

Dialectologia i recursos informàtics

a cura de

Maria Pilar Perea



DIALECTOLOGIA
I
RECURSOS INFORMÀTICS

Actes de l'11è Col·loqui Lingüístic
de la Universitat de Barcelona
i
de la IX Jornada
sobre la Variació Lingüística

DIALECTOLOGIA

I

RECURSOS INFORMÀTICS

A cura de
Maria Pilar Perea



Secció de Lingüística Catalana
Departament de Filologia Catalana



Barcelona, 2004

Primera edició, 2004

La reproducció total o parcial d'aquesta obra per qualsevol procediment, compresos la reprografia i el tractament informàtic, i la distribució d'exemplars mitjançant lloguer o préstec públics, resta rigorosament prohibida sense l'autorització escrita dels titulars del Copyright, i estarà sotmesa a les sancions establertes a la Llei.

© dels textos respectius: Gotzon Aurrekoetxea, Pilar García Mouton, Esteve Clua, Hans Goebel, Jaume Corbera, Maria Pilar Perea, Jacques Durand, Michel Aurnague, Julien Eychenne, Miquel Àngel Pradilla, Josefina Carrera, Ana M. Fernández, Joan Julià i Imma Creus.

© de l'edició: Maria Pilar Perea
(Secció de Lingüística Catalana, Departament de Filologia Catalana,
Universitat de Barcelona)

COL·LECCIÓ LINGÜÍSTICA CATALANA, 9

PPU, SA
Promociones y Publicaciones Universitarias, SA
Diputació, 213. 08009 Barcelona

Tel.: 934 516 570
Fax: 934 521 005
cristinappu@hotmail.com

DL: B-4.293-2016

Imprimeix: PPU

ÍNDIX

Presentació	9
-------------------	---

I

LES TECNOLOGIES INFORMÀTIQUES APLICADES A LA DIALECTOLOGIA

Gotzon Aurrekoetxea: <i>El atlas lingüístico vasco: 20 años de innovación tecnológica</i>	15
Pilar García Mouton: <i>El Atlas Lingüístico y Etnográfico de Castilla-La Mancha (ALECMAN) y el Atlas Linguistique Roman (ALIR)</i>	43
Esteve Clua: <i>El mètode dialectomètric: aplicació de l'anàlisi multivariant a la classificació de les varietats del català</i>	59
Hans Goebel: <i>Presentació de l'atles sonor en el marc de l'«Atles lingüístic del ladí dolomític i dels dialectes veïns, primera part» (ALD-I)</i>	89
Jaume Corbera: <i>L'«Arxiu audiovisual dels dialectes catalans de les Illes Balears»</i>	117
Maria Pilar Perea: <i>Mapes electrònics i mapes sonors</i>	135

II

L'EXPLOTACIÓ DE CORPUS LINGÜÍSTICS

Jacques Durand, Michel Aurnague i Julien Eychenne: <i>La phonologie du français contemporain au Pays basque et son contexte sociolinguistique</i>	155
M. Àngel Pradilla, Josefina Carrera i Ana M. Fernández: <i>La sociolingüística de la variació i la fonètica experimental. Per a una complementarietat metodològica</i>	199
Joan Julià i Imma Creus: <i>Els (sub)corpus lingüístics del Diccionari de la pronúncia catalana</i>	237

Presentació

La dialectologia, que prengué volada a mitjan segle XIX amb la finalitat d'estudiar la variació en el llenguatge en termes de distribució geogràfica dels parlants, és avui en dia una disciplina científica de plena actualitat. Els elements bàsics en la recollida de dades s'han mantingut des dels orígens: cal un qüestionari més o menys extens, uns informants que confirmin amb llurs respostes un determinat estat de la llengua objecte d'estudi i uns entrevistadors que duguin a terme l'enquesta. Ha canviat, però, el suport on s'enregistren les dades: els quaderns de camp han estat substituïts pel DAT, pel minidisc, pel DVD o pel vídeo digital, passant per un estadi intermedi, que ja ha quedat obsolet, d'ús del magnetòfon o de la casset convencional. Han canviat també els mètodes de processament de les dades: avui els resultats s'emmagatzemen en bases de dades, que incorporen registres lingüístics i sons, i, quan escau, imatges; i el conjunt de registres que formen part de la base de dades o l'aplec de textos orals de caràcter dialectal sol rebre el nom de "corpus".¹ I ha canviat igualment el sistema de difusió dels resultats: a les monografies específiques sobre una varietat o sobre un dialecte i als atles de gran format en paper, s'hi ha afegit la possibilitat de

¹ En sentit estricte, el terme *corpus* se sol aplicar, des de l'àmbit de la lingüística de corpus, a la recopilació estructurada de textos, orals i escrits, que s'han seleccionat d'acord amb uns criteris lingüístics determinats i s'han codificat d'una manera estàndard i homogènia, amb la finalitat que es puguin tractar amb tècniques informàtiques.

difondre les dades, per tal que siguin assequibles a tota la comunitat científica, a través d'Internet o mitjançant un suport digital, com ara el CD-ROM o el DVD.

La primera part del volum que presentem recull les actes de les ponències que es van presentar en l'onzena edició del Col·loqui Lingüístic de la Universitat de Barcelona, que va tenir lloc el 15 de desembre de 2003. El Col·loqui pretenia d'examinar la importància de l'aplicació de les noves tecnologies en la dialectologia. Les intervencions es van centrar no solament en el procediment de recollida de materials dialectals, sinó també en l'edició de les dades, mitjançant els atlas electrònics, i en els nous mètodes d'interpretació dels resultats, com ara la dialectometria.

Pel que fa a la temàtica de les presentacions, en primer lloc, Gotzon Aurrekoetxea va fer constar la contribució de la informàtica —els avantatges, però també certes dificultats— en la materialització del projecte de l'atles basc. A continuació, Pilar García Mouton va descriure les innovacions tecnològiques que s'han aplicat en l'elaboració del *Atlas de Castilla la Mancha* i de l'*Atlas Linguistique Roman*. Esteve Clua tancà les sessions expositives amb una síntesi de l'aplicació de l'estadística en la dialectologia, mitjançant un procediment que ha rebut el nom de *dialectometria*. Les tres darreres contribucions del Col·loqui es van basar en sengles descripcions monogràfiques relacionades amb la cartografia lingüística. Hans Goebel presentà el projecte de confecció d'un atlas sonor en el marc de l'«Atles lingüístic del ladí dolomític i dels dialectes veïns», la versió informàtica del qual s'ha adjuntat, en format DVD, en la contratapa d'aquest volum, gràcies a la gentilesa de l'autor. Jaume Corbera va fer una exposició de les característiques i del funcionament del DVD que conté els enregistraments que configuren el projecte multimèdia que rep el nom d'«Arxiu audiovisual dels dialectes catalans de les Illes Balears». En darrer

lloc, Maria Pilar Perea va descriure l'elaboració de mapes sonors en el marc del projecte VALDIC (“Anàlisi i interpretació de la variació lingüística dialectal a partir de l'explotació d'un corpus oral”).

La segona part del volum recull les contribucions que es van exposar en la novena Jornada sobre la variació lingüística, que s'esdevingué, en el marc de la xarxa temàtica «La variació lingüística. Dialectologia, sociolingüística i pragmàtica», el 17 de setembre de 2004. El fil conductor que va vincular les tres presentacions fou l'explotació de corpus lingüístics de característiques diverses. Cal destacar que, d'uns anys ençà, l'elaboració de corpus —escrits i orals— és una activitat més i més estesa que permet d'extreure informació rellevant i pertinent en relació amb determinats aspectes de la llengua.

En la tasca de bastir els corpus amb dades representatives, és ben cert que el procés de recollida de dades esdevé una tasca llarga i feixuga i el processament informàtic dels materials requereix grans esforços humans i temporals. Tanmateix, ambdues etapes són necessàries per arribar a l'etapa definitiva: la presentació de resultats. La IX Jornada sobre la Variació Lingüística va voler posar a disposició de tots els membres de la xarxa les activitats que diferents grups que la integren desenvolupen actualment en el terreny general de la lingüística de corpus. Els corpus que es van descriure en aquell acte tenien el denominador comú d'estar relacionats, poc o molt, amb la dialectologia. En primer lloc, Jacques Durand, Michel Aurnague i Julien Eychenne, en el marc del projecte «Phonologie du français contemporain: usages, variétés et structure», van relacionar algunes realitzacions fonològiques, obtingudes mitjançant enquesta en el País Basc, amb llur context sociolingüístic; en segon lloc, Miquel Àngel Pradilla, Josefina Carrera i Ana M. Fernández van associar metodològicament la sociolingüística de la variació amb la fonètica experimental des de l'àmbit

de diversos projectes de recerca que tenen a veure especialment amb la percepció del contrast fònic i amb la prosòdia; en tercer lloc, Joan Julià i Imma Creus van presentar els subcorpus lingüístics que intervenen en l'elaboració del *Diccionari de la pronúncia catalana*.

Les informacions que els diversos articles contenen, relacionades tant amb els recursos informàtics com amb l'explotació dels diferents corpus, permeten de conèixer els objectius de cadascun dels projectes, l'estat de la qüestió i la prospectiva.

Maria Pilar Perea
desembre de 2004

I

LES TECNOLOGIES INFORMÀTIQUES APLICADES A LA DIALECTOLOGIA

El atlas lingüístico vasco: 20 años de innovación tecnológica¹

Gotzon Aurrekoetxea

Departamento de Lingüística y Estudios Vascos

EHU-UPV

1. Introducción

Agraeixo la invitació de Maria Pilar Perea per a participar en aquesta jornada monogràfica sobre les noves tecnologies aplicades a la dialectologia, agraïment que s'estén a tot el Departament de Filologia Catalana.²

Mis rudimentarios conocimientos del catalán me obligan a utilizar un idioma extraño (en el sentido de no usual) tanto para mí como para los que han tenido la amabilidad de invitarme a esta jornada.

En consecuencia, utilizaré un idioma que podría llamar “hermanado”, porque aun conociendo la lucha de lenguas siento necesario, justo y

¹ El atlas lingüístico vasco (EHHA) es un proyecto que se desarrolla en el seno de la Academia de la Lengua Vasca-Euskaltzaindia.

² Y agradezco a Lluís Gonzales i Segura, lector de catalán en nuestra facultad, la ayuda que me ha proporcionado para traducir al catalán estas líneas de agradecimiento.

saludable la convivencia entre los idiomas. Y en defensa de esta convivencia tanto vosotros como nosotros (catalanes y vascos) somos adalides, porque por mucho que hablen o callen, y en el caso vasco hablo con conocimiento de causa, no podemos olvidar que hemos sido objeto de persecución por motivos idiomáticos. Porque para convivir hay primero que vivir y en muchos casos nuestros idiomas tienen problemas incluso de supervivencia.

Entrando en materia de este *col·loqui lingüístic* y como punto de partida de mi exposición he de decir que la utilización de la informática es inherente al proyecto del atlas vasco, es inseparable de él desde la misma concepción, al menos en la mente y en los deseos de los responsables académicos del mismo. Y no porque seamos forofos o tengamos especial simpatía por la informática, sino simplemente porque entendemos que la informática nos ofrece los medios más adecuados para tratar y elaborar la información recogida, por una parte; y porque entendemos que la publicación del atlas pasa inexorablemente por la publicación electrónica.

Pero, si de nuestra experiencia hemos de hablar, me temo que no sea quizá la persona más adecuada, porque si algo he aprendido en los 19 años de convivencia con los informáticos y de utilización de la informática diría que ha sido justo lo que no se debe hacer. No os hagáis, pues, ilusiones pensando que en estas líneas se encuentra la solución de todos los problemas que los lingüistas solemos tener cuando utilizamos aplicaciones informáticas. Lo único que hemos aprendido es lo que no hay que hacer; me falta aún por aprender lo más importante, lo que hay que hacer. Pero somos afortunados, porque esta vida nos ha dado una tercera oportunidad para aprender y hacer las cosas bien; y estamos en ello, con la innovación de todo el sistema informático del atlas vasco que se halla en curso desde hace unos meses.

