/LAUNIVERSIDAD/UBICACIONES/04/DOCENTE/MARI>].

A_VICTORIA_ESCANDELL_VIDAL/PUBLICACIONES/CORTESÍA,%20FÓRMULAS%20CONVENCIÓNES%20Y%20ESTRATEGIAS%20INDIRECTAS.PDF>, 08/15].

ESCANDELL-VIDAL, V. (1996): *Introducción a la pragmática*, Ariel Lingüística: Barcelona.

FERNÁNDEZ-ORDÓÑEZ, I. (dir.) (2005-): *Corpus Oral y Sonoro del Español Rural*. [En: <<http://www.uam.es/coser>>, 05/15].

FERNÁNDEZ-ORDÓÑEZ, I. (2011): “Nuevos horizontes en el estudio de la variación gramatical del español: el *Corpus Oral y Sonoro del Español Rural*” en G. Colón y Ll. Gimeno (eds.), *Noves tendències en la dialectologia contemporània*, Castelló de la Plana: Universitat Jaume I, pp. 173-203.

FIRTH, J. R. (1956): “Linguistic analysis and translation” en F.R. Palmer (ed.), *For Roman Jakobson* (1968), Mouton: The Hague, pp. 84-95.

GARCÍA, C. (1989): “Disagreeing and requesting by Americans and Venezuelans” en *Linguistics and Education*, núm. 1, pp. 299-322.

GARCÍA, M. (2006): *Corpus de Conversaciones en Español como Lengua Extranjera en Linred XI* (Lingüística en la red), núm. 11. [En: <http://www.linred.es/informacion_pdf/informacion11_24012006.pdf>, 04/15]

GILLARD, P. y GADSBY, A. (1998): “Using a learners’ corpus in compiling ELT dictionaries” en S. Granger (ed.) *Learner English on Computer*. London: Addison Wesley Longman, pp. 159-171.

GILG, R.; LARSEN SERIGOS, J. y TROVATO, A. (2014): “Spanish in Texas: Open learning tools for exploring language diversity” en *ACTFL Annual Convention & World Languages Expo*, San Antonio, TX. [En: <http://www.slideshare.net/spanish_in_texas/spanish-in-texas-open-learning-tools-to-explore-language-diversity>, 07/15].

GOFFMAN, E. (1967): *Interaction ritual. Essays on face-to-face behaviour*, New York: Doubleday.

GONZÁLEZ et al. (2010): “ClInt: a Bilingual Spanish-Catalan Spoken Corpus of Clinical Interviews” en *Procesamiento del Lenguaje Natural*, Revista núm. 45, septiembre 2010. [En: <<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/viewFile/796/650>>, 04/15].

GONZÁLEZ ROYO, C. (2011a): “La problemática de la afinidad entre el español y el italiano en la enseñanza/aprendizaje desde la fraseología: el corpus de interlengua oral CORINEI” en *redELE*, núm. 20. [En: <<http://www.mecd.gob.es/dctm/redele/Material->

RedEle/Numeros%20Especiales/2011_ESP_11_VIIIEncuentropracticoELE/2011_ESP_05_01C ONFERENCIA%20APERTURA%20.pdf?documentId=0901e72b810d90c3>, 04/15].

GONZÁLEZ ROYO, C. (2011b): "Skype y la interacción oral nativo/no nativo: funciones y rutinas conversacionales en Corinéi, un corpus de interlengua español italiano" en *XXII Congreso Internacional de la ASELE. La red y sus aplicaciones en la enseñanza-aprendizaje del español como lengua extranjera*, Valladolid. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/asele_xxii.htm>, 04/15].

GRANGER, S. (2002): "A Bird's-eye view of learner corpus research" en S. Granger; J. Hung y S. Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, Amsterdam: John Benjamins, pp. 3-33.

GRANGER, S. (2003): "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research" en *TESOL Quarterly*, vol. 37, núm. 3, pp. 538-546. [En: <<http://web.international.ucla.edu/media/files/KISSELEV--International-Corpus-of-Learner-Eng-2w-4xg.pdf>>, 04/15].

GRANGER, S.; DAGNEAUX, E.; MEUNIER, F. y PAQUOT, M. (2009): *International Corpus of Learner English v2 (ICLEv2)*, Louvain-la-Neuve: Presses universitaires de Louvain.

GRICE, H.P. (1981): GRICE, H.P. (1981), "Presuposición e implicatura conversacional" en M. T. Julio y R. Muñoz (comp.) (1998), *Textos clásicos de pragmatic*, Madrid: Arco/Libros.

GRIES, S. Th. (2010): "Corpus linguistics and theoretical linguistics. A love-hate relationship? Not necessarily..." en *International Journal of Corpus Linguistics*, vol. 15, núm. 3, pp. 327-343, John Benjamins Publishing Company. [En: <http://www.linguistics.ucsb.edu/faculty/stgries/research/2010_STG_CorpLingLingTheory_IJCL.pdf>, 10/14].

GRUPO ILSE (2009): *Corpus del habla de Almería*, Almería: Universidad de Almería. [En: <<http://www.grupoilse.org/>>, 05/15].

GUMPERZ, J. y HYMES, D. (1964): "The ethnography of communication" en *American Anthropologist* 66, 6, parte 2.

GUTIÉRREZ, E. (2005): "Corpus de textos escritos por universitarios italianos estudiantes de ELE" en *Linred (Lingüística en la red)*, 1, 11. [En: <http://www.linred.es/informacion_pdf/informacion9_01112005.pdf>, 04/15].

GUTIÉRREZ QUINTANA, E. (2007): *Enseñar español desde un enfoque funcional*, Madrid: Arco/Libros.

HALLIDAY, M. A. K. (1973): *Explorations in the Functions of Language*. London: Edward Arnold.

- HAVERKATE, H. (1994): *La cortesía verbal. Estudio pragmalingüístico*, Madrid: Gredos.
- HERNÁNDEZ, M.; VILLALBA, F.; AGUIRRE, C. y NAJT, M. (2008a): “Corpus oral del proceso de adquisición del español como L2 para una joven inmigrante búlgara” en *Recursos para la Investigación Adquisición /Aprendizaje de Segundas Lenguas. La expresión oral en L2 (I)*, núm. 1, mayo, pp. 17-82. [En: <http://crei.centros.educa.jcyl.es/sitio/upload/Recursos_para_la_investigacion_en_adquisicionaprendizaje_de_segundas_lenguas._La_expresion_oral_en_L2_I.pdf>, 04/15].
- HERNÁNDEZ, M.; VILLALBA, F.; AGUIRRE, C. y NAJT, M. (2008b): “Estudio Comparativo entre la adquisición del Español como Primera Lengua y la Adquisición del Español como Segunda Lengua para su aplicación metodológica en la enseñanza del español a Inmigrantes” (1999) en *Dosieres Segundas Lenguas e Inmigración*, núm. 1. [En: <<http://www.segundaslenguaseinmigracion.com/Dosieres/Dosier1.pdf>>, 04/15].
- HUI-CHUAN LU (2010): “An annotated Taiwanese Learners’ Corpus of Spanish, CATE” en *Corpus Linguistics and Linguistic Theory*, núm. 6-2, pp. 297-300. [En: <<http://www.degruyter.com/view/j/cllt.2010.6.issue-2/cllt.2010.011/cllt.2010.011.xml?formatXEQINT>>, 09/2015].
- HULSTIJN, J. H. (2014): “The Common European Framework of Reference for Languages: a challenge for applied linguistics” en *ITL: International Journal of Applied Linguistics*, vol. 165, núm. 1, pp. 3-18.
- HYMES, D. (1972): *Directions in sociolinguistics. The ethnography of communication*. Nueva York: Basil Blackwell.
- IDE, S. (1989): “Formal forms and discernment: Two neglected aspects of linguistic politeness” en *Multilingua*, núm. 8, pp. 223-248.
- IFE, A. (2004): “The L2 learner corpus: reviewing its potential for the early stages of learning” en M. Baynham; A. Deignan y G. White (eds.), *Applied Linguistics at the Interface: British Studies in Applied Linguistics*, London: Equinox, pp. 91-104.
- INSTITUTO CERVANTES (2007): *Plan curricular del Instituto Cervantes. Niveles de referencia para el español, niveles B1-B2*. Madrid: Biblioteca Nueva. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/>, 06/15].
- JAKOBSON, R. (1963): “Linguistique et poétique” en *Essais de linguistique générale*, 1.
- JØRGENSEN, A. M. (2008): “COLA, Un Corpus Oral de Lenguaje Adolescente” en *Anejos de Oralía*, 3/1, pp. 225-234. [En: <<http://www.colam.org/ArTICULO%20Oralia%202006,%20Almeria.pdf>>, 05/15].

- KAMAKURA, Y. (2005): "Japanese Learner Corpus of Spanish – L2 and L2 interlanguage contrastive analysis" en *4th International Contrastive Linguistics Conference (ICLC4)*, Santiago de Compostela.
- KASPER, G. (1990): "Linguistics Politeness", *Journal of Pragmatics*, núm. 14, pp. 193-218.
- KASPER, G. (1992): "Pragmatic Transfer" en *Second Language Research*, núm. 8, pp. 203-231.
- KERBRATT-ORECCHIONI, C. (2004): "¿Es universal la cortesía?" en D. Bravo y A. Briz (eds.), *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, Barcelona: Ariel Lingüística.
- KHON, K. (2011): *Corpus BACKBONE*. [En: <<http://projects.ael.uni-tuebingen.de/backbone/moodle/>>, 05/15].
- KOHN, K., HOFFSTAEDTER, P. y WIDMANN, J. (2010): "BACKBONE - pedagogic corpora for content and language integrated learning" en *Eurocall Conference, 9-12 September 2009. Conference Proceedings. Macmillan ELT*. [En: <http://projects.ael.uni-tuebingen.de/backbone/moodle/file.php/1/publications/Eurocall09_proceedings_54695.pdf>; 05/15].
- KRAMSCH, C. (1993): *Context and culture in language teaching*, Oxford: Oxford University Press.
- KRAMSCH, C. (2001): "El privilegio del hablante intercultural" en M. Byram y M. Fleming (eds.), *Perspectivas interculturales en el aprendizaje de idiomas*, Cambridge University Press, pp. 23-37.
- KRASHEN, S. (1985): *The Input Hypothesis: Issues and Implications*, London: Longman.
- LARSEN-FREEMAN, D. (2001): "Teaching grammar" en M. Celce-Murcia (ed.), *Teaching English as a second or foreign language*, 3ª ed., pp. 251-266. Boston: Heinle & Heinle.
- LEECH, G. (1983): *Principles of pragmatics*, London and New York: Longman.
- LEECH, G. (1997): "Teaching and Language Corpora: A Convergence" en Wichmann et al. 1997, pp. 1-23.
- LEVINSON, S. C. (1983): *Pragmática*, (1989). Barcelona: Teide.
- LÓPEZ, Mª P. (2013): "Aplicaciones de los corpus orales a la didáctica de ELE: repertorios léxicos y contextos pragmlingüísticos" en *XXIV Congreso Internacional de la ASELE. La enseñanza del español como LE/L2 en el siglo XXI*, Jaen. [En: <http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/24/24_397.pdf>, 07/15].