No me acuerdo quién decía aquello de que de los errores se aprende más que de los aciertos. Por razones que no vienen al caso, yo no aceptaría esta aseveración en todos sus términos, pero si fuera verdad los responsables del atlas vasco nos encontraríamos en inmejorable posición para llevar a cabo un buen programa informático para los profesionales de la geolingüística.

2. Innovaciones informáticas del proyecto del atlas vasco (EHHA)

El atlas vasco (EHHA) ha conocido 4 aplicaciones o innovaciones informáticas diferentes a lo largo de los 20 años de trabajos.

2.1. Sistema de gestión de datos del EHHA (1984-1988)

Como he dicho antes, el atlas vasco nació con vocación de utilizar la tecnología en curso en todas las fases del proyecto, desde el mismo nacimiento del proyecto. Una de las primeras decisiones del equipo directivo fue la compra de un ordenador, que se convirtió evidentemente en el centro de toda la actividad (ordenador que costó 1.500.000 pesetas; estoy hablando del año 1984), con una versión del sistema operativo que no tenía absolutamente nada que ver, como podéis imaginar, con las versiones actuales.

La confección del cuestionario, la relación de las localidades, de los informantes, etc. se realizó con la ayuda de una base de datos. Gracias a la ayuda inestimable de Iñaki Ibarra, asesor informático en aquel tiempo, que fue capaz de proporcionarnos un sistema de gestión de datos en la fase de encuesta.

Fue el primer contacto con la informática, en la que no jugaba un papel determinante en los trabajos que el equipo llevaba a cabo; sin lugar a dudas se convirtió en una ayuda inestimable, pero no indispensable, para seguir con los trabajos preparatorios. La verdad es que nos desentendimos del tema por la confianza casi ciega en nuestro asesor. Tenía a su favor nuestra prácticamente nula preparación en informática. Pero a pesar de ello, el sistema de gestión de datos nos fue muy útil.

Mediante este sistema de gestión de datos se elaboraron todos los trabajos preliminares del EHHA. Las características más sobresalientes del atlas vasco son las siguientes:

a) El cuestionario consta de 2.857 preguntas (2.162 sobre el léxico, 336 sobre la morfología nominal, 260 sobre la morfología verbal, 62 sobre la sintaxis y 37 sobre fonética sintáctica). Este cuestionario se completa con una ficha de encuesta compuesta por:

- Número de pregunta
- Título de la pregunta
- Información sobre el apoyo iconográfico de la pregunta
- Perífrasis o el modo de pregunta en distintos idiomas
- Orientaciones sobre el valor etno-lingüístico de la pregunta
- Léxico susceptible de aparecer en las respuestas ordenado por dialectos

b) Localidades: El atlas vasco recoge información de 145 localidades:

- 73 localidades en la Comunidad Autónoma Vasca
- 27 localidades en la Comunidad Foral de Navarra
- 45 localidades en Pays Basque

c) Metodología de las encuestas

-
- Grabación íntegra de las encuestas. Se han recogido más de 4.000 horas de material audio. Material que digitalizado está disponible para los investigadores en la “Azkue Biblioteka” de la Academia de la Lengua Vasca.
 - Idioma de encuesta: utilización del idioma de los informantes para la recogida del léxico y de otro idioma (español o francés) para la recogida de los datos gramaticales.
 - Informantes: la selección de los informantes obedece a un perfil clásico en los atlas tradicionales: edad avanzada, oriundos de la localidad, no alfabetizados en su idioma, etc.
 - Preguntas y propuestas de vocablos: el atlas vasco recoge información proporcionada por los informantes respondiendo a una pregunta e información sobre los vocablos propuestos por los encuestadores a los informantes indagando si son utilizados en la localidad, si son palabras arcaicas, etc.
 - La transcripción de las encuestas se hace por medio de una ficha de respuesta que se halla en un cuaderno de respuestas. La información correspondiente a cada localidad se transcribe en cuatro cuadernos de respuestas. Cada ficha de respuestas recoge información concerniente a una pregunta. La ficha contiene información del informante, número de cinta magnética y localización de la respuesta en la cinta, el contexto de la respuesta, la respuesta en alfabeto fonético, precisiones sobre la elocución, el grado de aceptabilidad, la categoría gramatical, etc.; un lugar destinado para transcribir los textos lingüísticos producidos por los informantes o por el transcriptor y, finalmente, los vocablos propuestos con información sobre el grado de aceptación por parte de los informantes.
 - Utilización del alfabeto fonético IPA para las respuestas y del alfabeto normativo para los textos.

- Recogida de material iconográfico (fotografías, diapositivas, etc.), complemento necesario para la información etnográfica recogida.

La metodología en la recogida de los datos utilizada en el atlas vasco ha estado enfocada a la obtención de dos tipos de información:

- Información lingüística
- Información etno-lingüística

En lo que se refiere a la información lingüística entendemos que toda persona tiene una competencia lingüística activa mediante la que expresa sus pensamientos y realiza los actos comunicativos con sus paisanos. Pero también es poseedora de otra competencia lingüística, que denominamos pasiva, por medio de la cual es capaz de entender los mensajes de otros hablantes, a pesar de que utilicen palabras o formas gramaticales que jamás ha pronunciado en su vida o que, aunque en algún momento los haya utilizado, actualmente no los utiliza y no es capaz de utilizar.

La metodología para recoger estas dos competencias se traduce en un doble sistema de recogida de datos:

- Pregunta-respuesta: la primera de ellas es el binomio pregunta-respuesta utilizado en toda historia de la atlantografía.
- Contra-pregunta o utilización de las propuestas: la segunda es el método de la contra-pregunta, mediante el que el encuestador proporciona o sugiere al informante una serie de vocablos preguntando si los utiliza, si han sido utilizados por sus antepasados, si son utilizados en su localidad o en localidades vecinas, etc. Se trata de recoger palabras susceptibles de ser usadas por los lugareños pero que no las ha pronunciado de *motu proprio* el informante; metodología utilizada en la segunda fase del ALG por X. Ravier, pero haciendo un uso más intensivo y regularizado.

El atlas lingüístico de las variedades del euskara ha recogido abundante información etno-lingüística. Éramos conscientes del fin de la vida tradicional en nuestros pueblos. La irrupción de la industrialización y de los medios de comunicación ha hecho desaparecer prácticamente la vida y las costumbres tradicionales: han desaparecido las técnicas artesanales (elaboración de los cestos, queso, etc.), han desaparecido los medios tradicionales de transporte (el asno se ha convertido en objeto turístico, el caballo no encuentra trabajo, los bueyes actúan sobre empedrado, el carro es objeto de museo, etc.); la lavadora ha hecho desaparecer las técnicas tradicionales de la limpieza de la ropa, sábanas, mantas, etc.; la televisión ha cambiado radicalmente el modo de relacionarse entre los miembros de una misma familia. Los niños ya no tienen quien les cuente cuentos (de ello se encargan la televisión, los CD-s, los ordenadores, los libros, etc.); han desaparecido las tertulias, ya no se cantan canciones populares como antaño, etc. No creáis que estáis ante un nostálgico. Es la pura, cruda y, si queréis, auténtica realidad.

Conscientes de que esta pérdida era irremediable se cuidó de recoger el saber popular que aún guardaban nuestros informantes. De esta forma se ha recogido abundante material etno-lingüístico que está parcialmente transcrito pero que en su mayor parte se halla en las cintas y en los CD-ROM que he mencionado más arriba.

La información que se ha recogido se puede clasificar en:

- Sabiduría popular
- Cuentos
- Narraciones
- Refranes
- Canciones
- Remedios y recetas

- Definiciones y precisiones semánticas
- Etc.

Una pequeña muestra de esta información transcrita se publicará en formato libro. Se publicará más información en formato CD-ROM.

2.2. Programa de informatización de los datos (1988-1996)

La confección del programa para el procesamiento de los datos de las encuestas ya fue otro cantar. Se nos antojaba determinante en el devenir del proyecto.

El programa fue realizado por la empresa informática del Gobierno Vasco llamada EJIE. Quisiera que mi intervención en este *col·loqui* sirviera para explicar los pasos que hemos dado en la informatización de los datos, pero que también sirviera para reflexionar sobre la relación entre los lingüistas, dialectólogos en este caso, y los informáticos encargados de la confección de los programas informáticos.

En el caso que nos ocupa, se puede afirmar que fue una larga conversación entre sordos, que al final llegan a entenderse, aunque sea parcialmente. Partíamos de dos mundos casi-casi antagónicos, en el que el dicho de la “en informática se puede hacer todo” se convirtió para nosotros en otro que podríamos resumir en que “la informática puede fastidiar todo”. Era evidente que debíamos trabajar mucho y duro para hacernos entender, para hacerles ver nuestras necesidades, y comprender lo que nos trataban de hacer con sus propuestas informáticas.

Nuestro conocimiento del uso de la informática en la atlantografía en aquella época se limitaba al proyecto del atlas de Alsacia en Estrasburgo.³ El retorno de la visita que les rendimos se convirtió en una pesadilla; en nuestras mentes sólo ronroneaba una pregunta: ¿Cómo introducir en el ordenador los datos fonéticos?

Para la primera entrevista con el informático que EJIE había destacado para nuestro proyecto habíamos adoptado dos decisiones claras: *a)* debíamos introducir datos fonéticos; *b)* la introducción de cada signo fonético se debería hacer con una única pulsación.

La aplicación informática para el procesamiento de los datos se decidió que funcionara sobre “dBase 3+”. Si bien tanto el diseño como la creación del programa de introducción de los datos como la estructura de la base fueron muy laboriosas, hubo aspectos en los que dejamos parte de nuestra vida:⁴

a) Uno de estos aspectos fue la creación de los signos fonéticos. El problema surgió debido a que no existía en aquel tiempo un tipo de letra adecuado para los datos fonéticos y tuvimos que diseñar los signos fonéticos mediante un editor de letras. Tarea evidentemente que escapa a la de un dialectólogo, pero que a la vista de que los informáticos no la asumían recayó sobre mis hombros. Fue un auténtico martirio diseñar píxel a píxel de entre los 256 caracteres ASCII, que tenía el editor de caracteres, los que necesitaba nuestro alfabeto fonético.

³ Me refiero al atlas lingüístico de M. Philip, *Atlas Linguistique et Ethnographique de la Lorraine Germanophone*.

⁴ Una exposición más detallada de la estructura de la base de datos se puede ver en Aurrekoetxea (1992) o en Aurrekoetxea y Videgain (1994).

b) El problema de la introducción de los datos fonéticos al ordenador se resolvió con una tableta digitalizadora (de unos 60 cm x 35 cm.), en la que se colocaron todos los signos fonéticos que necesitaba nuestro proyecto. El diseño de las coordenadas y la disposición de los signos fonéticos en dicha tableta recayó también en los dialectólogos.

c) El tercer y gran escollo fue la creación de una tabla informática en la que queríamos registrar la reacción del informante a las propuestas hechas por el investigador en las encuestas; lo que en nuestra jerga denominamos aún “tabla de conocimiento”. Por medio de esta tabla el investigador puede clasificar la reacción del informante ante la palabra propuesta: si conoce el vocablo propuesto, si considera que tiene significado diferente, si se trata de una palabra arcaica, etc.

La delimitación de las funciones de cada participante en un proyecto de esta envergadura es fundamental, para el correcto desarrollo del mismo. En este caso los dialectólogos se vieron involucrados en trabajos de los informáticos, haciendo trabajos que evidentemente no son de su área de conocimiento. Es la primera gran lección que aprendimos: el lingüista debe abstenerse de hacer el trabajo del informático.

Esta aplicación informática fue robusta y tuvo larga vida, puesto que estuvo vigente desde 1989 hasta 1996. Se trataba de una aplicación mono-puesto, evidentemente, que nos permitía la informatización por cuadernos y por localidades: primeramente se informatizaron todas las respuestas correspondientes al primer cuaderno de respuestas de cada localidad. Una vez informatizados los datos correspondientes al primer cuaderno, se abordaba el segundo cuaderno y así sucesivamente.

El sistema de copia de seguridad requería sistemas de almacenaje con mayor capacidad que los disquetes; en nuestro caso se recurrió al *Streamer*.

Esta aplicación informática únicamente contemplaba la informatización de los datos. No se ocupaba en absoluto de la publicación y, por tanto, no se había previsto el sistema y el modo de aglutinar todas las pequeñas bases de datos en una única que albergara todas ellas.

Pero fue un sistema que utilizando la tecnología vigente en la época fue capaz de proporcionarnos una aplicación que nos permitía la informatización de los datos que habíamos recogido en las encuestas y teníamos transcritos en nuestros cuadernos de respuestas, localidad a localidad.

2.3. Nuevo programa de informatización de los datos y programa para la elaboración y cartografiado (1996-2004)

En el año 1996 se decidió acometer los trabajos de la edición de los datos. Si bien los trabajos de la informatización no habían concluido, los responsables del proyecto comenzaron a diseñar la publicación del atlas.

Como consecuencia de ello se decidió renovar el sistema informático del proyecto. La nueva aplicación, implantada sobre el sistema Windows, está compuesta en realidad por dos aplicaciones: *a)* aplicación para el procesamiento de los datos; y *b)* aplicación para la publicación. Esta última tiene módulos diferentes: *a)* módulo de elaboración o puesta a punto de los datos; *b)* módulo de la lematización; *c)* secuencialización de las respuestas; *d)* módulo cartográfico; y *e)* módulo de autoedición.

Esta innovación tecnológica vino con la implantación y universalización del sistema Windows95, con herramientas Microsoft Office (Access, Word y Publisher) y programas creados para el proyecto.

Las dos aplicaciones informáticas siguen siendo mono-puesto, por dos razones: por una parte porque Euskaltzaindia no contaba con un sistema informático integrado; por otra, porque el equipo del atlas no tenía un único local de trabajo (trabajaba y sigue trabajando en sus distintas sedes y delegaciones de Bilbao, Donostia, Baiona e Iruñea).

Si la creación del anterior programa fue un quebradero de cabeza éste se convirtió en un calvario a la hora de testar las distintas versiones informáticas que nos llegaban. La aplicación corrió a cargo de la *Fundación Aurten Bai*. Porque a la dificultad de entendimiento entre dialectólogos e informáticos, se añadía en este caso el volcado de las pequeñas bases creadas en la fase de la introducción de datos, organizadas por cuadernos de respuesta y por localidades, a una única base de datos.

A decir verdad, pasamos de un extremo a otro: de una situación en la que se debía de dar ideas al informático y hacer incluso tareas de su incumbencia, pasamos a otra en el que el informático se adelanta continuamente y da soluciones a problemas que ni siquiera se habían imaginado. En efecto, el informático nos presentaba sus propuestas informáticas con programas hechos, terminados, adelantándose sin cesar a nuestras necesidades, dando soluciones a problemas jamás planteados: en definitiva, el informático metido a dialectólogo.

Seguro que por vuestras mentes ha pasado la idea de “¡Qué bien!”, “¡Qué informático más majo!”. Pues ni sí y ni no, ni todo lo contrario, por contradecir el dicho popular. Pero, de hecho, de todo un poco.

Sí, porque evidentemente facilitó mucho nuestra labor. Una persona inmersa en el mundo cultural vasco, con conocimientos suficientes de las tareas de la dialectología vasca, con acceso relativamente fácil a los diversos estamentos de la Academia de la Lengua Vasca-Euskaltzaindia, que presentó un proyecto atractivo, factible y alcanzable para la Academia, un

“chollo” de programa, “una golosina”, en definitiva, que no se dudó en coger.

No, porque la aplicación o las aplicaciones presentaban lagunas y puntos débiles; porque no tuvimos el tiempo y la tranquilidad deseada para testar la aplicación, y se han tenido que ir parcheando según surgían los problemas; porque el mantenimiento se ha convertido en el talón de Aquiles, con una dependencia superior a lo que se puede dar por norma en una relación comercial habitual. Con la perspectiva que nos da el tiempo, se puede decir que no es la mejor manera de trabajar. Esta es la segunda gran lección que hemos aprendido.

Después de este preámbulo, paso a explicar las virtudes del sistema informático vigente actualmente. Se compone de dos grandes núcleos:

- El núcleo del procesamiento de los datos.
- El núcleo de la elaboración y edición del atlas.

El procesamiento de los datos se hace en Donostia-San Sebastián. La elaboración y edición se realiza en Bilbao, Iruñea-Pamplona y Baiona-Bayonne.

2.3.1. El programa del procesamiento de los datos

La aplicación del procesamiento de los datos se compone de un único módulo que permite el procesamiento de los datos localidad a localidad.

Este programa permite la introducción tanto de los datos ortográficos como de los fonéticos. El procesamiento de los datos fonéticos se realiza mediante una pantalla que se activa al poner el cursor en un campo específico y en la que están registrados todos los signos fonéticos que se

utilizan. Una vez situado en esta pantalla es suficiente pinchar con el ratón sobre una cadena de signos y aceptar. Un sistema cómodo y rápido para trabajar en una base de datos y registrar datos fonéticos. Como se puede comprobar se utiliza el alfabeto IPA.

i	e	ø	ə	ɛ	œ	a	ɔ	o	u	y	p	p	t	t	ʈ	ʈ	tt	tt	k	k
ĩ	ẽ	ø̃	ə̃	ẽ	œ̃	ã	õ	õ	ũ	ỹ	ph	ph	th	th					kh	kh
ĩ	ĩ	u	u	y	y				ẽ	o	b	b	d	d	ɖ	ɖ	dd	dd	g	g
j	j	w	w	ɥ	ɥ				ə	ə	ɸ	f	θ	s	ʃ				x	h
ĩ	e	ø	ə	ɛ	œ	a	ɔ	o	u	y	ɸ	f	θ	s	ʃ				x	h
ĩ:	e:	ø:	ə:	ɛ:	œ:	ã:	õ:	õ:	ũ:	ỹ:	β	v	ð	z	ʒ		ʒ	ʝ	Y	
ĩ̃	ẽ̃	ø̃	ə̃	ẽ̃	œ̃	ã̃	õ̃	õ̃	ũ̃	ỹ̃	β	v	ð	z	ʒ		ʒ	ʝ	Y	ɸ
ĩ̃:	ẽ̃:	ø̃:	ə̃:	ẽ̃:	œ̃:	ã̃:	õ̃:	õ̃:	ũ̃:	ỹ̃:			ts	ts	tʃ	tʃ				ɹ
í	é	ó	ó	é	œ	á	ó	ó	ú	ý			dz	dz	dʒ	dʒ	dʒ	dʒ		ɹ
í:	é:	ó:	ó:	é:	œ:	á:	ó:	ó:	ú:	ý:	m	m			n	ɳ		ɳ	ɳ	ʃ
ĩ̃	ẽ̃	ø̃	ə̃	ẽ̃	œ̃	ã̃	õ̃	õ̃	ũ̃	ỹ̃	m	m			n	ɳ		ɳ	ɳ	ʃ
ĩ̃:	ẽ̃:	ø̃:	ə̃:	ẽ̃:	œ̃:	ã̃:	õ̃:	õ̃:	ũ̃:	ỹ̃:				l	l	ɭ	ɭ	ɭ	ɭ	tʃ
e	e	é		/	(	)	-	*	.	;	v:		ɹ	ɹ	ɹ	ɹ	ɹ	ɹ	R	R
a	a	á	o	o	ó		!	?	:	.	ʃ	s	f	x	f	l	ɹ	n	n	ɹ

Gráfico 1. Pantalla para el procesamiento de los datos fonéticos.

La introducción de las respuestas se hace en alfabeto fonético, pero el sistema informático tiene instalado una tabla de transliteración mediante la cual vuelca los datos al alfabeto ortográfico.

El procesamiento de los datos se ha acometido dividiendo el cuestionario en cinco partes: primeramente se han informatizado los datos recogidos en todas las localidades correspondientes sólo a la primera parte del

cuestionario. Posteriormente se ha obrado de la misma manera con los datos correspondientes a la segunda parte. Es un trabajo que comenzó en 1991 y del que quedan aún unos años, pues actualmente falta por informatizar el 30 % de los datos.

Esta informatización escalonada de los datos permite la edición de aquellos que están informatizados.

2.3.2. Puesta a punto de los datos

La elaboración y edición de los datos se compone de 4 módulos, a saber: la puesta a punto de los datos, la lematización, la secuencialización y la cartografía propiamente dicha.

La elaboración y edición comienza con la puesta a punto de los datos. Se trata de una lectura horizontal, pregunta a pregunta. El editor ve los datos desde una perspectiva diferente a la del procesador. Éste introduce los datos localidad a localidad, pero el editor analiza los datos pregunta a pregunta. Decimos que la visión del procesador se halla en el plano vertical, mientras que la del editor se halla en el horizontal.

Por ello se hace necesaria una confirmación de los datos localidad a localidad y respuesta a respuesta, resolviendo las dudas de los encuestadores y/o de los informantes, armonizando las distintas sensibilidades de los encuestadores, seleccionando los datos más interesantes de la base de datos para la creación de las notas de los mapas, etc.

2.3.3. Lematización de las respuestas

Después de la confirmación de los datos, el editor encara la cartografía lingüística. Esta labor requiere tres pasos: *a)* la lematización de las respuestas; *b)* la secuencialización; y *c)* la creación del mapa.

La lematización obedece a criterios estrictamente lingüísticos. Criterios que se aplican con cierta flexibilidad en cada una de las preguntas tratadas. Se trata de una clasificación tipológica de los datos.

El punto de partida de esta tarea es un listado de respuestas en alfabeto ortográfico. El editor clasifica las respuestas en grupos más o menos afines, atendiendo a diversos criterios. Pero antes de entrar a enumerarlos hay que hacer una aclaración: el atlas vasco no utiliza este término en la acepción con que se utiliza en lexicología, por poner un ejemplo, en la confección de las entradas de los diccionarios. En nuestro caso no se trata de un problema lexicológico, sino más bien lingüístico: de delimitar los rasgos más sobresalientes que se encuentran entre las respuestas. Los criterios son, pues, estrictamente lingüísticos. Los más comúnmente utilizados son:

- etimología
- ausencia o presencia de un morfema
- ausencia o presencia de un sonido determinado
- ausencia o presencia de una regla fonológica

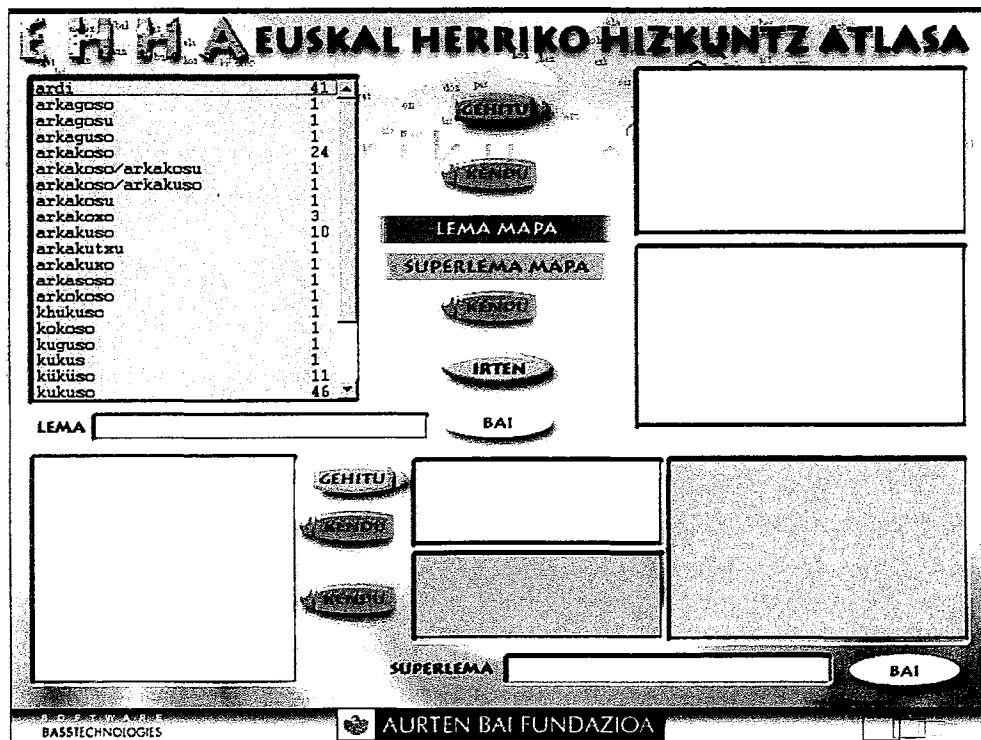


Gráfico 2. Pantalla de la lematización.