- LOZANO, C. (2009): "CEDEL2: Corpus Escrito del Español L2", en C. M. Bretones et al. (eds.), *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*, Almería: Universidad de Almería, pp. 187-212. [En: <<http://digibug.ugr.es/bitstream/10481/22177/6/Lozano%20%282009%29%20CEDEL2-Almer%C3%ADa.pdf>>, 04/15].
- LOZANO, C. y MENDIKOETXEA, A. (2013): "Learner corpora and Second Language Acquisition: The design and collection of CEDEL2" en N. Ballier; A. Díaz-Negrillo y P. Thompson (eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins.
- LLISTERRI, J. (2015): *Corpus de lengua oral en español*. [En: <http://liceu.uab.es/~joaquim/language_resources/spoken_res/Corp_leng_oral_esp.html>, 04/15].
- LLISTERRI, J. y GARRIDO, J.M. (1998): "La ingeniería Lingüística en España" en *El español en el mundo. Anuario*, Madrid: Instituto Cervantes. [En: <http://cvc.cervantes.es/lengua/anuario/anuario_98/llisterri/p01.htm>, 05/15].
- LLORIÁN, S. (2007): *Entender y utilizar el Marco común europeo de referencia desde el punto de vista del profesor de lenguas*, Madrid: Santillana educación.
- MARCOS MARÍN, F. (1991): "Corpus lingüístico de referencia de la lengua española" en *Boletín de la Academia de la lengua española*, núm. 56, pp. 129-155.
- MARCOS MARÍN, F. y MORENO, A. (1992): CORLEC (Corpus Oral de Referencia de la Lengua Española Contemporánea). [En: <<http://www.llf.uam.es/ESP/Corlec.html>>, 04/15].
- MÁRQUEZ REITER, R. (2000): *Linguistic Politeness in Britain and Uruguay*, Amsterdam/Filadelfia: John Benjamins.
- MARTÍ, M.A. y TAULÉ, M. (2014): "«M'exalta el nou i me'enamora el vell»: els corpus, passat i futur de la lingüística" en M.A. Martí y M. Taulé (coords.), *Homenatge a Sebastià Serrano*, Col·lecció Homenatges: 44, Barcelona: Universitat de Barcelona, pp. 177-204.
- MARTINET, A. (1970): *Éléments de linguistique générale*, Paris: Armand Colin.
- MATTE BON, F. (2004): "Los contenidos funcionales y comunicativos" en J. Sánchez Lobato y I. Santos Gargallo, *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid: Edelsa.
- MCCARTHY, M. y CARTER, R. (1995): "Spoken grammar: what is it and how we teach it?", *ELT J.*, pp. 207-218. [En: <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=319A4820B95C215144C7C2A419E20CFD?doi=10.1.1.458.7590&rep=rep1&type=pdf>>, 04/15].

- MCCULLOUGH, J.L. (2001): “ Los usos de los corpóra de textos en la enseñanza de lenguas” en M. Trenchs (ed.), *Nuevas Tecnologías para el autoaprendizaje y la didáctica de lenguas*, Lleida: Milenio, pp. 125-140.
- MCENERY, T., XIAO, R. y TONO, Y. (2006): *Corpus-based Language Studies: An Advanced ResourceBook*, London/New York: Routledge.
- MCWHINNEY, B. (1995, 2000), *The CHILDES project: tools for analysing talk*, Hillsdale, NJ, LEA.
- MENDIKOETXEA, A.; O’DONNELL, M. y ROLLINSON, P. (2009): “WriCLE: a learner corpus for second language acquisition research” en M. Mahlberg; V. González-Díaz y C. Smith (eds), *Proceedings of Corpus Linguistics*, University of Liverpool, artículo 351. [En: <<http://ucrel.lancs.ac.uk/publications/cl2009/>>, 04/15].
- MEURERS, W. D. (2015): “Learner Corpora and Natural Language Processing” en S. Granger; G. Gilquin y F. Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, Cambridge University Press. [En: <<http://www.sfs.uni-tuebingen.de/~dm/publications.html>>, 04/15].
- MITCHELL, R.; DOMÍGUEZ, L.; ARCHE, M^a; MYLES, F. y MARSDEN, E. (2008): “Research Report: Linguistic Development in L2 Spanish: Creation and analysis of a learner corpus” en L. Roberts; F. Myles y A. David (eds.), *EuroSLA Yearbook*, 8, Amsterdam: John Benjamins; Philadelphia, pp. 287-304. [En: <http://www.splloc.soton.ac.uk/doc/SPLLOC_1_Report_Final.pdf>, 04/15].
- MORENO, F. (2001): “El corpus ACUAH: Análisis de los clíticos” en J. De Kock (ed.) *Gramática española enseñanza e investigación. Apuntes metodológicos. Lingüística con corpus: catorce aplicaciones sobre el español*, Salamanca: Universidad de Salamanca, pp. 353-368.
- MORENO, F. (2005): “Corpus para el estudio del español en su variación geográfica y social. El corpus “PRESEEA”” en *Oralia*, vol. 8, pp. 125-139.
- MORENO SANDOVAL, A. (2002): “La evolución de los corpus de habla espontánea: la experiencia del LLI-UAM” en *II Jornadas de Tecnologías del Habla*, Granada, España. [En: <<https://uam.academia.edu/AntonioMorenoSandoval>>, 04/15].
- MORENO SANDOVAL, A.; DE LA MADRID, G.; ALCÁNTARA, M.; GONZÁLEZ, A.; GUIRAO, J.M. y DE LA TORRE, R. (2005): “Chapter 4: The Spanish Corpus” en E. Cresti y M. Moneglia, *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: John Benjamins.
- MORENO, A., y GUIRAO, J. M. (2006): “Morphosyntactic Tagging of the Spanish C-ORAL-ROM Corpus: Methodology, Tools and Evaluation” en Y. Kawaguchi; S. Zaima y T.

- Takagaki (eds.), *Spoken Language Corpus and Linguistic Informatics*, pp. 199-218. Amsterdam: John Benjamins.
- MUÑOZ, C. (2014): “Español oral en contexto. Campillos Llanos, L., Gozalo Gómez, P., Guirao, J. M^a y Moreno Sandoval, A., I.S.B.N. 978-84-8344-181-7; Madrid: UAM Ediciones, 2010” en *redELE*, núm. 26. [En: <http://www.mecd.gob.es/dctm/redele/Material-RedEle/Revista/2014_26/2014_redELE_26_04MUNOZreseña.pdf?documentId=0901e72b818b81c9>, 04/15].
- NESSELHAUF (2004): “Learner corpora and their potential for language teaching” en J. M. Sinclair (ed.), *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp. 187-199.
- NICA, V. (2012): “Los corpus discursivos orales como herramienta didáctica para los niveles iniciales” en *Revista Foro de Profesores de E/LE*, vol. 8. [En: <<http://foroele.es/revista/index.php/foroele/article/view/25>>; 05/15].
- NICOLÁS, C. (2002): “Una propuesta de utilización de corpus orales en la enseñanza de segundas lenguas” en *Congreso CLIP 2002*, Lamusa Digital, núm. 3 (monográfico). [En: <<http://www.uclm.es/lamusa/index.asp?accion=misce&id=52&lengua=es>>, 04/15].
- NICOLÁS, C. (2012a): *Corpus C-Or-DiAL. Corpus Oral Didáctico Anotado Lingüísticamente*. [En: <<http://lablita.dit.unifi.it/corpora/cordial/>>, 04/15].
- NICOLÁS, C. (2012b): *Corpus C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*, Madrid: Liceus Lengua Española.
- NORTH, B., ORTEGA, A. y SHEEHAN, S. (2010): *A Core Inventory for General English*. British Council/EAQUALS. [En: <http://clients.squareeye.net/uploads/eaquals2011/documents/EAQUALS_British_Council_Core_Curriculum_April2011.pdf>, 09/15].
- NUMAN, D. (1991): *Methods in second language classroom oriented research: a critical review*. *Studies in second language acquisition*, vol. 13, pp. 249-274.
- OROL, A. y ALONSO RAMOS, M. (2013): “A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish” en *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 563-570. [En: <<http://www.dicesp.com/app/webroot/files/file/Orol%20González%20and%20Alonso%20Ramos%202013.pdf>>, 08/14].
- ORTEGA, A. (2010): “¿Qué son en realidad los niveles C? Desarrollo de sus descriptores en el MCER y el PEL”, en Ortega Calvo (ed.) *Niveles C: Currículos, programación enseñanza y certificación*, Madrid: IFIIE – Ministerio de Educación, pp. 21-85.
- PADRÓ, L. (s.f.): *Demonstration. FreeLing 3.1. An open-source suite of language analyzers*. Barcelona: TALP - Tecnologies i Aplicacions del Llenguatge i de la Parla, Universitat

- Politécnica de Catalunya. [En: <<http://nlp.lsi.upc.edu/freeling/demo/demo.php>>, 09/15].
- PARODI, G. (2008): “Lingüística de corpus: una introducción al ámbito” en *Revista de Lingüística Teórica y Aplicada*, vol. 46, núm. 1, pp. 93-119. [En: <http://www.scielo.cl/scielo.php?pid=S0718-48832008000100006&script=sci_arttext>, 10/14].
- PASCUAL, C. (2013): “Análisis del marcador discursivo bueno en las conversaciones de estudiantes italianos y españoles hablando en español” en *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*, núm. 13. [En: <<http://www.nebrija.com/revista-linguistica/analisis-del-marcador-discursivo-bueno-en-las-conversaciones-de-estudiantes-italianos-y-espanoles-hablando-en-espanol>>, 04/15].
- PASCUAL, C. (2015): “Corpus de conversaciones en italianos y en español LE (CIELE)” en *Linred XII* (Lingüística en la red), núm. 12. [En: <http://www.linred.es/numero12_corpus-1.html>, 04/15].
- PAZ, Y. (2013): “Aplicación pedagógica de la lingüística de corpus a una clase de gramática avanzada de español como segunda lengua” en *MarcoELE*, núm. 16. [En: <<http://marcoele.com/linguistica-de-corpus/>>, 04/15].
- PÉREZ HERNÁNDEZ, M.C. (2002): “Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento” en *Estudios de lingüística Española* (ELiEs), Universidad de Málaga. [En: <<http://elies.rediris.es/elies18/>>, 04/15].
- PÉREZ-PAREDES, P. (2010): “Corpus linguistics and language education in perspective: appropriation and the possibilities scenario” en T. Harris y M. Moreno (eds.), *Corpus Linguistics in Language Teaching*, Peter Lang AG: Bern, pp. 53-77.
- PÉREZ RUIZ, J. (2011): “Análisis de errores y de la interacción oral como bases para una propuesta didáctica en la clase de conversación: primeros resultados” en *Revista Nebrija de Lingüística aplicada a la enseñanza de las lenguas*, núm. 10, pp. 27-51. [En: <www.nebrija.com/revista-linguistica/revistanebrija10/revistaNebrija10.pdf>, 05/15].
- PINO, A. (2009): “Palabras en interacción: un corpus de aprendices de E/LE” en P. Cantos y A. Sánchez (eds.), *A Survey on Corpus-based Research/Panorama de investigaciones basadas en corpus*. I Congreso Internacional de Lingüística de Corpus (CILC-09), AELINCO, 7-9 mayo, Universidad de Murcia, 2009. [En: <<http://www.um.es/lacell/aelinco/contenido/pdf/33.pdf>>, 05/15].
- PINTO, D. (2002): *Perdóname ¿llevas mucho esperando? Conventionalized language in L1 and L2 Spanish*. Unpublished doctoral dissertation, Davis: University of California.