Partiendo de la lista de respuesta (izquierda / arriba de la pantalla) que le proporciona el programa, el investigador tiene que ir aglutinando y clasificando las respuestas de acuerdo a los criterios descritos hasta vaciar el espacio, y pasando por una serie de filtros llegar a llenar el espacio de los “superlemas”, base de la leyenda de los mapas.

Esta tipologización de los datos puede obedecer a criterios más o menos subjetivos, pero siempre se tratará de una clasificación. Es por ello por lo que nosotros utilizamos el término de “elaboración” y no de “interpretación”.

Consideramos el término “interpretación” como un procedimiento mediante el cual se pregunta el porqué de una palabra, morfema o rasgo lingüístico determinado recogido en una localidad, atendiendo a la historia, cultura, geografía, vertebración política, etc. de una región. Nosotros no preguntamos por la razón de la aparición de una palabra en una determinada localidad; simplemente constatamos tal hecho lingüístico.

Esta lematización es un intento para hacer más visibles y leíbles los datos. Es un intento de superar el dicho popular “que los árboles no nos dejen ver el bosque”. Queremos ver el bosque en su conjunto, las distintas variedades de árboles que encierra, y no árbol por árbol.

Esta tarea tiene también sus límites. El número máximo de superlemas permitido se eleva a 15.

El programa prevé diversos tipos de lematización:

- Lematización ortográfica para el mapa grande
- Lematización ortográfica para el mapa pequeño
- Lematización fonética para el mapa grande
- Lematización fonética para el mapa pequeño

El diseño de la publicación de los datos prevé un mapa por pregunta basándose en la lematización de las respuestas en alfabeto ortográfico. Es por ello que el investigador está obligado a la lematización ortográfica de las respuestas.

Hasta el momento (se han elaborado unos mil mapas) no se ha creado ningún mapa grande basándose en las respuestas fonéticas, a pesar de que es posible hacerlo.

Si el investigador cree conveniente que el mapa usual no representa de manera correcta o adecuada alguna de las peculiaridades que se encuentra entre las respuestas tiene la posibilidad de hacer una nueva lematización,

bien sea partiendo de los datos fonéticos o de los ortográficos, para un mapa de reducidas dimensiones que denominamos “mapa pequeño” y que se halla en el vértice derecho bajo de la página del listado de las respuestas.

Para llevar a cabo la lematización se ha creado un programa *ad hoc* para el proyecto.

2.3.4. Secuencialización de las respuestas

Una vez terminada la lematización se acude de nuevo a la base de datos para ordenar las respuestas en orden de preferencia para el mapa.

El mapa da la posibilidad de publicar los datos de dos respuestas y una propuesta. Si en una localidad se ha recogido más de una respuesta a una pregunta determinada, tenemos la opción de invertir el orden mediante la secuencialización.

Así mismo, si en una localidad se ha aceptado más de una propuesta tenemos la opción de elegir cuál de ellas debe ser publicada en el mapa.

Todas aquellas respuestas o propuestas aceptadas que no tengan cabida en el mapa se publican en otro apartado en forma de listado.

2.3.5. Programa de cartografía lingüística

Es un programa gráfico diseñado *ad hoc*, en el que la base de los mapas se sustenta en la poligonación de las localidades, siguiendo el modelo creado por el profesor Hans Goebel, en sus trabajos de dialectometría.⁵

Las 145 localidades se convierten en otros tantos polígonos en los que se colocan las características de las respuestas obtenidas.

⁵ Ver por ejemplo “La dialectométrie appliquée à l’ALF (Normandie)”.

Antes de dar respuesta a estas preguntas ya habíamos dado un no rotundo a la publicación típica de grandes dimensiones. Queríamos un libro “normal”, de dimensiones normales, vendible, almacenable, utilizable, etc.

Esta decisión nos ayudó mucho en definir las características de la publicación: en cuanto a las dimensiones de la publicación en papel se decidió utilizar el formato A4, lo que nos indicaba la dimensión máxima del mapa.

A pesar de las reducidas dimensiones tenemos la opción de crear dos mapas por cada pregunta: un mapa grande y un mapa pequeño. El primero de ellos se convierte en obligatorio para el editor; el segundo es opcional.

2.3.5.1. Mapa grande

Como se puede observar en el gráfico los mapas del EHHA son mapas de áreas, coloreadas según el color asignado a cada superlema.

Las reducidas dimensiones del mapa no impiden la introducción de tres respuestas en una misma localidad por cada pregunta: dos respuestas y una propuesta.

Los superlemas creados en la lematización son la base de la leyenda del mapa. A cada superlema se corresponde un color que puede ser elegido y cambiado por el dialectólogo.

El rango de las respuestas tiene una visualización diferente:

- A la primera respuesta se le asigna el color de fondo.
- La segunda respuesta se visualiza en forma de trama.
- La propuesta está representada por un pequeño círculo del color correspondiente.

En aras a una mejor lectura del mapa se incrustan unas etiquetas con textos de las respuestas.

2.3.5.2. Mapa pequeño

Los mapas pequeños son opcionales para el investigador. Si el mapa grande no refleja claramente algunos aspectos recogidos en las respuestas el investigador puede hacer una nueva lematización atendiendo a un único y unos pocos criterios: en la pregunta 01300 ‘pulga / puce / flea’ los 16 superlemas del mapa grande han sido reducidos a tres superlemas en el mapa pequeño: *ardi*, *arkakuso*, *kukuso*.

2.3.6. Edición

En la versión actual es el dialectólogo el encargado de la autoedición. La autoedición se realiza en Microsoft Publisher.

El dialectólogo utiliza una plantilla que irá rellenando con información apropiada en cada caso. Esta información se extrae de la base de datos.

La autoedición se encarga solamente del formato libro y abarca los siguientes aspectos:

- listado de las respuestas ordenadas alfabéticamente por territorios históricos (ver gráfico núm. 4).
- Cabecera de la pregunta con el número de mapa, el título en español, francés, inglés y la denominación científica cuando es menester; referencia del número de la pregunta del cuestionario, referencias de distintos atlas lingüísticos (ALEANR, ALG, ALE).
- El mapa correspondiente (y en su caso un mapa pequeño adicional).

- La leyenda del mapa.
- Notas lingüísticas al mapa.

[illegible]

Gráfico 4. Listado de respuestas

2.4. Informatización y elaboración en red (2004-)

Esta versión informática tiene ya los días contados. Actualmente se está trabajando en el diseño de una nueva actualización del sistema informático.

En el momento actual no se puede adelantar gran cosa, puesto que estamos iniciando los trabajos y hace falta tomar muchas decisiones. Entre las características más sobresalientes que se pueden adelantar figuran las siguientes:

- No será un sistema mono-puesto. Será un programa cliente-servidor.
- Se trabajará en red. El programa cliente se encargará de descargar los datos a través de Internet, sin intervención de los usuarios, desde las distintas delegaciones. Así los miembros del grupo trabajarán en una única base de datos. En cualquier momento se podrá consultar el trabajo de los compañeros sin tener que esperar a la reunificación de los datos.
- Será un sistema integral, en que se englobarán la introducción y la edición de los datos. Todo el proceso, desde la introducción de datos hasta la elaboración del mapa, se realizará en una misma aplicación. Por ello, ya no serán necesarios conocimientos de autoedición para elaborar los mapas.
- Será un programa que generará mapas independientemente del ámbito geográfico que se quiera usar; de tal forma que posibilitará la edición de listados de respuestas o mapas parciales, de un territorio histórico, de una región, etc.
- Con esta nueva innovación se resolverán los problemas técnicos que se han localizado en la anterior versión.
- Se constituirá un corpus unificado y clasificado, con toda la información del proyecto y no varias bases de datos dispersas. Así, se facilitará su consulta y su explotación a través de Internet.
- Esta aplicación será flexible y estará preparada para trabajar con cualquier corpus y entorno informático. Será multilingüístico, multiplataforma, multiusuario.
- Será un programa modular; es decir, que se le podrán añadir o quitar funcionalidades diferentes.
- Será una aplicación distribuida: los miembros pueden estar distribuidos geográficamente, pero trabajarán en una única base de datos. Para el usuario es como si trabajara en local.

3. La publicación del EHHA

El procedimiento de publicaciones de Euskaltzaindia delega la decisión de la publicación en órganos superiores a la comisión del atlas. La función de la comisión es preparar el original para la impresión y su trabajo termina al entregar el original a la comisión de publicaciones. Actualmente se están corrigiendo las pruebas de imprenta del primer tomo.

La Academia tiene previsto la publicación del atlas en diversos soportes: libro, CD y Internet.

La publicación tendrá las siguientes características:

- Respuestas en listados organizados por localidades con respuestas en alfabeto fonético.
- Mapas con el título, leyenda y notas.
- Índice de las respuestas en alfabeto ortográfico con indicación del número de mapa.
- Índice de los lemas que aparecen en las leyendas de los mapas, con indicación del número de mapa.
- Índice de las preguntas (en español, francés e inglés; la denominación científica en su caso).

Bibliografia

AURREKOETXEA, Gotzon (1986): «Euskal Herriko Hizkuntz Atlasa (EHHA): inkesta metodologia eta ezezko datuak» [el atlas lingüístico vasco (EHHA): metodología de encuesta y los datos negativos], *Euskera* XXXI-2, 413-424.

_____ (1992): «Euskal Herriko Hizkuntz Atlasaren informatizapena», en Aurrekoetxea, Gotzon & Xarles Videgain (ed.), *Actas del Congreso Internacional de Dialectología (Bilbao, 21/25-X-1991)*, Iker 7, Euskaltzaindia, Bilbao, 319-337.

_____ (2002a): «El atlas lingüístico vasco (EHHA)», en Michel Aurnague & Michel Roché, *Hommage à Jacques Allières, I. Domaines basque et pyrénéen*, Atlantica, Biarritz, 63-71.

_____ (2002b): «Algunas consideraciones sobre la contrapregunta en las encuestas lingüísticas», en Lidia Rabassa (ed.), *Mélanges offerts à Jean-Louis Fossat*, Cahiers d'Etudes Romanes, CERCLiD 11/12, 57-65.

AURREKOETXEA, Gotzon & Xarles VIDEGAIN (1989): «L'atlas linguistique du Pays Basque: Euskal Herriko Hizkuntza Atlasa», *Hommage au Musée Basque*, Baiona, 495-510.

_____ (1992) (ed.): *Actas del Congreso Internacional de Dialectología (Bilbao, 21/25-X-1991)*, Iker 7, Euskaltzaindia, Bilbao.

_____ (1993): «Euskal Herriko Hizkuntz Atlasa (EHHA): galdesorta / cuestionario / questionnaire», *Euskera* XXXVIII-3, 529-647.

_____ (1994): «Historia y futuro del Atlas Lingüístico Vasco (EHHA)», en Pilar García Mouton (ed.), *Geolingüística. Trabajos europeos*, Biblioteca de Filología Hispánica, CSIC, 79-97.

_____ (en prensa) «L'interprétation dans les atlas linguistiques: le cas de l'Atlas linguistique basque (EHHA)», *Proceedings of 4. International Congress of Dialectologists and Geolinguistics, Riga, 2003*.

GOEBL, Hans (1976): «La dialectométrie appliquée à l'ALF (Normandie)», en A. Varvaro (ed.), *XIV Congresso internazionale di linguistica e filologia romanza. Atti. II*, Napoles (G. Macchiaroli) - Amsterdam (J. Benjamins), 165-195.