- PLACENCIA, M. E. y BRAVO, D. (2002): “Panorámica sobre el estudio de los actos de habla y la cortesía lingüística” en M. E. Placencia y D. Bravo (eds.), *Actos de habla y cortesía en español*, Lincom Europa, Munich: Lincom Studies in Pragmatics 05, pp. 292-305.
- PRESEEA (2015-): *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*, Alcalá de Henares: Universidad de Alcalá. [En: <<http://preseea.linguas.net>>, 05/15].
- PROYECTO COLA (2002-): *Corpus Oral de Lenguaje Adolescente*. [En: <<http://www.colam.org/>>, 05/15]
- PROYECTO SACODEYL (s.f.): *System Aided Compotation and Open Distribution of European Youth Language* (SACODEYL). [En: <<http://www.um.es/web/filologia-inglesa/contenido/investigacion/proyectos/sacodeyl>>, 04/15].
- REAL ACADEMIA ESPAÑOLA (2004): Banco de datos CREA. *Corpus de Referencia del Español Actual*. [En: <<http://corpus.rae.es/cranet.html>>, 04/15].
- ROBINSON, P y ELLIS, N. C., eds. (2008): *A handbook of cognitive linguistics and second language acquisition*. London: Routledge.
- ROJO, G. (2015): “Hispanic Corpus Linguistics” en M. Lacorte (ed.), *The Routledge Handbook of Hispanic Applied Linguistics*, Routledge Handbooks: University of Mariland.
- RÖMER, U. (2004): “A corpus-driven approach to modal auxiliaries and their didactics” en J. M. Sinclair (ed.), *How to Use Corpora in Language Teaching*, Amsterdam: John Benjamins, pp.187-199.
- RÖMER, U. (2008): “7. Corpora and language teaching” en L. Anke y M. Kytö (eds.), *Corpus Linguistics. An Interantional Handbook* (volume 1), Berlin/New York: Mouton de Gruyter, pp. 112-131.
- RUBIO, M. (2009): “Creación del Corpus Oral del español en Taiwán (COET). Primeros textos” en *Linred VI (Lingüística en la Red)*, núm. 6. [En: <http://www.linred.es/informacion_pdf/LR_informacion17_15022009.pdf>, 04/15].
- RUBIO, M. (2013a): *El mecanismo de las interrupciones en la conversación de alumnos taiwaneses de español como lengua extranjera*, Tesis doctoral inédita, Alcalá de Henares: Universidad de Alcalá.
- RUBIO, M. (2013b): “Problemas en el aprendizaje del mecanismo de las interrupciones en la conversación en español de alumnos taiwaneses de ELE” en *Linred XI (Lingüística en la Red)*, núm. 11. [En: <http://www.linred.es/241rticulos_pdf/LR_articulo_30052013.pdf>, 04/15].

- SAMPER, A.; HERNÁNDEZ; C. y TROYA, M. (1998): *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispano (MC-NLCH)*, Las Palmas de Gran Canaria: Servicio de Publicaciones de la Universidad de Las Palmas de Gran Canaria-ALFAL. Edición en CD-ROM.
- SANCHEZ, A. (2001): “Las simulaciones en la didáctica de la lengua” en M. Trenchs Parera, (ed.), *Nuevas Tecnologías para el autoaprendizaje y la didáctica de lenguas*, Lleida: Milenio, pp. 51-70.
- SCHEGLOFF, E. A. (2007): *Sequence Organization in Interaction: Introduction to sequence organization*. Cambridge: Cambridge University Press.
- SEARLE, J. (1969): *Speech Acts*. Traducción española *Actos de habla* (1980), Cátedra: Madrid.
- SECCHI, D. (2014): “La conversación coloquial en E/LE: carencias, necesidades y propuestas” en *Anuario brasileño de estudios hispánicos*, XXIV, pp. 11-26. [En: <<http://www.mecd.gob.es/brasil/dms/consejerias-exteriores/brasil/2014/publicaciones/abehxxiv.pdf>>, 04/15].
- SELINKER, L. (1972): “Interlanguage” en *International Review of Applied Linguistics*, vol. 10, núm. 3, pp. 209-301. Traducción española: “La interlengua”, en J. Muñoz Liceras (1992) *La Adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*. Madrid: Visor, pp. 79-101.
- SIEBOLD, K. (2006): “¿Cómo se piden las cosas? Estudio pragmlingüístico de las peticiones en español y en alemán” en C. Mourón y T.I. Moralejo (eds.), *Studies in Contrastive Linguistics, Proceedings of the 4th International Contrastive Linguistics Conference* (2005), Universidad Santiago de Compostela, pp. 953-960. [En: <http://www.upo.es/export/portal/com/bin/portal/upo/profesores/ksiex/profesor/1317216475682_siebold-peticiones_xespaxolx.pdf>, 06/15].
- SIEBOLD, K. (2007): “El papel de la pragmática en la comunicación intercultural entre españoles y alemanes” en L. Luque y A. Pamies (eds.), *Interculturalidad y Lenguaje*, Lingüística: Granada, pp. 281-290. [En: <http://www.upo.es/export/portal/com/bin/portal/upo/profesores/ksiex/profesor/1317216456183_siebold-papel_pragmatica_2007.pdf>, 06/15].
- SINCLAIR, J.M., (1996): “Preliminary Recommendations on Corpus Typology” en *EAGLES Document eag-tcwg-ctyp/p* [En: <<http://www.ilc.cnr.it/EAGLES/corpus/typ/corpus/typ.html>>, 10/2014].
- SINCLAIR, J. (ed.) (2004): *How to use corpora in language teaching*, Amsterdam: John Benjamins.

- SLAGTER, P.J. (1979): *Un nivel umbral*, Estrasburgo: Consejo de Europa.
- STUBBS, M. (1987): *Análisis del discurso: análisis sociolingüístico del lenguaje natural*, Madrid: Alianza editorial.
- TAGNIN, S. E. O. (2002): *A Multilingual Learner Corpus (MLC) in Brazil*. [En: <<http://comet.fflch.usp.br/sites/comet.fflch.usp.br/files/u30/A%20multilingual%20learner%20corpus%20in%20Brazil.pdf>>, 05/15].
- TELETÁNDEM-CORINÉI: Publicaciones. [En: <<http://dti.ua.es/va/teletandem/publicaciones/publicaciones.html>>, 04/15].
- TEI Consortium (eds.) (2015): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [Version 2.8.0] [En: <<http://www.tei-c.org/Guidelines/P5/>>, 09/2014].
- THORNBURY, S. (2011): “C is for Core Inventory” en *An A-Z of ELT. Scott Thornbury’s blog*. [En: <<https://scottthornbury.wordpress.com/2011/02/06/c-is-for-core-inventory/>>, 10/14].
- TOGNINI-BONELLI E. (ed.) (2001): *Corpus linguistics at work. Studies in corpus linguistics*, núm. 6, Amsterdam: John Benjamins.
- TORRUELLA, J. y LLISTERRI, J. (1999): “Diseño de corpus textuales y orales” en J.M. Bleca; G. Clavería; C. Sánchez y J. Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Seminario de Filología e Informática de la Universidad Autónoma de Barcelona, Barcelona: Millenio, pp. 45-46. [En: <http://latel.upf.edu/traductica/lc/material/torruella_llisterri_99.pdf>, 10/14].
- VAN EK, J.A. y TRIM, J.L.M. (2001): *Waystage*. Council of Europe. Cambridge: Cambridge University Press.
- VAN EK, J.A. y TRIM, J.L.M. (2001): *Threshold 1990*. Council of Europe. Cambridge: Cambridge University Press.
- VAN EK, J.A. y TRIM, J.L.M. (2001): *Vantage Level*. Council of Europe. Cambridge University Press [En: <https://www.coe.int/t/dg4/linguistic/Vantage_CUP.pdf>, 10/14].
- VILLAYANDRE LLAMAZARES, M. (2008): “Lingüística de corpus” en *E.H. Filología*, núm. 30, pp. 329-349.
- VV.AA. (2008): *Destino Erasmus 2. Niveles intermedio y avanzado (B1-B2)*. Estudios Hispánicos, Universidad de Barcelona, Barcelona: SGEL.
- WIERZBICKA, A. (1985): “Different cultures, different languages, different speech acts” en *Journal of Pragmatics*, núm. 9, pp. 145-281.

WIERZBICKA, A. (1991): *Cross-Cultural Pragmatics. The Semantics of Human Interaction*, Berlín-Nueva York: Munton de Gruyter, pp. 255-281.

WILKINS, D. A. (1972): "Grammatical, situational and notional syllabuses" en C. J. Brumfit y K. Johnson (1979), *The Communicative Approach to Language Teaching*, Oxford: Oxford University Press, pp. 82-90.

WILKINS, D. A. (1976): *Notional Syllabuses*, Oxford: Oxford University Press.

ZANÓN, J. (2007): "Psicolingüística y didáctica de las lenguas: una aproximación histórica y conceptual" en *MarcoELE*, núm. 5. [En: < <http://marcoele.com/gramatica-cognitiva-y-ele/>>, 10/14].

Fuentes URLs:

ALTE (*The Association of Language Testers in Europe*): <<http://www.alte.org>>

Archivo vídeo SpinTX: <<http://www.coerll.utexas.edu/spintx/>>

Diccionario de términos clave de ELE del Centro Virtual Cervantes:

<http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario>

Corpus de aprendices de español (CAES): <<http://galvan.usc.es/caes>>

Corpus CbLLE: <<http://cbll.tufs.ac.jp/tag/es/search.php?menulang=en>>

Corpus de conversaciones en español de Columbia University:

<<http://corpusdeconversaciones.com>>

Corpus Escrito Del Español L2 (CEDEL2):

<<http://www.uam.es/proyectosinv/woslac/cedel2.htm>>

English Profile. The CEFR for English: <<http://www.englishprofile.org/english-grammar-profile/egp-online>>

Etiquetas EAGLES para el español: <<http://nlp.lsi.upc.edu/freeling/doc/tagsets/tagset-es.html>>

GRAMPAL: <<http://cartago.llf.uam.es/grampal/grampal.cgi>>

The Longman Learners' Corpus (LLC):

<<http://www.pearsonlongman.com/dictionaries/corpus/learners.html>>

Louvain Corpus Of Native English Essays (LOCNESS):

<<http://www.learnercorpusassociation.org/resources/corpora/locness-corpus/>>

Reference Levels Descriptors en ingles: <http://www.coe.int/t/dg4/linguistic/DNR_EN.asp>

Proyecto SACODEYL: <<http://www.um.es/web/filologia-inglesa/contenido/investigacion/proyectos/sacodeyl>>

Proyecto europeo Word Order in Second Language Acquisition Corpora (WOSLAC):

<<http://www.uam.es/proyectosinv/woslac/>>

Proyecto Talkbank (Carnegie Mellon University y University of Pennsylvania):

<<http://talkbank.org>>

TRANSCRIBER: <<http://trans.sourceforge.net/en/presentation.php>>

Transana: <www.transana.org>

Santiago University Learner of English Corpus (SULEC): <<http://www.sulec.es>>

Web Learner Corpora around the world: <<http://www.uclouvain.be/en-cecl-lcworld.html>>

WINPICH: <<http://www.winpitch.com>>

Written Corpus of Learner English (WriCLE):

<http://web.uam.es/proyectosinv/woslac/Wricle/>

ANEXOS¹²⁶

ANEXO I. Epígrafes del MCER

En el MCER, 4.4.3.1. Actividades de interacción oral, incluido el cuadro en el apartado 4.4. Actividades comunicativas. La letra negrita se refiere al escenario común para la práctica de la interacción oral de una transacción.

INTERACTUAR PARA OBTENER BIENES Y SERVICIOS ¹²⁷	
B1	Se desenvuelve bien en la mayoría de las transacciones que pueden surgir mientras viaja , organiza el viaje o el alojamiento o trata con las autoridades competentes durante un viaje al extranjero.
	Se desenvuelve en situaciones menos habituales en tiendas, oficinas de correo, bancos; por ejemplo, devuelve una compra con la que no queda satisfecho. Sabe plantear una queja o hacer una reclamación.
	Se desenvuelve en la mayoría de las situaciones que suelen surgir cuando se organiza un viaje a través de una agencia o cuando se está viajando; por ejemplo: preguntarle a un pasajero dónde debe bajarse cuando el destino es desconocido.
	Se desenvuelve en aspectos comunes de la vida cotidiana como son los viajes , el alojamiento, las comidas y las compras. Consigue toda la información que necesita de una oficina de turismo siempre que esa información sea sencilla y no especializada.

FIGURA 37. Descriptores del MCER: Epígrafe 4.4.3.1

En el MCER, 5.2.3.2. La competencia funcional¹²⁸.

¹²⁶ Todos los anexos, también, pueden consultarse en el cd que se incluye en este trabajo de investigación.

¹²⁷ Solo hemos reproducido el nivel B1, a modo de ejemplo:

<http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cap_04.htm#p443>

¹²⁸ <http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cap_05.htm#p523>

(...)

Son comunes los tríos en los que el primer hablante admite la respuesta del interlocutor o responde a ella. Los pares y los tríos suelen ir adheridos a transacciones e interacciones más extensas. Por ejemplo, en transacciones de colaboración que son más complejas y que están orientadas a la meta, la lengua se hace necesaria para:

- formar el grupo de trabajo y establecer relaciones entre los participantes;
- establecer el conocimiento común de las características relevantes de la situación actual y llegar a una lectura común;
- identificar lo que podría y debería ser cambiado;
- establecer un acuerdo común sobre los objetivos y la acción requerida para cumplirlos;
- acordar roles para realizar la acción;
- controlar las acciones prácticas que hay que realizar a la hora de, por ejemplo:
 - identificar y abordar los problemas que surjan;
 - coordinar y secuenciar las intervenciones;
 - alentarse mutuamente;
- reconocer el logro de metas intermedias;
- reconocer la consecución de la tarea;
- evaluar la transacción;
- completar y terminar la transacción.

El proceso total se puede representar esquemáticamente. Un ejemplo es el esquema general que se ofrece para la compra de bienes y servicios en el capítulo 8 de *Threshold Level 1990*:

Esquema general para la compra de bienes y servicios

1. Desplazarse al lugar de la transacción:
 - 1.1. encontrar el camino a la tienda, almacén, supermercado, restaurante, estación, hotel, etc.;
 - 1.2. encontrar el camino al mostrador, al departamento, a la mesa, a la taquilla, a la recepción, etc.
2. Establecer el contacto:
 - 2.1. intercambio de saludos con el tendero o el dependiente o el camarero o el recepcionista, etc.:
 - 2.1.1. el dependiente saluda;
 - 2.1.2. el cliente saluda.
3. Elegir bienes o servicios:
 - 3.1. Determinar la categoría de los bienes o servicios requeridos:
 - 3.1.1. buscar información;
 - 3.1.2. dar información.
 - 3.2. Identificar opciones.
 - 3.3. Discutir las ventajas y los inconvenientes de las opciones (por ejemplo: calidad, precio, color, tamaño de los bienes):

- 3.3.1. buscar información;
- 3.3.2. dar información;
- 3.3.3. pedir consejo;
- 3.3.4. dar consejo;
- 3.3.5. preguntar por preferencias;
- 3.3.6. expresar preferencias, etc.
- 3.4. Identificar los bienes concretos requeridos.
- 3.5. Examinar bienes.
- 3.6. Acordar la compra.
- 4. Intercambio de bienes por dinero:
 - 4.1. acordar los precios de los objetos;
 - 4.2. acordar la suma del total;
 - 4.3. recibir o entregar el pago;
 - 4.4. recibir o entregar los bienes (y el recibo);
 - 4.5. intercambio de agradecimientos:
 - 4.5.1. el dependiente da las gracias;
 - 4.5.2. el cliente da las gracias.
- 5. Despedida:
 - 5.1. Expresión de satisfacción (mutua):
 - 5.1.1. el dependiente expresa satisfacción;
 - 5.1.2. el cliente expresa satisfacción.
 - 5.2. Intercambio de comentarios interpersonales (por ejemplo: sobre el tiempo, cotilleos de la zona).
 - 5.3. Intercambio de saludos de despedida:
 - 5.3.1. el dependiente se despide;
 - 5.3.2. el cliente se despide.

FIGURA 38. Esquema general del proceso de la compra de bienes y servicios

5.1.2. Las destrezas y las habilidades¹²⁹

5.1.2.1. Las destrezas y las habilidades prácticas

- Destrezas sociales: la capacidad de actuar de acuerdo con los tipos de convenciones presentados en 5.1.1.2. (el conocimiento sociocultural) y de comportarse de la forma esperada, siempre que los foráneos, y en especial los extranjeros, lo consideren apropiado.
- Destrezas de la vida: la capacidad de llevar a cabo con eficacia las acciones

¹²⁹ <http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cap_05.htm#p512>

rutinarias que se requieren en la vida diaria (bañarse, vestirse, pasear, cocinar, comer, etc.), el mantenimiento y la reparación de los aparatos de la casa; etc.

- Destrezas profesionales: la capacidad de realizar acciones especializadas (mentales y físicas) que se necesitan para realizar los deberes del (auto)empleo.
- Destrezas de ocio: la capacidad de realizar con eficacia las acciones necesarias para las actividades de ocio; por ejemplo:
 - artes (pintura, escultura, tocar instrumentos musicales, etc.);
 - trabajos manuales (hacer punto, bordado, tejer, cestería, carpintería, etc.);
 - deportes (juegos de equipo, atletismo, carrera, escalada, natación, etc.);
 - aficiones (fotografía, jardinería, etc.).

Los usuarios del *Marco de referencia* pueden tener presente y, en su caso, determinar:

- Qué destrezas y habilidades prácticas necesitará el alumno o cuáles se le exigirá que tenga con el fin de comunicarse con eficacia en un área de interés.

5.1.2.2. Las destrezas y las habilidades interculturales

Este apartado incluye:

- La capacidad de relacionar entre sí la cultura de origen y la cultura extranjera.
- La sensibilidad cultural y la capacidad de identificar y utilizar una variedad de estrategias para establecer contacto con personas de otras culturas.
- La capacidad de cumplir el papel de intermediario cultural entre la cultura propia y la cultura extranjera y de abordar con eficacia los malentendidos interculturales y las situaciones conflictivas.
- La capacidad de superar relaciones estereotipadas.

Los usuarios del *Marco de referencia* pueden tener presente y, en su caso, determinar:

- Qué papeles y funciones de intermediario cultural tendrá que cumplir el alumno, cómo se le capacitará para ello o qué se le exigirá al respecto.
- Qué características de las culturas propia y extranjera tendrá que distinguir el alumno, cómo se le capacitará para ello o qué se le exigirá al respecto.
- Qué se le puede ofrecer al alumno para que experimente la cultura objeto de estudio.
- Qué posibilidades tendrá el alumno de actuar como intermediario cultural.

FIGURA 39. *Descriptor del MCER: Epígrafe 5.1.2. Las destrezas y las habilidades*

ANEXO II. Diseño de la actividad *role-play* 1: diálogo tren

1. CONTEXTUALIZACIÓN

Visualizar fragmento de la película “Con la muerte en los talones” de Hitchcock, del 41’:18’’al 41’:48’’.

2. EJEMPLO

TAQUILLERO: Buenas tardes.

VIAJERO: Buenas tardes, ¿podría informarme del tren Barcelona-Madrid?

TAQUILLERO: ¿Quieres viajar por la mañana?

VIAJERO: Sí, por la mañana.

TAQUILLERO: a ver, por la mañana, el primero es a las 6 de la mañana y llega a las 8h55min. Es un AVE.

VIAJERO: Vale y más tarde, a las 9 de la mañana.

TAQUILLERO: Sí, hay uno a las 9h que llega a las 11h 45min. También en AVE.

VIAJERO: Vale, ya me está bien. Sería para el 24 de mayo.

TAQUILLERO: No habría ningún problema, hay todavía plazas.

VIAJERO: De acuerdo

TAQUILLERO: Vale.

VIAJERO: Gracias.

TAQUILLERO: Hasta luego.

3. PRÁCTICA

Ahora, por parejas, vais a preparar una conversación similar. Uno será **el taquillero**, otro **el viajero**. Os ayudamos con la siguiente información:

TAQUILLERO PARA VIAJERO A: Dar información, tren Barcelona-Granada

TREN/RECORRIDO	SALIDA	LLEGADA	DURACIÓN	PRECIOS	PRESTACIONES
Talgo (Con enlace) 1er. trayecto Bcn-Linares (8:30h-16:35h) 2º trayecto Linares-Granada (16: 45h-19:08h)	8:30	19:08	10h.38min.	Turista:61€ Preferente: 80 €	-Climatización -Servicio cafetería
Tren hotel (directo)	21:30	08:05	10h.35min.	Turista(sentado): 64€ Preferente(sentado): 85€ Cama individual turista: 92€ Cama preferente individual: 166€	-Climatización -Servicio cafetería -Servicio restaurante -Prensa en preferente -Bebida de bienvenida -Incluida cena y desayuno en coche restaurante

PREGUNTAR:

-Recorrido: ¿solo ida? - ¿ida y vuelta? ¿Directos o no? (Con el Talgo hay que *cambiar de tren* en Linares y hay una espera de 15 minutos).
 -Día
 -Tipo de tren
 -Precio
 -Descuentos: solo ida, no descuentos – ida y vuelta, descuento del 10%. *Carnet de estudiante* para hacer un descuento del 10%.