-
- _____ (2003): «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics*, XXV, 59-121.
- RAVIER, Xavier (1965): «Le traitement des données négatives dans l'Atlas linguistique et ethnographique de la Gascogne», *Revue de Linguistique Romane*, XXIX, 262-274.

**El *Atlas Lingüístico y Etnográfico de Castilla-La Mancha*
(ALECMAN) y el *Atlas Linguistique Roman* (ALIR):
un acercamiento a las nuevas tecnologías***

Pilar García Mouton

CSIC

1. Introducción

Mi presencia en este Coloquio Lingüístico de la Universidad de Barcelona, que se celebra bajo el sugerente título de *Nuevas tecnologías aplicadas a la dialectología*, en realidad no se debe a mis méritos en el uso de esas *nuevas tecnologías*, sino a la amabilidad de Maria Pilar Perea, quien, desde el principio y a pesar de mi relativa resistencia, se empeñó en incluirme en él. Por eso no quiero dejar de darles las más efusivas gracias a ella y también al equipo que lo organiza. Mi siguiente obligación antes de comenzar consistirá en excusarme desde el punto de vista científico, porque mi trabajo geolingüístico y dialectológico ha tenido, tiene y forzosamente tendrá que ver con las nuevas tecnologías por voluntad y por cercanía, pero

* Este trabajo se ha redactado en el marco del proyecto BFF2002-01087 del MCYT, desde el que coordino la colaboración al *Atlas Linguistique Roman* y al *Atlas Linguarum Europae*.

nunca por una preparación especial ni por unos conocimientos directos en este campo. De manera que en este coloquio vendré a representar el papel que me corresponde, el de la dialectóloga que se formó en la metodología más tradicional y que poco a poco ha pasado a conocer y a valorar las ventajas que las nuevas tecnologías pueden proporcionar a nuestro trabajo. Hablaré, pues, de la experiencia de un cambio.

Hace ya muchos años, mi colaboración hasta entonces exclusivamente epistolar con el equipo que dirigía los primeros pasos del *Atlas Linguistique Roman*, el ALiR —el hijo románico del *Atlas Linguarum Europae*—, me llevó a una reunión celebrada en París, a la que me hizo invitar su actual director, Michel Contini, sobre la informatización de los atlas lingüísticos franceses. En ella, bajo la dirección de Jean Le Dù, los dialectólogos franceses reunidos en el proyecto GRECO —el que acogía los atlas lingüísticos franceses continuadores de la iniciativa del *Nouvel atlas linguistique de la France par régions* de Albert Dauzat—, se replanteaban el futuro de la publicación en papel de esos atlas, que el CNRS consideraba demasiado caros de editar. Para entonces yo llevaba ya bastantes años trabajando en el CSIC, en el Departamento de Geografía Lingüística que dirigía Manuel Alvar, desde donde atendíamos, y seguimos atendiendo, la colaboración con el ALE y el ALiR, elaborando las síntesis que se enviaban a las Redacciones de los dos atlas, tarea para la que se promovieron encuestas en zonas peninsulares que carecían de estudios geolingüísticos.

En aquel Departamento se habían hecho gran parte de los trabajos de elaboración de varios atlas regionales por el procedimiento tradicional, el de copiar en cuadernillos de formas, uno por cada concepto cartografiable, las respuestas en transcripción fonética de los cuestionarios previamente llenos por los encuestadores en su trabajo de campo. Cada cuadernillo de formas suponía así el primer borrador de cada uno de los futuros mapas, y a partir

de ahí intervenía ya el dibujante, que *dibujaba* la transcripción fonética o los símbolos en los mapas elaborados. Pero en aquel Departamento conocíamos la colaboración temprana que M. Alvar había buscado con el Centro de Cálculo de la Universidad Complutense,¹ para elaborar el que entonces llamábamos *Atlas Lingüístico y Etnográfico de Santander* (ALES), y que sólo mucho después, precisamente por peripecias relacionadas con la informática aplicada, se editó como *Atlas Lingüístico y Etnográfico de Cantabria* (ALECant). También sabíamos que los materiales que enviábamos al *Atlas Linguarum Europae*, el ALE, se estaban cartografiando en Nimega de forma “automatizada”, ya no por los penosos procedimientos manuales. Por otra parte, Manuel Alvar y Antonio Quilis —que vino a aquella reunión de París— dirigían entonces desde el CSIC el proyecto del *Atlas Lingüístico de Hispanoamérica*, en el que yo colaboraba, de ahí también nuestro interés en conocer cómo estaban resolviendo los colegas franceses los problemas de informatización, porque sus experiencias podían ayudarnos a desbrozar el camino.

En nuestro concepto de Geografía Lingüística el recurso a las nuevas tecnologías se planteaba entonces como un instrumento mucho más ágil para la edición, que entonces no se imaginaba en otro soporte que no fuese el papel. El cartografiado automático se buscaba básicamente para agilizar el largo proceso de dibujado y posterior corrección que había que recorrer antes de ver impreso un atlas. Esta situación puede resultar difícil de comprender desde la óptica de los investigadores y los alumnos más jóvenes, que siempre han convivido con los ordenadores y se relacionan con ellos con naturalidad, pero estamos hablando de los últimos años de la década de los setenta y de los primeros años ochenta, donde hubo que

¹ En el libro *Informática y Lingüística* (1984), recogió seis artículos sobre el tema, el primero de ellos ya de 1976.

asimilar su existencia, e incorporarlos a la “cultura” del entorno, antes de pensar en cómo aprovechar sus posibilidades. Por otra parte, hay que decir que, como es lógico, la relación con el mundo de la informática tenía entonces muy poco que ver con la actual en todos los ámbitos, era mucho menor, especialmente la de los investigadores de Humanidades. Y precisamente esa distancia contribuía a que, por nuestra parte, se idealizasen las posibilidades de una colaboración que, a la hora de la verdad, se hacía muy dura al intentar, desde el desconocimiento mutuo, las primeras puestas en común de los objetivos de nuestros proyectos.

En aquella reunión del CNRS se habló de cartografiado automático, pero también de las posibilidades que ofrecían a la larga otras aplicaciones de los instrumentos informáticos; se discutió sobre cómo conseguir unos índices de *todos* los atlas franceses, no sólo de los que todavía estaban sin publicar, sino sobre todo de los muchos que ya estaban editados en papel, y recuerdo que se consideraron los problemas que en ellos planteaba el fondo de mapa, con las divisiones de los departamentos y las líneas de los ríos, que dificultaban enormemente el reconocimiento óptico de las formas transcritas, ya de por sí difíciles de reconocer. Recuerdo que, a la vuelta de aquel viaje, transmitimos con tal entusiasmo aquellas informaciones al profesor Alvar, bastante decepcionado entonces de sus contactos con la informática, que el proyecto siguiente que se solicitó desde el CSIC fue para emprender el “cartografiado automático” del *Atlas Lingüístico de Hispanoamérica*.²

² Esos trabajos se intentaron avanzar desde el CSIC en colaboración con la Universidad de Albany, donde Claire Ziamandanis se encargó de ellos con M. Alvar; cf. Ziamandanis (1996).

2. ALeCMan

A finales de los años ochenta presentamos el proyecto del *Atlas Lingüístico (y etnográfico) de Castilla-La Mancha*, el ALeCMan, García Mouton-Moreno Fernández (1988), que desde hace unos meses puede ser consultado parcialmente en Internet, en la dirección www.uah.es/otrosweb/alecman. Como explicamos entonces, se trataba de hacer un atlas regional, en la línea de los atlas regionales de Alvar, para llenar un hueco evidente, el que existía entre el *Atlas Lingüístico y Etnográfico de Aragón*, el ALEANR, y el *Atlas Lingüístico y Etnográfico de Andalucía*, el ALEA. No hace falta insistir en la necesidad de contar con información detallada de la zona centro peninsular, clave para conocer los límites o la continuidad de tantos fenómenos “meridionales”. Sus innovaciones metodológicas se centraban en dos puntos: en cada localidad se encuestaría a una mujer y a un hombre, no solamente a un hombre, como era lo habitual hasta entonces, y en las ciudades se haría una encuesta de corte sociolingüístico, con metodología diferente y se mantendría un pequeño cuestionario léxico para permitir la comparación con el resto de localidades; cf. García Mouton-Moreno Fernández (1993). Desde el principio pensamos en cartografiar el ALeCMan informáticamente, siguiendo los avances en el cartografiado de los atlas más cercanos, el *Atlas Lingüístico de Hispanoamérica* y algún atlas europeo más pequeño, como el de Córcega, cuyos responsables parecían decididos a ofrecernos su experiencia informática. Lamentablemente estas vías no cuajaron y, después de otros intentos, a partir de 1996, Francisco Moreno Fernández, José Enrique Moreno Fernández y Antonio García de las Heras diseñaron un programa propio llamado *Atlante*, que presentaron en el Congreso Internacional de Dialectología de Amsterdam.

Atlante es el programa que genera, a partir de su base de datos, los mapas del ALeCMan que ahora se pueden consultar, a medida que se van elaborando y colgando en Internet. De acuerdo con lo que explican allí sus autores, el programa está especialmente concebido para gestionar bases de datos que sirvan para cartografiar de forma automática y conseguir “el máximo aprovechamiento de toda la información lingüística que contiene una obra de estas características”, es decir, índices de atlas lingüísticos del tipo de los regionales españoles ya publicados. Busca, pues, *a posteriori* “la posibilidad de trabajar y manipular la información lingüística directamente en medios informáticos”. El alfabeto que usa no es propiamente el de la *Revista de Filología Española*, sino el adaptado habitual en los atlas regionales, si bien puede utilizar uno similar al Internacional. Cuenta con dos fuentes de letra diferenciadas para permitir la inclusión de todos los caracteres fonéticos.

El programa está diseñado para entornos Windows a partir de la versión 95. Desarrollado en Visual Basic 4, utiliza como base de datos Microsoft Access, realizando la comunicación a través del Motor Microsoft Jet. Los requerimientos de hardware son mínimos para los equipos que utilizamos hoy; basta un procesador Pentium y 32 Megabytes de memoria. Sí se necesita gran capacidad de almacenamiento en disco duro, pues cada uno de los mapas generados ocupa varios megas.

En realidad *Atlante* no cuenta con una sola base de datos, sino con tres teóricas, que en la práctica resultan ser siete bases de datos MS Access —a las que se accede, una vez abierto el programa, a través de la interfaz del programa desde el menú principal—, y su proyección en un mapa: la *base de localidades* (que comprende el nombre de los puntos del mapa con su código, los nombres de los informantes y las coordenadas de cada punto); la *base de informantes* (que incluye todos los datos relativos a los informantes: de dónde son sus

padres, su nivel de instrucción, sus viajes, su contacto con los medios de comunicación, etc.) y la *base de respuestas* (en ortografía normal o en alfabeto fonético, que contiene la información obtenida a lo largo de la encuesta). Para la transcripción, tiene instaladas dos fuentes diferentes y complementarias como ya se ha visto, utilizando un campo RTF (Rich Text Format) de Visual Basic. De esta forma se hace posible “mezclar” las fuentes en la introducción de los datos, su manejo y su visualización. Al ser datos en formato RTF, que debe “traducir” el programa, las transcripciones fonéticas no son directamente legibles en la base de datos.

La base de información lingüística es la más compleja y, en realidad, consta de varias bases que fraccionan la información, clasificándola por los campos que se estime conveniente —morfología, sintaxis, fonética, léxico— que, a su vez, como en estos últimos ejemplos, se pueden subdividir en campos más pequeños para agilizar el manejo de las bases.