VIAJERO A: Características

-Estudiante, tienes poco dinero. Pides información:
 -Viajar por la noche.
 -Recorrido: Barcelona-Granada, solo ida.
 -Elegir:
 *dormir en cama/ ir sentado
 *billete barato /caro

TAQUILLERO PARA VIAJERO B: Dar información, tren Barcelona-Sevilla

TREN/RECORRIDO	SALIDA	LLEGADA	DURACIÓN	PRECIOS	PRESTACIONES
Alaris	8:30	19:20	10h.50min	Turista:68€ Preferente: 90 €	-Climatización -Servicio cafetería
AVE	8:30	13:57	5h.27min.	Turista: 142€ Business: 256€	-Climatización -Desayuno en asientos preferentes -Bebida bienvenida en preferente -Prensa en preferente, gran clase y club -Vídeo por canal individual -Servicio cafetería
AVE	15:50	21:15	5h.25min.	Turista: 142€ Business: 256€	-Climatización -Merienda en asientos preferentes -Bebida bienvenida en preferente -Prensa en preferente, gran clase y club -Vídeo por canal individual -Servicio cafetería

PREGUNTAR:

- Recorrido: ¿solo ida? - ¿ida y vuelta? .
- Día
- Tipo de tren
- Precio: Pagar con *tarjeta de empresa*, con *tarjeta personal*, en *metálico*.
- Descuentos: solo ida, no descuentos – ida y vuelta, descuento del 10% en trenes normales, en AVE descuento del 20%.

VIAJERO B: Características

- Eres un hombre/mujer de negocios, tienes mucho dinero.
- hacer una reserva (comprar un billete)
- Viajar rápido. No necesitas dormir.
- No te importa el precio.
- Recorrido: Barcelona-Sevilla.
- Quieres todos los servicios:
 - *zona preferente
 - *restaurante
 - *enchufes para el portátil (laptop) y wifi

ANEXO III. Diseño de la actividad *role-play* 2: para entrenar

A continuación, te explicamos cómo funcionan en la actualidad los trenes en España.

Los **billetes** de tren se pueden comprar en la **taquilla** dentro de la estación de tren, en una **agencia de viajes** o por **Internet**.

Para ir de Barcelona a Granada se compra un **billete sencillo** Barcelona-Granada. Para ir de Barcelona a Granada y de Granada a Barcelona se compra un **billete de ida y vuelta**.

Para asegurarse de tener una **plaza** se puede comprar un billete con **antelación** o se puede hacer una **reserva**.

Cuando se hacen **largos recorridos**, se puede viajar durante el día o durante la noche.

- Si se viaja durante el día, se puede elegir ir sentado en **clase preferente** o **clase turista** (la opción más barata, pero no tan cómoda).
- Si se viaja por la noche, se va más cómodo en el **coche-cama**. Se puede elegir entre **cabins de camas** (se puede viajar en **gran clase** o en **clase preferente** y en **individual** o **doble**) y **coches de literas** (hasta cuatro personas en un departamento). También, se puede ir en una **plaza sentado**, pero no se va tan cómodo.

En los trenes de largo recorrido suele haber un **coche-bar** que es similar a una cafetería, con barra y un **coche-restaurant** que tiene mesas fijas para servir comidas completas.

Durante el viaje el **revisor** comprobará todos los billetes y, si el billete no es el correcto, pondrá una **multa**.

¿Qué tipos de trenes hay?

- **AVE** (=Alta Velocidad Española).
- Grandes líneas como el **Talgo**, **Intercity**, **Euromed**, etc. Son trenes de **largo recorrido** con una velocidad media-alta.
- **Regionales**: son trenes de cercanías que realizan largos y medios recorridos. Son más baratos que el Talgo, Intercity, Euromed, etc.
- **Cercanías**: son trenes que realizan cortos recorridos, muchas paradas, pasan frecuentemente y sólo hay una clase.

Avisos por los megáfonos de la estación

🔊 El Talgo **procedente de Bilbao** va a efectuar su **entrada** por la vía 2.

🔊 El Intercity **con destino Valencia** va a efectuar su **salida** por la vía 7.



1. *¿Cómo es el interior del AVE?* <<http://www.youtube.com/watch?v=Tsqq54jlg4>>
2. *Completa los siguientes ejercicios y practica el vocabulario aprendido*¹³⁰.

A. Mira el siguiente billete de tren y responde si es verdadero (V) o falso (F).



1. Es un billete de ida y vuelta Santiago de Compostela-Madrid.
2. Es un billete para el 17 de julio de 2009.
3. Es un billete de clase turista.
4. El billete se pagó con tarjeta de crédito.
5. El tren sale a las 21:08 de Santiago de Compostela.
6. El billete se compró por Internet.

B. Elige la respuesta correcta.

1. Para llegar lo antes posible a nuestro destino, tomamos_____
 - a) un talgo
 - b) un regional
2. Si quiero comprar un billete más barato, compro un billete en clase_____
 - a) preferente
 - b) turista
3. Somos un grupo de cuatro amigos y queremos dormir todos juntos; reservamos_____
 - a) una cabina de camas
 - b) un cabina de literas
4. El tren de Alta Velocidad Española se llama...
 - a) Euromed
 - b) AVE

¹³⁰ La información del sistema ferroviario —primera hoja— y el ejercicio 2#A y B está tomado de Bellver y San Andrés (2010), pp. 68-69. El resto es nuestra aportación.

3. COMPRAR UN BILLETE DE TREN

Ya conoces cómo funcionan *los trenes de larga distancia*. Así que, ahora, en parejas, imaginad que estáis en la estación de tren de Sants (Barcelona). Elegid uno de los dos roles (viajero o taquillero) y usad el vocabulario que habéis aprendido para comprar-vender un billete de tren.

VIAJERO: Quieres hacer una reserva. Estás en Barcelona y viajas a Granada para asistir a un concierto de tu grupo favorito que es a las 22h. Tienes que decidir:

→ ¿Vas solo y allí te encuentras con tus amigos o viajas con tus amigos?

→ ¿Recorrido:

-Barcelona-Granada, billete sencillo o Barcelona-Granada/Granada-Barcelona, billete de ida y vuelta?

-Barcelona-Málaga en el AVE (tren de alta velocidad), billete sencillo o billete de ida y vuelta y después Málaga-Granada en bus (no hay conexión de tren entre estas dos ciudades)?

→ ¿Viajas de día o de noche?

→ ¿Qué día y hora quieres viajar?

→ ¿Cuánto quieres gastar?

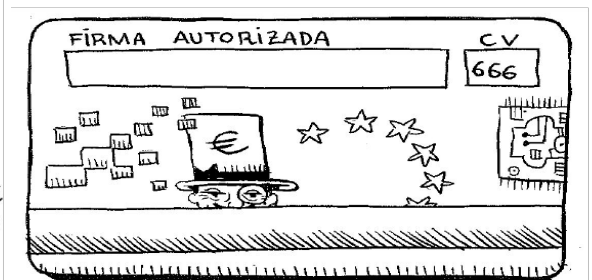
Te dejamos un esquema con los tipos de trenes para ayudarte cuando el taquillero te explique:

+ barato	Turista sentado	☀	Talgo (con enlace)
	Turista sentado	☾	Tren Hotel (directo)
	Preferente sentado	☀	Talgo (con enlace)
	Preferente sentado	☾	Tren Hotel (directo)
	Cama individual turista	☀	Talgo (con enlace)
	Cama preferente individual	☾	Tren Hotel (directo)
+ caro	Turista Sentado	☀	AVE
	Preferente sentado	☀	AVE

Esta es tu tarjeta de crédito. Completa con tus datos para poder pagar.



ESCRIBE
TU NOMBRE
Y APELLIDO



TAQUILLERO: Eres muy amable y quieres ayudar al cliente. Tienes que informar y preguntar al viajero para saber qué quiere.

→ Informa: hay tres tipos de trenes: Talgo (con enlace), Tren hotel y AVE.



→ El viajero quiere: ¿solo ida? o ¿ida y vuelta? ¿Tren directo o no directo? (Recuerda, con el Talgo hay que cambiar de tren en una estación pequeña, en el campo, llamada Linares y hay una espera de 15 minutos).

→ Comprueba: ¿Qué día quiere viajar?


→ ¿Precio? Solo ida, no hay descuentos – Ida y vuelta, sí hay descuento: del 10%.

→ Tienen que pagar obligatoriamente con la tarjeta de crédito.

BARCELONA-GRANADA

TREN/RECORRIDO	SALIDA	LLEGADA	DURACIÓN	PRECIOS	PRESTACIONES
1. Talgo (Con enlace) 1er. trayecto <i>Bcn-Linares</i> (8:30h-16:35h) 2º trayecto <i>Linares-Granada</i> (16: 45h-19:08h)	8:30 	19:08	10h.38min.	Turista: 61€ Preferente: 80 €	-Climatización -Servicio cafetería y restaurante
				Cama individual turista (cabina literas 4 camas): 92€	-Climatización -Servicio cafetería y restaurante
2. Tren hotel (directo) <i>Bcn-Granada</i>	21:30 	08:05	10h.35min.	Turista (sentado): 64€	-Climatización -Servicio cafetería y restaurante
				Preferente (sentado): 85€	- Bebida de bienvenida y prensa -Incluida cena y desayuno en coche restaurante
				Cama individual turista (cabina literas 4 camas): 92€	-Climatización -Servicio cafetería y restaurante
				Cama preferente individual o doble: 166€	- Bebida de bienvenida y prensa -Incluida cena y desayuno en coche restaurante

BARCELONA-MÁLAGA-GRANADA

TREN/RECORRIDO	SALIDA	LLEGADA	DURACIÓN	PRECIOS	PRESTACIONES
3. AVE <i>Barcelona-Málaga*</i>	8:30 	14:24	5h.54min.	Turista: 146€	-Climatización -Vídeo y música por canal individual -Servicio cafetería y restaurante
				Preferente: 219 €	-Climatización -Vídeo y música por canal individual -Prensa en preferente -Servicio cafetería y restaurante -Almuerzo para clientes con billete preferente

* De Málaga a Granada no hay conexión de tren. Así que la mejor opción es ir en autobús que sale cada hora de las 9h a las 23h. Tarda dos horas y cuesta el billete 12 euros.

ANEXO IV. Cuestionario contacto lingüístico: institución y alumnos

NECESITAMOS VOLUNTARIOS PARA UN ESTUDIO SOBRE EL PROCESO DE COMUNICACIÓN EN ESPAÑOL EN LA SITUACIÓN DE COMPRAR BILLETES DE TREN

Se trata de un proyecto de investigación dentro del programa de doctorado Lingüística Teórica, Computacional y Aplicada, Departamento de Lingüística General, Universitat de Barcelona.

Esto es una solicitud de colaboración entre la UB (CLiC-UB), UPF y su institución para su proyecto de investigación en lingüística aplicada.

Se está llevando a cabo un estudio interlingüístico sobre diálogos para modelizarlos informáticamente. El ámbito es la negociación en viajes (compra-venta, solicitud de información, condiciones, etc.). Necesitamos muestras de interacciones lo más reales posible, dentro de la simulación, entre hablantes nativos y no nativos. Por esta razón, solicitamos la colaboración de centros docentes con alumnos internacionales, además de la de centros locales, para elaborar el modelo.

La participación de los alumnos implica:

- 1) Desplazarse a la UB a realizar una conversación telefónica que se grabará y de la que pueden tener copia, tanto el alumno como la institución.
- 2) El tiempo estimado de cada entrevista es de 15-30 minutos.
- 3) No se remunera la colaboración del alumno ni a la institución colaboradora, pero constara en la "tabula gratulatoria" o créditos de los trabajos que se deriven. Asimismo, no duden en que si de la sistematización resultante, deriva material útil para la docencia, se les hará entrega como reconocimiento.
- 4) Cuestionario que se administrará a los alumnos para solicitar voluntarios:
Dado que no se puede indicar el objeto investigado al alumno porque alteraría los datos, hemos previsto una situación ficticia, inocua, que es la de proponer que el objetivo es *entrenar a futuros operadores* de empresas de transporte. Esto evita que se sientan "analizados", "examinados" o "juzgados".
Les adjuntamos, para su consideración institucional, el cuestionario. En caso de que hubiera algún cambio que ustedes estimaran oportuno, estaremos gustosos de asumirlo.

Muchas gracias por su colaboración en nombre del equipo de las dos instituciones (UB y UPF).

NECESITAMOS VOLUNTARIOS PARA UN TRAINING DE OPERADORES TELEFÓNICOS DE COMPAÑÍAS FERROVIARIAS (tren): ASISTENCIA AL VIAJERO

Se trata de un proyecto de investigación dentro del programa de doctorado Lingüística Teórica, Computacional y Aplicada, Departamento de Lingüística General, Universitat de Barcelona.

SOLICITUD DE VOLUNTARIOS

Si decides colaborar con nosotros, la actividad tiene las siguientes características:

1. ¿Qué actividad?

Pedir información de trenes por teléfono: hacer preguntas y contestar preguntas cortas. Grabaremos todo en audio para entrenar a los operadores. El tiempo total que necesitaremos en la actividad será entre 15'- 30' máximo.

2. ¿Cuándo y dónde?

- Quedaremos una vez.
- Se hará entre marzo y mayo.
- Llamaréis por teléfono desde la Universidad de Barcelona, departamento de Lingüística (Plaza Universidad, Metro Línea 1 o roja) y la llamada será de 15' a 30'.
- Horario previsto para la actividad: se enviará un e-mail con vuestra cita.

3. Contraprestación: ¿Qué obtengo yo a cambio?

- Puedes tener una copia de la conversación.
- Bebida refrescante, galletas, bizcocho, cafés... durante la colaboración.

4. ¿Con quién hablarás?

Con un trabajador de la compañía de trenes.

5. Garantía del anonimato de los voluntarios y su imagen en la recogida de datos

Cada voluntario tendrá un código de identificación que usaremos en lugar de su nombre para garantizar el anonimato. El material recogido llevará solamente el código como identificación y ningún otro dato personal.

¡MUY IMPORTANTE!

La actividad y el resultado no influyen en vuestro curso habitual, en el que estáis matriculados. NO son actividades complementarias del curso, NO suben nota, NO están relacionadas con el programa. Solo os necesitamos con vuestra experiencia de clientes viajeros.

Por eso, todos los estudiantes de todos los niveles que participan en el proyecto hacen la misma actividad, independientemente del nivel, curso o escuela a la que pertenezcan. Así, si encontráis la actividad muy difícil o muy fácil, TRANQUILOS. Se trata de "viajar".

Si queréis una copia del material grabado, pedidla.

Si queréis que aparezca vuestro nombre en el apartado de agradecimientos del informe final, marcadlo a continuación:

Sí No

Por último, contesta la siguiente información para poder ponernos en contacto contigo y empezar. No olvides escribir tu e-mail.

!!!!MUCHAS GRACIAS!!!!

BLOQUE 1. PERFIL GENERAL

1. Nombre: hombre mujer
2. Edad:
3. Nacionalidad:
¿Dónde naciste (ciudad)?
4. Dirección de e-mail:
5. ¿En qué lugar te educaste (brought up) –ciudad, país:-
6. ¿Qué has estudiado?
 primaria secundaria universidad (si has marcado esta opción, contesta la pregunta 6.1)
- 6.1. ¿Qué especialidad has estudiado en la universidad?
- ¿Dónde? (universidad)
- ¿Cuántos años de universidad has estudiado?
- ¿Has terminado tus estudios universitarios?
 Sí Todavía no

BLOQUE 2. PERFIL LINGÜÍSTICO

1. Lengua materna del padre:
- Lengua materna de la madre:

2. Otras lenguas que aprendiste en orden cronológico: usa el cuadro de abajo. Si estás todavía aprendiendo alguna lengua, escribe "todavía aprendiendo".

Lenguas	Año que empecé	Año que terminé	Muy bien	Bien	Regular	Mal
a.						
b.						
c.						
d.						
e.						

3. ¿Cuánto tiempo vas a estudiar español en tu escuela de Barcelona (EOI, Arcadia University, PEI,...)?
 1 curso (3 o 4 meses)
 2 cursos (6 o 8 meses)
 más de dos cursos
 meses:
4. ¿Piensas quedarte en Barcelona permanentemente? Sí No

5. Voy a quedarme en Barcelona:

- 4 meses 5 meses 1 año 2 años más de 2 años
 Otros (indica)

BLOQUE 2.1. CONTACTO

1. ¿Usualmente...?

- Vivo con españoles o latinoamericanos.
 Tengo un/a novio/a español/a o latinoamericano/a.
 Trabajo con españoles o latinoamericanos.
 Tengo amigos españoles o latinoamericanos.
 No hablo casi nunca con españoles ni latinoamericanos.
 Hablo solo con extranjeros, excepto en clase.

2. ¿Cuánto tiempo dedicas a las siguientes actividades?

nunca	un poco	bastante	mucho	muchísimo	
0	1	2	3	4	
					a. leer en español: periódicos, revistas, cómics
					b. leer en español: novelas, ensayos
					c. ver en español: películas, links en youtube
					d. ver en español: televisión
					e. escuchar en español: música y canciones
					f. hablar en español con personas que conoces (vecinos, compañeros de universidad, otros)
					g. hablar en español con personas que NO conoces

BLOQUE 3. INFORMACIÓN DE VIAJES

1. ¿Viajas mucho? No Sí (si has marcado esta opción, contesta el siguiente cuadro)

	EN TU PAÍS	EN EUROPA	EN ESPAÑA
1. VIAJAS en transporte público (bus, metro, autocar –URBANOS–)	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
2. VIAJAS en tren, pullman, voyageur... –INTERURBANOS–	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
2.1. ¿A dónde?			
2.2. ¿Ida y vuelta (return ticket)?			
2.3. ¿Solo ida o solo vuelta (one way)?			
2.4. ¿Quién compra el billete, tú o un amigo?			
2.5. ¿Compras billetes/tickets...?	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)
3. VIAJAS en avión	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No	<input type="checkbox"/> Sí <input type="checkbox"/> No
3.1. ¿Compras billetes/tickets...?	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)	<input type="checkbox"/> por teléfono <input type="checkbox"/> por internet <input type="checkbox"/> en ventanilla (desk-office)

BLOQUE 4. VOLUNTARIADO

1. Los operarios de información de transportes:

• ¿Te parecen amables y colaboradores en España? Sí No

• ¿Te parece que ayudan a los extranjeros? Sí No

• ¿Crees que pueden mejorar?

Sí. ¿Cómo?

No

• ¿Quieres ayudar a entrenarlos como extranjeros? Sí No

ANEXO V. Nociones simples

Véase el cd que se incluye.

ANEXO VI. Nociones combinadas

Véase el cd que se incluye.

ANEXO VII. Un diálogo en diferentes formatos: diálogo *FerrovieLE-L_trained-multiling_1*

VII.a. Formato word

C8: Hola señor.

O8: Hola, ¿puedo ayudarte?

C8: Sí, es que estoy un poco confundido sobre los tipos de billetes que puedo comprar. Me gustaría viajar a Granada, ¿qué es la manera más fácil?

O8: Vale, hay tres trenes hoy a Granada. ¿Prefieres un tren directo?

C8: Eh, sí por favor.

O8: Eh hay un tren directo esta noche, a las nueve y media.

C8: ¡Ah, pues, sí gracias! ¿Pero hay distintos tipos de billetes, no?

O8: Sí, eh ¿quieres un billete turista, preferente o en en cama?

C8: ¡Ah, vale entiendo! Entonces, un billete turista, por favor.

O8: Eh, bueno eh cuesta sesenta y cuatro euros.

C8: ¡Uf! ¿No hay descuentos para estudiantes?

O8: Eh, no hay descuentos en los billetes de ida, pero te puedo ofrecer un descuento de diez por ciento si compras un billete de ida y vuelta.

C8: ¡Ah, vale! Por eso dame una ida y vuelta, por favor. ¿Puedo pagar en efectivo?

O8: Eh, no. Tienes que pagar con la tarjeta de crédito.

C8: ¡Agh! Pues, sólo tengo mi tarjeta británica. Eh ¿ésta funciona?

O8: Ah sí, eh simplemente ingresas su código pen.

C8: ¡Ah vale!, gracias.

O8: Exacto, adiós.

C8: Adiós, gracias por la ayuda.

O8: Espero que tengas un buen viaje.

C8: Gracias.

VII.b. Formato html

Program of
Transcribed by Mònica, version 9 of 150105

report - Initiating conversation

C8

- Hola señor.

report - Initiating conversation

O8

- Hola, [*lex=change function-*]

report - Offer

O8

- *[-lex=change function]* ¿puedo ayudarte?

report - Accepting

C8

- Sí, *[lex=change function-]*

report - Expressing reason, cause, excuse

C8

- *[-lex=change function]* es que estoy un poco confundido sobre los tipos de billetes que puedo comprar. *[lex=change function-]*

report - Request a service: buying a ticket

C8

- *[-lex=change function]* Me gustaría viajar a Granada, ¿qué es la manera más fácil?

report - Confirming

O8

- Vale, *[lex=change function-]*

report - Giving precise information

O8

- *[-lex=change function]* *[ent=time-]* hay tres trenes hoy *[-ent=time]* *[ent=loc.destination.place-]* a Granada. *[-ent=loc.destination.place]* *[lex=change function-]*

report - Enquiring about train trajet

O8

- *[-lex=change function]* ¿Prefieres un tren directo?

report - Accepting

C8

- Eh, sí por favor.

report - Giving precise information

O8

- *[ent=prod.train.traject-]* Eh hay un tren directo *[-ent=prod.train.traject]*
[ent=time-] esta noche, *[-ent=time]* *[ent=time.departure-]* a las nueve y media.
[-ent=time.departure]

report - Accepting

C8

- ¡Ah, pues, sí gracias! *[lex=change function-]*

report - Enquiring about fare

C8

- *[-lex=change function]* ¿Pero hay distintos tipos de billetes, *[lex=change function-]*

report - Inviting agreement

C8

- *[lex=change function]* no?

report - Confirming

O8

- Sí, *[lex=change function-]*

report - Enquiring about fare

O8

- *[-lex=change function]* eh ¿quieres un billete turista, preferente o en en cama?

report - Accepting

C8

- ¡Ah, vale entiendo! [*lex=change function-*]

report - Expressing preference: fare

C8

- [*-lex=change function*] Entonces, un billete turista, por favor.