La introducción de datos se realiza por pasos. Lo primero son los datos “estáticos”: localidades, informantes, etc. Posteriormente se introducen los materiales recogidos en las encuestas. Mi papel aquí es el de mera usuaria de la base de datos, en la que procedo a cargar mis materiales —con la colaboración inestimable de Carmen Bordón— y a elaborar en algún aspecto los mapas que ésta genera. Como no he intervenido en el diseño de *Atlante*, hablo del programa a partir de las informaciones que de él proporcionan sus autores y de mi experiencia como usuaria-autora de los mapas que corresponden a los cuestionarios de mujer. Supongo que los problemas que encuentro en la versión con la que trabajo para la elaboración son semejantes a los que se encuentran en otras bases de datos: esas típicas cosas que serían fáciles de corregir cuando ya se conoce a fondo el programa, debidas básicamente a una utilización complicada de la interfaz, que obliga a dar de forma repetitiva muchos pasos

que podrían haberse automatizado. Fundamentalmente el tener que abrir para cada localidad las ventanas que podrían haberse podido dejar abiertas para facilitar el mismo trabajo, y, en el proceso de volcado a mapa, el hecho de que las respuestas se superpongan en algunos puntos y sea necesario sacarlas fuera del mapa una a una, no automáticamente, lo mismo que la redacción física de las notas. El que las bases de datos no sean relacionales explica que los mapas sean estáticos, lo que obliga a generar un mapa nuevo cada vez que hay que hacer una mínima corrección, ya que *Atlante* no permite modificar datos en el mapa. A pesar de lo que pensábamos en los comienzos, cargar las bases de datos y acoplar sus contenidos armónicamente en un mapa legible es un trabajo tan lento y penoso como podía ser el del cartografiado tradicional y, sin embargo, es evidente que las posibilidades que este procedimiento permite a medio y largo plazo son enormes. Estoy segura de que, de aquí a que se publique el atlas completo, seremos capaces de apurarlas en muchos sentidos.

3. *www*

Pero una de las grandes ventajas que han venido a apoyar la utilización de las nuevas tecnologías en nuestra materia es la posibilidad de combinarlas con la ubicación de los materiales dialectales, en este caso de nuestro atlas, *on line*, en la red, en una página “en construcción”. Esto supone un acceso directo e inmediato a los datos, sin que sea necesario poseer el CD-ROM que lo contenga, ni tener totalmente terminado el atlas. A medida que vamos avanzando en la elaboración de mapas, estos mapas pueden ser consultados en Internet, y, al tiempo, pueden irse completando o corrigiendo, pueden añadirse notas, imágenes, etc., de modo que el atlas resulta así mucho más dinámico y

perfeccionable de lo que puede ser cualquier atlas en proceso de ser editado en papel. En el caso del ALeCMan, de momento sólo está prevista la publicación en Internet de un número importante de mapas, no el acceso directo a las bases de datos; por otra parte, se han colgado también materiales bibliográficos que explican la metodología y los avances de los trabajos. Sin duda de aquí a que terminemos la elaboración de los materiales surgirán posibilidades nuevas de aprovechar la tecnología. Hoy aquí van a tener oportunidad de ver atlas que han sacado gran provecho de ella: el pionero atlas sonoro “parlante” de Hans Goebel; ya han visto el cartografiado del *Atlas Lingüístico Vasco* de Gotzon Aurrekoetxea y Charles Videgain; podríamos acceder al CD-ROM del *Atlas Linguistique de la Corse* de Marie-Jo Steffanaggi Dalbéra o al elaborado por los investigadores del *Atlante Linguistico della Sicilia* con los mapas del trompo, y los vídeos en los que se ve cómo lo juegan los habitantes de la isla italiana, por no hablar de los trabajos que esta tarde les mostrarán los integrantes del equipo que trabaja en la Universidad de Barcelona sobre estos temas.

4. ALiR

Desde el año 1996 se está publicando el *Atlas Linguistique Roman* (ALiR), Contini (1996; 2001), publicación de un atlas interpretativo de tercera generación en soporte convencional. Organizado sobre un fascículo de mapas elaborados que se apoyan en un libro en el que se reúnen los estudios monográficos dedicados a esos mapas, supone el trabajo de distintos Comités lingüísticos que elaboran las síntesis de cada una de sus áreas que constituyen la base de la síntesis románica final; cf. García Mouton (2003).

El *Atlas Linguistique Roman* se hace con un cartografiado automático en la línea del experimentado en el ALE. El Instituto Poligráfico e Zecca dello Stato

lo publica en Roma previa entrega de los textos definitivos, incluidos los *tableaux de synthèse*, en los que relacionamos las formas fonéticas, codificadas por columnas, con los números que representan los puntos de la red románica, precedidos del código de cada dominio, y los listados que emparejan un símbolo determinado (identificado con un código) con los puntos del mapa en los que debe aparecer. Así se hacen los mapas del ALiR.

HÊTRE

TABLEAU DE SYNTHÈSE

1.1.1 Continuateurs du latin *FAGA (f.)	
itm. [faya] (f.)	ITA. 108.
1.2.1 Continuateurs de l'adjectif latin FAGEA (f.), substantivé	
prt. [fajɐ]	PRT. 5, 12, 26, 28, 29, 31, 33, 34, 39-45, 47, 49, 51, 54.
glc. [faja]	
cast. [aja]	
cal. [ʔaja]	GLC. 2, 2b, 3, 4, 4b, 20, 20b, 21, 22, 24, 24b, 25b, 26, 27, 27b, 47.
occ. [hʔaja]	
its. [ʔaza]	
itc. [fʔaʒa]	ESP. 5, 6, 7, 8, 11, 12, 13, 17, 17b, 35, 36, 39b, 40, 41, 42b, 43b, 48, 55, 58, 59, 60, 64, 65, 75, 81, 82, 83, 85, 97, 99, 101, 105, 121.
itm. [fʔajə]	
	CAT. 222, 241.
	ITA. 40, 48, 49, 53-55, 104.
1.3.1 Idem. Avec un suffixe diminutif -ŌLA (f.)	
occ. [haʒʔlə]	FRA. 247, 263.
1.4.1 Idem. Avec un suffixe -ĪCIA: FAG-ĪCIA (f.)	
oïl [fejʔis]	FRA. 62-64, 78, 79.
1.5.1 Dérivé tardif sur le radical [fu] < FAGUM + suffixe [ɛl] (f.)	
oïl [fwel]	FRA. 105, 119.
2.1.1 Continuateur de l'accusatif latin FAGUM devenu masculin	
arag. [faʒo]	ESP. 17b, 42, 44, 44b, 45, 45b, 67b.
bsq. [bʔaypa]	

En los últimos años se ha planteado, como parece lógico, el aprovechamiento conjunto de todos esos datos a partir de procedimientos informáticos que permitan elaboraciones posteriores y búsquedas no encaminadas únicamente al cartografiado. Ya en la reunión del Comité de Redacción que se celebró en Niza en diciembre del año 2000, se habló por primera vez de cómo proceder en ese sentido y se encargó a Manuel González González, director del Comité gallego, con experiencia en estos temas, que estudiase las posibilidades de generar una base de datos retroactiva que se pueda adoptar para el futuro. En el último coloquio, el número dieciséis, que nos ha reunido a todos los participantes en el Piamonte el pasado mes de octubre de 2003, Manuel González comunicó que ya estaba muy adelantado el programa básico de recuperación de formas. De este modo se abre una nueva vía para el aprovechamiento de todos los datos reunidos en el ALiR con unos índices que estarán disponibles en la red.

5. Madrid

En el marco del ALiR, estamos completando desde el CSIC las encuestas dialectales que curiosamente faltaban por hacer precisamente en la Comunidad de Madrid. Es evidente que nuestra Comunidad no es una provincia cualquiera, sino que presenta unas características que la hacen interesantísima, por estar localizada en una zona de transición, por estar permeada casi toda ella de las influencias que irradian desde la capital, por constituir en un radio muy amplio de kilómetros alrededor de la gran ciudad un crisol donde se mezclan las hablas de las gentes venidas de zonas rurales a veces lejanas, y ahora también extranjeras.

Estamos haciendo las encuestas de la Comunidad con un cuestionario redactado por Esther Hernández, Isabel Molina y por mí, cuestionario que debe mucho a los cuestionarios de los atlas regionales anteriores, y que aprovecha especialmente nuestra experiencia —la de Isabel Molina y la mía— en las encuestas del ALeCMan, al tiempo que incluye, cosa que no ocurre con los cuestionarios de otras áreas, todas las preguntas que contiene el cuestionario del ALiR. Con la ayuda de Carlos García, ingeniero informático que ha trabajado con nosotras y que ahora colabora desde el Cindoc, estamos desarrollando un programa propio con el que esperamos poder proporcionar en poco tiempo una información muy ágil que comprenda no sólo mapas, sino también textos, sonido, imágenes, etc. Está montado sobre una base de datos relacional, que permite la utilización de cualquier motor de base de datos (Access, MySQL, etc.) y facilita el acceso y la corrección de datos. Como parece deseable, recalcula la posición de los puntos en el mapa para que no se superpongan las respuestas y permite la corrección de los datos sobre los mapas.

6. Los atlas en la red. Inconvenientes y ventajas

Como se ve, ahora ya no hablamos de utilizar las nuevas tecnologías sólo para organizar, acortar o agilizar el proceso de hacer un atlas. Es evidente que el valor añadido de este tipo de edición no lo constituyen sólo las posibilidades de visualización que permite, sino la capacidad de investigar dentro de los datos: preguntas, estadísticas, frecuencias, dialectometría. Introducidos en la base de datos, éstos serán accesibles en el futuro para cualquier estudio.

Pero hay más. Ahora las nuevas tecnologías, combinadas con Internet, permiten otro tipo de edición, no sólo la edición electrónica, sino la edición en la red, con todo lo que eso supone. Supone, en primer lugar, un abaratamiento grande de los costes reales de edición; supone unas posibilidades desconocidas hasta hace unos años, como puede ser el editar en color, con imágenes, con sonido; puede suponer, si así se concibe el proceso, la generación de mapas “a medida” a partir de las bases de datos, no la simple consulta de mapas fijos, y muchas otras ventajas, como las de trabajar en grupo a través de la red. Pero todo esto podría cargarse en un CD-ROM. La enorme ventaja de ubicar los mapas en la red es que de ese modo casi se eliminan los problemas de distribución de unas obras raras, caras y, por eso mismo, difíciles de encontrar y de consultar.

Esto plantea una pregunta: ¿se acabaron los atlas en papel? Es una probabilidad, pero lo deseable sería compatibilizar los dos tipos de edición. Personalmente prefiero trabajar sobre papel, no es igual trabajar en la pantalla. Es cierto que a partir de las bases de datos se pueden conseguir muchas cosas que no da el inmovilismo del papel, sobre todo informaciones organizadas, pero no hay nada para un geolingüista como poder ver la disposición de las respuestas en el espacio.

Deberíamos conseguir editar fácilmente también en papel a partir del soporte informático. Porque supongamos que la red siempre será la red, pero ¿y los servidores? ¿Estarán siempre ahí?, ¿mantendrán nuestra información?, ¿no dependerán de organismos y de personas que cambian, y cuyos intereses cambian también? Y, si es así, ¿no pueden hacer perecer nuestros atlas? ¿Y los temas de seguridad? ¿Y los derechos de propiedad intelectual? Los CD-ROM o sus desarrollos podrían ser la respuesta: equivalentes al papel, sólo suponen otro soporte, no sabemos si

tan duradero; pero, a pesar de ser mucho más ágiles y teóricamente mucho menos deteriorables que el papel, ¿no estarán sometido a la larga a la tiranía con la que la imparable innovación tecnológica va dejando atrás tantos productos que hace obsoletos en su avance? Esa advertencia que dice *requisitos del sistema* ¿no podrá llegar a ser una trampa? Porque para leer lo impreso en papel no necesitamos más que tenerlo; pero no podemos asegurar que lo que hoy está colgado en Internet vaya a estarlo en el futuro, ni que dentro de veinte años cualquier ordenador pueda leer lo que lee hoy, ni que dentro de un tiempo las grandes bibliotecas vayan a dedicar una parte importante de sus recursos a convertir esas obras pasándolas a los soportes del momento para mantenerlas vivas. Saludemos pues los avances que nos permiten las nuevas tecnologías, pero aprovechémoslas también para conservar nuestros trabajos en soportes alternativos.