report - Giving precise information: price

O8

- Eh, bueno eh cuesta sesenta y cuatro euros.

report - Expressing reaction

C8

- ¡Uf! [*lex=change function-*]

report - Enquiring about fare

C8

- [*-lex=change function*] ¿No hay descuentos para estudiantes?

report - Giving precise information: fare

O8

- Eh, no hay descuentos en los billetes de ida, pero te puedo ofrecer un descuento de diez por ciento si compras un billete de ida y vuelta.

report - Accepting

C8

- ¡Ah, vale! [*ent=prod.buying.a.ticket-*] Por eso dame [*lex=check_short distance-*] una ida y vuelta, por favor. [*-lex=check_short distance*] [*-ent=prod.buying.a.ticket*] [*lex=change function-*]

report - Enquiring about payment conditions

C8

- *[-lex=change function]* ¿Puedo pagar en efectivo?

report - Refusing

O8

- Eh, no. *[lex=change function-]*

report - Obligation and necessity

O8

- *[-lex=change function]* Tienes que pagar con la tarjeta de crédito.

report - Expressing reaction

C8

- ¡Agh! *[lex=change function-]*

report - Giving information: credit card

C8

- *[-lex=change function]* Pues, sólo tengo mi tarjeta británica. *[lex=change function-]*

report - Asking for confirmation or denial

C8

- *[-lex=change function]* Eh ¿ésta funciona?

report - Confirming

O8

- Ah sí, *[lex=change function-]*

report - Giving precise information: credit card_PIN

O8

- [*-lex=change function*] eh simplemente ingresas su código pen.

report - Accepting

C8

- ¡Ah vale!, gracias.

report - Confirming

O8

- Exacto, [*lex=change function-*]

report - Closing conversation

O8

- [*-lex=change function*] adiós.

report - Closing conversation

C8

- Adiós, gracias por la ayuda.

report - Socialising

O8

- Espero que tengas un buen viaje.

report - Socialising

C8

- Gracias.

VII.c. Formato txt

C8: Hola señor.

O8: Hola, [lex=change function-]

O8: [-lex=change function] ¿puedo ayudarte?

C8: Sí, [lex=change function-]

C8: [-lex=change function] es que estoy un poco confundido sobre los tipos de billetes que puedo comprar. [lex=change function-]

C8: [-lex=change function] Me gustaría viajar a Granada, ¿qué es la manera más fácil?

O8: Vale, [lex=change function-]

O8: [-lex=change function] [ent=time-] hay tres trenes hoy [-ent=time]

[ent=loc.destination.place-] a Granada. [-ent=loc.destination.place] [lex=change function-]

O8: [-lex=change function] ¿Prefieres un tren directo?

C8: Eh, sí por favor.

O8: [ent=prod.train.traject-] Eh hay un tren directo [-ent=prod.train.traject] [ent=time-] esta noche, [-ent=time] [ent=time.departure-] a las nueve y media. [-ent=time.departure]

C8: ¡Ah, pues, sí gracias! [lex=change function-]

C8: [-lex=change function] ¿Pero hay distintos tipos de billetes, [lex=change function-]

C8: [lex=change function] no?

O8: Sí, [lex=change function-]

O8: [-lex=change function] eh ¿quieres un billete turista, preferente o en en cama?

C8: ¡Ah, vale entiendo! [lex=change function-]

C8: [-lex=change function] Entonces, un billete turista, por favor.

O8: Eh, bueno eh cuesta sesenta y cuatro euros.

C8: ¡Uf! [lex=change function-]

C8: [-lex=change function] ¿No hay descuentos para estudiantes?

O8: Eh, no hay descuentos en los billetes de ida, pero te puedo ofrecer un descuento de diez por ciento si compras un billete de ida y vuelta.

C8: ¡Ah, vale! [ent=prod.buying.a.ticket-] Por eso dame [lex=check_short distance-] una ida y vuelta, por favor. [-lex=check_short distance] [-ent=prod.buying.a.ticket] [lex=change function-]

C8: [-lex=change function] ¿Puedo pagar en efectivo?

O8: Eh, no. [lex=change function-]

O8: [-lex=change function] Tienes que pagar con la tarjeta de crédito.

C8: ¡Agh! [lex=change function-]

C8: [-lex=change function] Pues, sólo tengo mi tarjeta británica. [lex=change function-]

C8: [-lex=change function] Eh ¿ésta funciona?

O8: Ah sí, [lex=change function-]

O8: [-lex=change function] eh simplemente ingresas su código pen.

C8: ¡Ah vale!, gracias.

O8: Exacto, [lex=change function-]

O8: [-lex=change function] adiós.

VII.d. Formato xml

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Trans SYSTEM "trans-14.dtd">
<Trans scribe="Mònica" audio_filename="FerroviELE-A_trained-multiling_1"
version="9" version_date="150105">
<Topics>
<Topic id="to1" desc="Initiating conversation"/>
<Topic id="to2" desc="Offer"/>
<Topic id="to3" desc="Accepting"/>
<Topic id="to4" desc="Expressing reason, cause, excuse"/>
<Topic id="to5" desc="Request a service: buying a ticket"/>
<Topic id="to6" desc="Confirming"/>
<Topic id="to7" desc="Giving precise information"/>
<Topic id="to8" desc="Enquiring about train trajet"/>
<Topic id="to9" desc="Enquiring about fare"/>
<Topic id="to10" desc="Inviting agreement"/>
<Topic id="to11" desc="Expressing preference: fare"/>
<Topic id="to12" desc="Giving precise information: price"/>
<Topic id="to13" desc="Expressing reaction"/>
<Topic id="to14" desc="Giving precise information: fare"/>
<Topic id="to15" desc="Enquiring about payment conditions"/>
<Topic id="to16" desc="Refusing"/>
<Topic id="to17" desc="Obligation and necessity"/>
<Topic id="to18" desc="Giving information: credit card"/>
<Topic id="to19" desc="Asking for confirmation or denial"/>
<Topic id="to20" desc="Giving precise information: credit card_PIN"/>
<Topic id="to21" desc="Closing conversation"/>
<Topic id="to22" desc="Socialising"/>
</Topics>
<Speakers>
<Speaker id="spk1" name="C8" check="no" type="male" dialect="nonnative"
accent="English and German_B2 Level" scope="local"/>
<Speaker id="spk2" name="O8" check="no" type="male" dialect="nonnative"
accent="English and German_B2 Level" scope="local"/>
</Speakers>
<Episode>
<Section type="report" startTime="0" endTime="3.385" topic="to1">
<Turn startTime="0" endTime="3.385" speaker="spk1" mode="planned"
fidelity="high" channel="studio">
<Sync time="0"/>
Hola señor.
</Turn>
</Section>
<Section type="report" topic="to1" startTime="3.385" endTime="3.982">
<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="3.385" endTime="3.982">
```

<Sync time="3.385"/>

Hola,

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to2" startTime="3.982" endTime="5.818">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio" startTime="3.982" endTime="5.818">

<Sync time="3.982"/>

<Event desc="change function" type="lexical" extent="end"/>

¿puedo ayudarte?

</Turn>

</Section>

<Section type="report" topic="to3" startTime="5.818" endTime="6.131">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="5.818" endTime="6.131">

<Sync time="5.818"/>

Sí,

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to4" startTime="6.131" endTime="10.701">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="6.131" endTime="10.701">

<Sync time="6.131"/>

<Event desc="change function" type="lexical" extent="end"/>

es que estoy un poco confundido sobre los tipos de billetes que puedo comprar.

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to5" startTime="10.701" endTime="14.344">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="10.701" endTime="14.344">

<Sync time="10.701"/>

<Event desc="change function" type="lexical" extent="end"/>

Me gustaría viajar a Granada, ¿qué es la manera más fácil?

</Turn>

</Section>

<Section type="report" topic="to6" startTime="14.344" endTime="15.0">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio" startTime="14.344" endTime="15.0">

<Sync time="14.344"/>
Vale,
<Event desc="change function" type="lexical" extent="begin"/>

</Turn>
</Section>
<Section type="report" topic="to7" startTime="15.0" endTime="17.172">
<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="15.0" endTime="17.172">
<Sync time="15.0"/>

<Event desc="change function" type="lexical" extent="end"/>

<Event desc="time" type="entities" extent="begin"/>
hay tres trenes hoy
<Event desc="time" type="entities" extent="end"/>

<Event desc="loc.destination.place" type="entities" extent="begin"/>
a Granada.
<Event desc="loc.destination.place" type="entities" extent="end"/>

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>
</Section>
<Section type="report" topic="to8" startTime="17.172" endTime="19.422">
<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="17.172" endTime="19.422">
<Sync time="17.172"/>

<Event desc="change function" type="lexical" extent="end"/>
¿Prefieres un tren directo?
</Turn>
</Section>
<Section type="report" topic="to3" startTime="19.422" endTime="21.5">
<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="19.422" endTime="21.5">
<Sync time="19.422"/>
Eh, sí por favor.
</Turn>
</Section>
<Section type="report" topic="to7" startTime="21.5" endTime="24.889">
<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="21.5" endTime="24.889">
<Sync time="21.5"/>

<Event desc="prod.train.traject" type="entities" extent="begin"/>

Eh hay un tren directo
 <Event desc="prod.train.traject" type="entities" extent="end"/>

<Event desc="time" type="entities" extent="begin"/>
 esta noche,
 <Event desc="time" type="entities" extent="end"/>

<Event desc="time.departure" type="entities" extent="begin"/>
 a las nueve y media.
 <Event desc="time.departure" type="entities" extent="end"/>

</Turn>
 </Section>
 <Section type="report" topic="to3" startTime="24.889" endTime="26.583">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="24.889" endTime="26.583">
 <Sync time="24.889"/>
 ¡Ah, pues, sí gracias!
 <Event desc="change function" type="lexical" extent="begin"/>

</Turn>
 </Section>
 <Section type="report" topic="to9" startTime="26.583" endTime="28.868">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="26.583" endTime="28.868">
 <Sync time="26.583"/>

<Event desc="change function" type="lexical" extent="end"/>
 ¿Pero hay distintos tipos de billetes,
 <Event desc="change function" type="lexical" extent="begin"/>

</Turn>
 </Section>
 <Section type="report" topic="to10" startTime="28.868" endTime="29.78">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="28.868" endTime="29.78">
 <Sync time="28.868"/>

<Event desc="change function" type="lexical" extent="instantaneous"/>
 no?
 </Turn>
 </Section>
 <Section type="report" topic="to6" startTime="29.78" endTime="30.068">
 <Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
 startTime="29.78" endTime="30.068">
 <Sync time="29.78"/>
 Sí,

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to9" startTime="30.068" endTime="34.046">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio" startTime="30.068" endTime="34.046">

<Sync time="30.068"/>

<Event desc="change function" type="lexical" extent="end"/>

eh ¿quieres un billete turista, preferente o en en cama?

</Turn>

</Section>

<Section type="report" topic="to3" startTime="34.046" endTime="35.045">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="34.046" endTime="35.045">

<Sync time="34.046"/>

¡Ah, vale entiendo!

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to11" startTime="35.045" endTime="38.345">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="35.045" endTime="38.345">

<Sync time="35.045"/>

<Event desc="change function" type="lexical" extent="end"/>

Entonces, un billete turista, por favor.

</Turn>

</Section>

<Section type="report" topic="to12" startTime="38.345" endTime="41.095">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio" startTime="38.345" endTime="41.095">

<Sync time="38.345"/>

Eh, bueno eh cuesta sesenta y cuatro euros.

</Turn>

</Section>

<Section type="report" topic="to13" startTime="41.095" endTime="42.767">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio" startTime="41.095" endTime="42.767">

<Sync time="41.095"/>

¡Uf!

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to9" startTime="42.767" endTime="45.457">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="42.767" endTime="45.457">
 <Sync time="42.767"/>

<Event desc="change function" type="lexical" extent="end"/>
 ¿No hay descuentos para estudiantes?
 </Turn>
 </Section>

<Section type="report" topic="to14" startTime="45.457" endTime="53.806">
 <Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
 startTime="45.457" endTime="53.806">
 <Sync time="45.457"/>
 Eh, no hay descuentos en los billetes de ida, pero te puedo ofrecer un descuento de diez por ciento si compras un billete de ida y vuelta.
 </Turn>
 </Section>

<Section type="report" topic="to3" startTime="53.806" endTime="58.413">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="53.806" endTime="58.413">
 <Sync time="53.806"/>
 ¡Ah, vale!
 <Event desc="prod.buying.a.ticket" type="entities" extent="begin"/>
 Por eso dame
 <Event desc="check_short distance" type="lexical" extent="begin"/>
 una ida y vuelta, por favor.
 <Event desc="check_short distance" type="lexical" extent="end"/>

<Event desc="prod.buying.a.ticket" type="entities" extent="end"/>

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>
 </Section>

<Section type="report" topic="to15" startTime="58.413" endTime="59.963">
 <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
 startTime="58.413" endTime="59.963">
 <Sync time="58.413"/>

<Event desc="change function" type="lexical" extent="end"/>
 ¿Puedo pagar en efectivo?
 </Turn>
 </Section>

<Section type="report" topic="to16" startTime="59.963" endTime="61.014">
 <Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
 startTime="59.963" endTime="61.014">
 <Sync time="59.963"/>

Eh, no.

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to17" startTime="61.014" endTime="63.609">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="61.014" endTime="63.609">

<Sync time="61.014"/>

<Event desc="change function" type="lexical" extent="end"/>

Tienes que pagar con la tarjeta de crédito.

</Turn>

</Section>

<Section type="report" topic="to13" startTime="63.609" endTime="64.204">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="63.609" endTime="64.204">

<Sync time="63.609"/>

¡Agh!

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to18" startTime="64.204" endTime="66.942">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="64.204" endTime="66.942">

<Sync time="64.204"/>

<Event desc="change function" type="lexical" extent="end"/>

Pues, sólo tengo mi tarjeta británica.

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to19" startTime="66.942" endTime="68.475">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="66.942" endTime="68.475">

<Sync time="66.942"/>

<Event desc="change function" type="lexical" extent="end"/>

Eh ¿ésta funciona?

</Turn>

</Section>

<Section type="report" topic="to6" startTime="68.475" endTime="69.25">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="68.475" endTime="69.25">

<Sync time="68.475"/>

Ah sí,

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to20" startTime="69.25" endTime="71.779">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="69.25" endTime="71.779">

<Sync time="69.25"/>

<Event desc="change function" type="lexical" extent="end"/>

eh simplemente ingresas su código pen.

</Turn>

</Section>

<Section type="report" topic="to3" startTime="71.779" endTime="73.176">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="71.779" endTime="73.176">

<Sync time="71.779"/>

¡Ah vale!, gracias.

</Turn>

</Section>

<Section type="report" topic="to6" startTime="73.176" endTime="73.844">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="73.176" endTime="73.844">

<Sync time="73.176"/>

Exacto,

<Event desc="change function" type="lexical" extent="begin"/>

</Turn>

</Section>

<Section type="report" topic="to21" startTime="73.844" endTime="74.826">

<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="73.844" endTime="74.826">

<Sync time="73.844"/>

<Event desc="change function" type="lexical" extent="end"/>

adiós.

</Turn>

</Section>

<Section type="report" topic="to21" startTime="74.826" endTime="76.244">

<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="74.826" endTime="76.244">

<Sync time="74.826"/>

Adiós, gracias por la ayuda.

</Turn>

</Section>

<Section type="report" topic="to22" startTime="76.244" endTime="78.211">