Referencias bibliográficas

- ALVAR, Manuel (1984): *Informática y Lingüística*, Málaga: Ediciones Ágora.
- CONTINI, Michel (dir.) (1996): *Atlas Linguistique Roman (ALiR)*, vol. I, *Présentation; Commentaires; Cartes*, Roma: Istituto Poligrafico e Zecca dello Stato.
- _____ (dir.) (2001): *Atlas Linguistique Roman (ALiR)*, vol. IIa, *Commentaires; Cartes*, Roma: Istituto Poligrafico e Zecca dello Stato.
- GARCÍA MOUTON, Pilar (2003): «El *Atlas Linguistique Roman (ALiR)*: Historia y características de un proyecto europeo», en *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, Madrid: Arco/Libros, 611-624.
- GARCÍA MOUTON, Pilar; MORENO FERNÁNDEZ, Francisco (1988): «El Proyecto de un Atlas Lingüístico (y etnográfico) de Castilla-La Mancha», en *Actas del Primer Congreso Internacional de Historia de la Lengua Española*, II, Madrid: Arco/Libros, 1461-1480.

- _____ (1993): «Sociolingüística en el *Atlas Lingüístico (y etnográfico) de Castilla-La Mancha*», en *Actas del Primer Congreso Anglo-Hispano*, tomo 1, *Lingüística*, Madrid: Castalia, 139-149.
- _____ (2003): *Atlas Lingüístico y etnográfico de Castilla-La Mancha* (página en construcción), www.uah.es/otrosweb/alecman.
- ZIAMANDANIS, Claire (1996): «Dialectología y ordenadores», en Alvar, M. (dir.), *Manual de dialectología hispánica. El Español de España*, Barcelona: Ariel, 55-62.

El mètode dialectomètric: aplicació de l'anàlisi multivariant a la classificació de les varietats del català*

Esteve Clua

Universitat Pompeu Fabra

1. Naixement i evolució de la dialectometria

En l'àmbit de la variació dialectal, durant les darreres tres dècades s'han desenvolupat i aplicat un nombre considerable de mètodes de descripció i de classificació de varietats lingüístiques geogràfiques basats en criteris quantitius. Aquests mètodes solen compartir el fet de tractar de forma sintètica i global el conjunt de dades d'una enquesta o atlas dialectal amb finalitats classificatòries. Amb aquesta finalitat utilitzen anàlisis estadístiques multivariants per a la quantificació i el tractament de les dades; amb el terme anàlisi multivariant (*Multivariate Analysis*) hom fa referència a un nombre considerable de tècniques de descripció i d'anàlisi d'un conjunt

*Aquest treball forma part del projecte VALDIC ("Anàlisi i interpretació de la variació lingüística dialectal a partir de l'explotació d'un corpus oral"), finançat pel Ministerio de Ciencia y Tecnología (BFF2001-3798). També es beneficia de recursos obtinguts del Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (grup 2001SGR-00004). Es pot consultar més informació sobre el projecte a <<http://www.ub.edu/lincat>>.

d'elements a partir de diverses variables observades, sense que, a priori, s'atorgui un estatus qualitatiu especial a cap d'elles.

Tots aquests mètodes se solen agrupar sota el rètol de dialectometria. Aquesta disciplina apareix al principi de la dècada dels setanta en l'àmbit de la lingüística romànica, lligada estretament a la geolingüística. I es caracteritza per l'adopció del concepte de distància lingüística com a eina bàsica de la classificació dialectal. El concepte de distància fou manllevat de l'àmbit científic de l'anàlisi de dades, en el qual s'associa —en general— a la quantificació de les similituds o de les diferències entre individus, poblacions o grups de poblacions. L'aplicació de mètodes estadístics a ciències diverses és un fenomen molt corrent. La biologia, la medicina, l'economia, la sociologia, la psicologia —per esmentar-ne algunes— s'han beneficiat de l'aplicació de models propis de l'anàlisi de dades. En molts d'aquest casos es tractava d'establir classificacions de determinades entitats a partir de l'anàlisi multivariant. La dialectometria, en part, és el resultat d'aquesta comunicació interdisciplinària entre la dialectologia geogràfica i l'anàlisi de dades.

El primer a utilitzar el terme dialectometria va ser Jean Séguy, que juntament amb Henri Guiter ha estat considerat el pare de la disciplina (Séguy 1971, 1973; Guiter 1973). L'impuls més important de la disciplina es deu, però, al professor Hans Goebl de la Universitat de Salzburg, que amb una sòlida base de taxonomia numèrica ha anat polint els seus propis mètodes dialectomètrics a partir de treballs sobre materials de l'*Atlas Linguistique de la France* (ALF), de l'*Atlante Italo-svizzero* (AIS), de l'*Atlante linguistico del ladino dolomitico e dei dialetti limitrofi* (ALD) i del *Survey of English Dialects* (SED) (Goebl 1981, 1987, 1989, 1992, 1993 i 1997). La metodologia dialectomètrica de Goebl constitueix un aparell força complex que ha anat variant i millorant al llarg de les investigacions

realitzades, conscient com és aquest autor del fet que els possibles arranjaments classificatoris són múltiples i variables i depenen del context metodològic i de la finalitat de la investigació. En l'actualitat i en el marc del Projecte Dialectomètric de la Universitat de Salzburg, que dirigeix, s'ha desenvolupat el programa VDM (*Visual Dialectometry*),¹ que és el paquet d'eines electròniques per al tractament dialectomètric més complet que existeix en l'actualitat.

El procés d'anàlisi dialectomètrica s'inicia amb la transposició de les dades d'un determinat atlas lingüístic a la matriu de dades de comparació. La matriu de dades recull les diferents poblacions analitzades i, per a cada una d'aquestes, les variants que presenten en relació amb cada una de les variables —en la bibliografia dialectomètrica anomenades taxats i atributs, respectivament. El pas següent suposa la selecció de l'índex de similitud, l'aplicació del qual permetrà la transformació de la matriu de dades en la matriu de similitud. La utilització d'un determinat índex s'explica, entre altres motius, per la necessitat d'optimitzar els resultats de l'aplicació dialectomètrica, tot minimitzant aspectes negatius com poden ser els buits que apareixen en les respostes de les enquestes dialectals.

El darrer pas consisteix a representar de la manera més adequada possible la matriu de similituds. Els mètodes utilitzats per Goebl han estat força variats: cartogràfics, mitjançant l'aplicació dels polígons de Thiessen —i les anomenades *cartes choroplètes*— o el mètode d'interpunts; o dendrogràfics, mètode de representació geomètrica de la matriu de similitud, mitjançant un arbre d'estructura jeràrquica. Prèviament, però, l'investigador ha d'haver triat l'algorisme matemàtic —a partir del qual es genera la representació classificatòria— més adient, d'acord amb les finalitats de la seva recerca.

¹ Vegeu <http://ald.sbg.ac.at/dm/>.

Els estudis dialectomètrics s'han multiplicat darrerament arreu del món.² La possibilitat generalitzada de disposar d'eines informàtiques cada cop més potents, que permeten el tractament de grans quantitats de dades i la realització de càlculs extensos, ha facilitat aquesta proliferació. A la vegada, també s'ha ampliat molt l'espectre de tècniques que es fan servir per al càlcul de la distància lingüística. Així, al costat dels mètodes primerencs, com ara l'anomenat global de Gutter o l'interpuntual de Goebel, han aparegut nous mètodes basats en diferents tècniques de l'anàlisi estadística multivariant. Entre aquestes, les més utilitzades són les d'anàlisi de conglomerats (*Cluster Analysis*) i l'anàlisi de proximitats (*Multidimensional Scaling*).

2. Objectius de la dialectometria

Originàriament, la dialectometria va sorgir com un intent de defugir l'obligatorietat d'haver de triar entre determinades característiques lingüístiques a l'hora d'establir classificacions i delimitacions dialectals. El principal retret que s'ha fet a la noció d'isoglossa en tant que criteri bàsic per a la delimitació de varietats dialectals té a veure amb l'arbitrarietat que implica. Efectivament, les divisions d'àrees dialectals establertes mitjançant aquest mètode solen basar-se en un petit nombre de trets lingüístics, que són el resultat d'una tria entre el gran conjunt de trets que constitueixen els sistemes lingüístics de les àrees analitzades. El problema rau en el fet que no disposem d'una jerarquització d'isoglosses prou fonamentada que justifiqui una determinada tria de característiques lingüístiques a l'hora d'establir les

² Vegeu a Goebel (1992:433-434) un recull bibliogràfic de treballs dialectomètrics.

agrupacions dialectals; com a molt, es pot establir una ordenació de les isoglosses en funció de la seva rellevància estructural. Per tant aquesta metodologia comporta ineludiblement una dosi important de subjectivitat per part de l'investigador.

La voluntat d'eludir aquesta arbitrarietat en la selecció dels elements lingüístics, sobre la base dels quals s'estableix la delimitació dialectal, és una de les raons principals que ha portat els investigadors a basar la determinació de varietats lingüístiques en el criteri quantitatiu (Veny 1992: 205-206). Des d'aquesta perspectiva, la delimitació i la classificació dialectals es duen a terme a partir de l'aplicació de tractaments estadístics a tot el conjunt de dades que pot furnir una enquesta dialectal; a partir de la quantificació de les similituds o de les diferències entre les varietats estudiades es pot comprovar la distància lingüística que les separa i, per tant, establir-ne la classificació, una classificació que serà el resultat del tractament sintètic i global del conjunt de dades lingüístiques.

Tanmateix, seria il·lusori pensar que el fet de defugir l'arbitrarietat en aquest aspecte concret implica la manca total de tries subjectives durant tot el procés. La necessitat d'elecció amb un cert grau de subjectivitat acompanya l'investigador en molts moments del procés classificatori: en la definició de l'enquesta dialectal, en el tipus d'anàlisi lingüística a partir de la qual establirà la comparació, en la mesura de similitud utilitzada per definir la distància lingüística, en el tipus d'algorisme matemàtic utilitzat per establir els agrupaments de les diferents varietats lingüístiques, etc. D'altra banda els tractaments dialectomètrics no han estat exempts de mancances, que han hipotecat tant l'adequació dels seus resultats, com la seva pretesa objectivitat. Algunes d'aquestes mancances estan relacionades amb factors com els següents (Clua 2000):

a) La utilització indiscriminada de les dades d'estudis geolingüístics no concebuts per ser analitzats amb tècniques d'anàlisi multivariant com a punt de partida del tractament dialectomètric. De fet, una de les finalitats dels mètodes dialectomètrics —inicialment— era poder tractar la gran quantitat de dades dels atlas lingüístics, perquè aquests no es convertissin en 'cementiris de dades' (Vierek 1987:11). Un tractament quantitatiu requereix que el conjunt de dades sobre el qual s'aplica sigui representatiu del sistema lingüístic o de l'estrat lingüístic que es vol analitzar, sobretot si l'anàlisi de les dades es fa amb mètodes multivariants que valoren per igual les diferents variables tingudes en compte. I aquesta no és una característica gaire comuna en alguns dels estudis geolingüístics que han estat tractats amb mètodes dialectomètrics, molt centrats en aspectes lèxics no sistemàtics.

b) La manca d'una anàlisi lingüística coherent a l'hora de transformar les dades fonètiques, obtingudes a partir del qüestionari lingüístic, en les variables de comparació que serveixen de base per al tractament quantitatiu. Atesa la gran variació que es pot donar en qualsevol estudi dialectal, els tractaments dialectomètrics han tendit a una certa 'tipificació' de les dades —és a dir, a una certa simplificació de la variació observada—; si aquesta 'tipificació' no recolza sobre criteris lingüístics coherents, pot ocórrer que es barregin criteris sincrònics i diacrònics o que les variables en què el procés classificatori recolza no reflecteixin adequadament la variació real observada.

c) El considerable grau d'arbitrarietat que pot comportar la tria del tipus de mesura de similitud, a partir de la qual s'estableix la distància lingüística, i de l'algorisme de classificació que permet transformar la distància inicial en una distància ultramètrica i obtenir la posterior representació dendrogràfica. L'investigador pot triar entre diferents possibilitats de

mesura de similituds o d'algorisme classificatori, que donaran resultats amb diferències considerables. Per això és indispensable utilitzar mesures de control que permetin contrastar la fiabilitat de les classificacions obtingudes.