```
<Turn speaker="spk2" mode="planned" fidelity="high" channel="studio"
startTime="76.244" endTime="78.211">
<Sync time="76.244"/>
Espero que tengas un buen viaje.
</Turn>
</Section>
<Section type="report" topic="to22" startTime="78.211" endTime="79.882">
<Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
startTime="78.211" endTime="79.882">
<Sync time="78.211"/>
Gracias.
</Turn>
</Section>
</Episode>
</Trans>
```

VII.e. Análisis morfológico

Véase el cd que se incluye.

ANEXO VIII. Un ejemplo de diálogo de cada subcorpus en formato html

Véase el cd que se incluye.

ANEXO IX. Inventario de funciones y nociones del PCIC

1. Dar y pedir información

1.1. Identificar

1.2. Pedir información

1.2.1. Proponiendo alternativas

1.2.2. Expresando curiosidad

1.2.3. Solicitando una explicación

1.2.4. Con prudencia

1.3. Dar información

1.3.1. Corrigiendo otra información previa

1.4. Pedir confirmación

1.4.1. Cuestionando la información

1.5. Confirmar la información previa

1.6. Describir

2. Expresar opiniones, actitudes y conocimientos

2.1. Pedir opinión

2.2. Dar una opinión

(...)

2.8. Invitar al acuerdo

2.9. Expresar acuerdo

2.10. Expresar desacuerdo

(...)

2.15. Invitar a formular una hipótesis

2.16. Expresar posibilidad

2.17. Expresar obligación y necesidad

2.19. Preguntar por el conocimiento de algo

2.20. Expresar conocimiento

2.21. Expresar desconocimiento

3. Expresar gustos, deseos y sentimientos

3.1. Preguntar por gustos e intereses

3.2. Expresar gustos e intereses

(...)

3.4. Preguntar por preferencias

3.5. Expresar preferencia

3.6. Expresar indiferencia o ausencia de preferencia

3.7. Preguntar por deseos

3.8. Expresar deseos

3.9. Preguntar por planes e intenciones

3.10. Expresar planes e intenciones

4. Influir en el interlocutor

4.1. Dar una orden o instrucción

4.2. Pedir un favor

4.3. Pedir objetos

4.4. Pedir ayuda

4.5. Rogar

(...)

4.13. Proponer y sugerir

4.14. Ofrecer e invitar

4.15. Solicitar confirmación de una propuesta previa

4.16. Aceptar una propuesta, ofrecimiento o invitación

4.17. Rechazar una propuesta, ofrecimiento o invitación

4.18. Aconsejar

5. Relacionarse socialmente

5.1. Saludar

5.2. Responder a un saludo

(...)

5.10. Disculparse

5.11. Responder a una disculpa

5.12. Agradecer

5.13. Responder a un agradecimiento

(...)

<p>5.17. Formular buenos deseos</p> <p>5.18. Responder a felicitaciones y buenos deseos</p> <p>(...)</p> <p>5.21. Despedirse</p> <p>6. Estructurar el discurso</p> <p>(...)</p> <p>6.9. Indicar que se sigue el relato con interés</p> <p>6.14. Reformular lo dicho</p> <p>6.16. Introducir palabras de otros</p> <p>6.21. Interrumpir</p> <p>6.25. Indicar que se desea continuar el discurso</p>
--

FIGURA 40. *Inventario de las funciones del PCIC (2007)*

14.3. Sistema de transporte	
B1	B2
<p>14.3.1. Red de transportes [v. Saberes y comportamientos socioculturales 1.13.3.]</p> <ul style="list-style-type: none"> • autopista, vía, andén, túnel • semáforo, paso ~ de cebra/de peatones • carril ~ bus/bici • haber ~ caravana/(un) atasco/un embotellamiento • ir/salir/llegar ~ con retraso, parar <p>14.3.2. Tipos de transporte [v. Saberes y comportamientos socioculturales 1.13.4.]</p> <ul style="list-style-type: none"> -viajero -tranvía, moto, camión -asiento, fila, pasillo -maletero, cinturón 	<p>14.3.1. Red de transportes [v. Saberes y comportamientos socioculturales 1.13.3.]</p> <ol style="list-style-type: none"> 5. autovía, carretera ~ nacional/comarcal/principal/secundaria, autopista de peaje 6. curva, cruce, rotonda, bache 7. área de descanso, puesto de socorro, peaje 8. autobús ~ urbano/interurbano 9. IBERIA, RENFE (Red Nacional de Ferrocarriles Españoles) 10. haber ~ (mucho/poco) ~ tráfico/circulación/movimiento 11. cruzar la frontera, pasar la aduana, declarar 12. pasar un control ~ de carreteras/de aduana

<ul style="list-style-type: none"> -conducir, aparcar, parar, girar -subir(se) a/montar(se) en/bajar(se) de ~ una bicicleta/una moto/un coche -hacer transbordo, perder ~ el tren/el autobús • Transporte marítimo y fluvial <ul style="list-style-type: none"> -barca, piragua -marinero, capitán -navegar -subir(se) a/bajar(se) de ~ un barco • Transporte aéreo <ul style="list-style-type: none"> -helicóptero, globo, ala delta, parapente -terminal, punto ~ de información/de encuentro -tarjeta/puerta ~ de embarque -sala ~ de llegadas/de embarque -asiento, ventanilla -despegar, aterrizar, hacer escala -facturar/recoger ~ el equipaje/las maletas -subir(se) a/bajar(se) de ~ un avión (...) 	<p>14.3.2. Tipos de transporte</p> <ul style="list-style-type: none"> • Transporte por tierra <p>[v. Saberes y comportamientos socioculturales 1.13.4.]</p> <ul style="list-style-type: none"> -pasajero, peatón -coche ~ de carreras/oficial/particular -(número de) matrícula -compartimento, vagón -arrancar, acelerar, frenar, adelantar -poner el intermitente, tocar el claxon -circular en coche, llevar una moto, manejar <p>Transporte marítimo y fluvial</p> <ul style="list-style-type: none"> -barco ~ de pasajeros/de vela/pesquero, ferry -tripulación, pasajero -marítimo -hacer escala, (des)embarcar -hacer un crucero <p>Transporte aéreo</p> <ul style="list-style-type: none"> -tripulación, pasajero -puente/transporte/accidente ~ aéreo -vuelo ~ chárter/directo/con escalas/ (inter)nacional/de líneas regulares -pilotar -anunciar/cancelar/retrasar ~ un vuelo -arreglo, reparación, revisión ~ anual/periódica/a fondo -cambiar ~ el aceite/una rueda/una pieza -arreglar un pinchazo -tener garantía (...)
--	--

FIGURA 41. *Inventario de las nociones específicas para el ámbito de Viajes, alojamiento y transporte (II) del PCIC (2007)*