L'evolució i la millora de la disciplina ha estat molts cops relacionada amb la minimització dels aspectes que requereixen una intervenció amb un elevat grau de subjectivitat per part de l'investigador, amb l'augment de les mesures de control i de fiabilitat dels resultats, o amb el replantejament d'altres aspectes, com ara el tipus d'anàlisi lingüística en què es basa la transposició de les dades "brutes" de l'enquesta lingüística a la matriu de dades dialectomètrica, o el conjunt de dades en què es basa una determinada classificació. En una anàlisi quantitativa tots aquests aspectes s'han de tractar amb molta més cura que no pas en un tractament qualitatiu, ja que —cal recordar-ho— el tractament quantitatiu, en moltes ocasions, considera totes les variables de comparació com si fossin del mateix rang.

3. Aplicacions de la dialectometria a l'àmbit català

Les aplicacions dialectomètriques a les varietats dialectals catalanes han estat poc nombroses i més aviat esparses. Hi ha un primer grup de treballs dialectomètrics que coincideix amb els primers passos de la disciplina. Ens referim als treballs de Guiter (1973, 1978), Sardà & Guiter (1975) i Costa (1977). Es tracta d'aplicacions quantitatives a les dades de l'*Atlas Lingüístic dels Pirineus Orientals* i de l'*Atlas Lingüístic de Catalunya*, els resultats de les quals, realment sorprenents en alguns casos, han estat força contestats (vegeu Veny 1986: 25-27).

Després d'aquests treballs primerencs, cal destacar el treball de Polanco (1992),³ en què s'apliquen els mètodes dialectomètrics de Guiter i de Goebel a les dades de l'*Atlas Lingüístico de la Península Ibérica* (ALPI) corresponents al català. Concretament, l'autor se serveix del mètode d'interval·lització de Guiter i de l'Índex General d'Impermeabilitat de Goebel per descriure, amb un resultat molt aproximat als de la geolingüística tradicional, el grau de diferenciació lingüística existent entre les varietats del català. Els resultats serveixen a l'autor per confirmar i reforçar, a partir de dades estrictament lingüístiques, el continuum lingüístic entre les varietats valencianes i catalanes.

Un treball remarcable entre les aplicacions de mètodes dialectomètrics a les varietats del català és el de Ochoa (1998). Aquest autor fa una aplicació de l'anàlisi de conglomerats (*Cluster Analysis*) a les varietats lingüístiques de la Marina, a partir de dades de tipus fonètic, morfològic i lèxic. Però el que és més destacable d'aquest treball és la defensa clara i concisa de la necessitat d'ampliar les relacions interdisciplinàries entre la geolingüística i l'anàlisi de dades "perquè entre l'exposició descriptiva de les dades brutes i l'elaboració de les conclusions finals no hi intercedisca només el coneixement intuïtiu".

Aquesta tònica general de treballs puntuals i esparsos, que caracteritza les aplicacions dialectomètriques als dialectes catalans, sembla que ha començat a canviar d'ençà de finals dels anys noranta. Al voltant del Corpus Oral Dialectal (COD) de la Universitat de Barcelona, han aparegut diversos treballs que tenen com a finalitat analitzar la variació dialectal de manera global amb mètodes estadístics; a més a més, els projectes de recerca centrats en l'explotació de les dades del COD compten sempre entre els seus

³ Segons assenyala l'autor, aquest article havia estat presentat al XVII Congrés Internacional de Lingüística i Filologia Romàniques, que tingué lloc a Ais de Provença a l'agost-setembre de 1982.

objectius prioritaris amb la determinació de la distància lingüística entre les varietats catalanes mitjançant mètodes manllevats a l'anàlisi de dades (vegeu Clua *et. al.* en premsa). Entre aquests treballs destaquen el de Viaplana (1999) i el de Clua (1999). El primer és un estudi de caràcter interdisciplinari que combina diferents aspectes de la dialectologia geogràfica, de la dialectometria, de la lingüística generativa i de l'anàlisi de dades, amb l'objectiu de descriure els aspectes més rellevants de la morfologia regular de les varietats nord-occidentals, de reflectir-ne les relacions i les tendències de proximitat o de llunyania lingüística i de proposar-ne, finalment, una classificació a partir de la variació analitzada. El segon treball, que té com a punt de partida Viaplana (1999), analitza la morfologia flexiva de les varietats valencianes i en proposa una classificació basada en el concepte de distància lingüística, tot replantejant alguns aspectes del procés dialectomètric.

4. El procés dialectomètric

A continuació exemplificarem el procés de l'anàlisi dialectomètrica a partir d'aquest darrer treball. L'estudi de Clua (1999), en el qual es proposa una classificació de les varietats valencianes a partir de la morfologia flexiva, parteix de la idea que l'afinitat entre dues varietats dialectals és inversament proporcional a les divergències existents entre els sistemes lingüístics respectius. Aquestes divergències entre dos sistemes constitueix el que anomenem "distància lingüística", que, com hem vist, és una noció manllevada de l'anàlisi de dades. La major o menor distància lingüística és el factor que determina l'agrupament de les diferents varietats en classes (sub)dialectals. Atès que les característiques lingüístiques a partir de les

quals s'estableix la distància (variables de comparació) són múltiples i que les dades resultants de l'anàlisi lingüística són considerables, s'utilitzen tècniques d'anàlisi estadística multivariant per poder-les tractar adequadament.

Tot seguit presentem l'esquema del procés adoptat en aquest treball per a la definició de la distància lingüística de les varietats valencianes. I a continuació n'expliquem els aspectes més rellevants des de l'òptica de l'anàlisi quantitativa utilitzada.

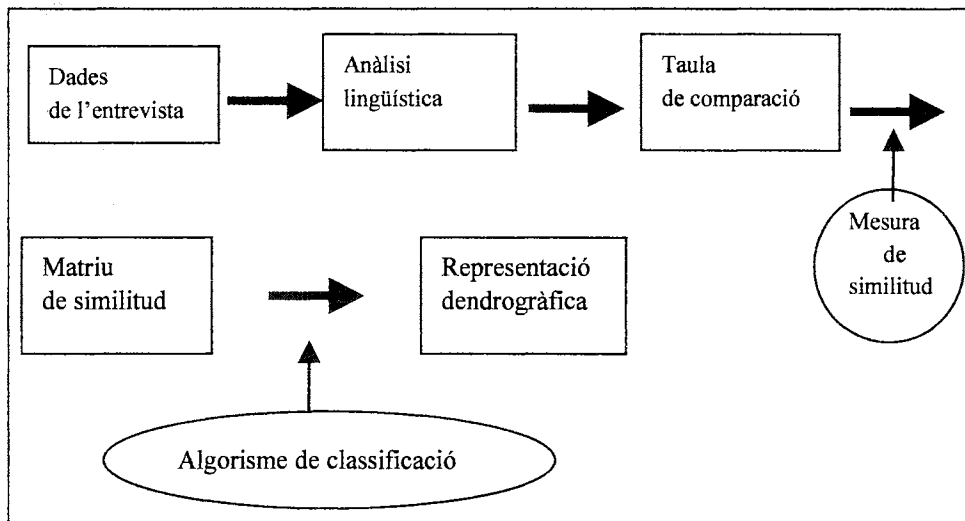


Figura 1.

En principi, en aquest estudi no es tractava d'aplicar una determinada tècnica quantitativa a les dades d'una entrevista preexistent o als mapes d'un determinat atlas dialectal, sinó que tant el disseny del qüestionari com la realització de l'entrevista i els altres estadis de la investigació s'han dut a terme tenint en compte que un dels objectius finals era el tractament quantitatiu de la variació lingüística observada. Això vol dir que, per

exemple, en el disseny del qüestionari —un cop descartat, per inviable, el tractament exhaustiu de la flexió— es va intentar que abastés una mostra prou representativa d'aquest àmbit lingüístic, que permetés arribar a conclusions generalitzables. El conjunt de dades analitzades comprèn exhaustivament les formes dels articles, les formes simples dels clítics pronominals i els diferents paradigmes dels verbs regulars; pel que fa als adjectius, ultra les formes regulars, s'analitzen els casos més generals entre els considerats "irregulars". Creiem, doncs, que es tracta d'un conjunt representatiu de la flexió valenciana.

Quant a l'entrevista, després d'una primera anàlisi global de les dades obtingudes en l'aplicació principal, se'n va fer una de comprovació. Entre els objectius no secundaris d'aquesta segona recollida de dades hi havia el d'intentar cobrir les respostes deficientes o nul·les de la primera entrevista, ja que aquest tipus de mancances pot distorsionar els resultats d'un tractament quantitatiu com el que s'ha utilitzat en aquest estudi.

4.1. L'anàlisi lingüística de les dades

L'anàlisi lingüística de les dades constitueix el següent pas del procés. Aquest es un factor força rellevant en l'anàlisi multivariant, ja que la matriu de comparació no està constituïda per les dades fonètiques resultants del procés de transcripció de les respostes de l'entrevista, sinó que els elements que la componen són el resultat de l'aplicació de l'anàlisi lingüística a aquestes dades. L'anàlisi lingüística emprada es basa en els principis de la fonologia generativa, que es caracteritza per analitzar la matèria sonora de les llengües a partir de dos nivells: el de les formes subjacents o bàsiques, constituïdes per segments fonològics, i el de les formes fonètiques o

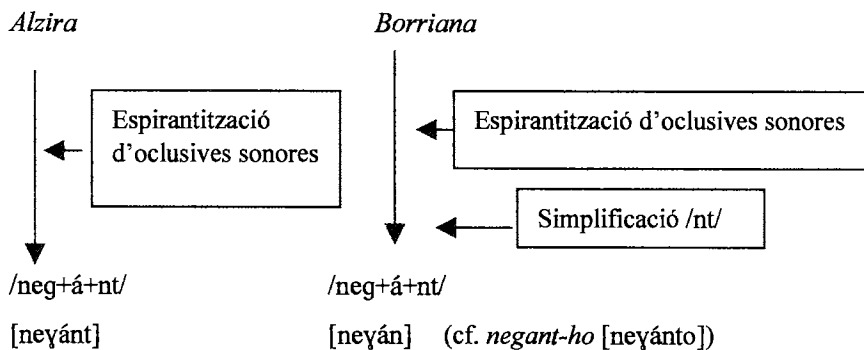
superficials. La relació entre ambdós nivells s'explica mitjançant els processos fonològics que permeten derivar les realitzacions fonètiques de les formes subjacents. D'aquesta manera s'aconsegueix simplificar la descripció lingüística i, a la vegada, es poden establir generalitzacions sobre el funcionament i l'estructura de les varietats lingüístiques.

Des d'aquest enfocament la variació morfològica existent entre dues varietats pot ser bàsicament de dos tipus: de tipus morfològic, que implica l'existència de segments morfològics subjacents diferents (1); o de tipus fonètic, en què els segments morfològics subjacents són coincidents i la diferència es produeix per divergències en els processos fonològics que deriven les realitzacions fonètiques a partir de les representacions fonològiques dels morfemes (2).

- (1) 1a persona del singular del present d'indicatiu:

<i>Vinaròs</i>	<i>l'Alcora</i>	<i>resta de varietats valencianes</i>
/kánt+o/	/kánt+a/	/kánt+e/
↓	↓	↓
[kánto]	[kánta]	[kánte]

- (2) Gerundi:



A (1) les divergències en la vocal final només es poden explicar a partir de formes subjacents diferents; altrament, per explicar l'alternança [o], [a] i [e], caldria postular uns processos fonològics, condicionats morfològicament, que resultarien totalment *ad hoc*. En canvi, a (2) el segment morfològic que marca el gerundi, /nt/, té una mateixa representació fonològica en les dues varietats —a Borriana, la presència del segment dental en la representació fonètica d'aquest gerundi, quan va seguit d'un pronom enclític, demostra la pertinència d'aquesta forma subjacent—; la diferència fonètica, en aquest cas, és deguda a l'existència, en la varietat de Borriana, d'un procés que simplifica els grups consonàntics /nt/ en situació final de mot, procés que no ocorre a Alzira.

D'altra banda, una anàlisi lingüística d'aquest tipus permet discriminar clarament el caràcter lingüístic dels diferents fenòmens que entren en joc en la variació. Com s'afirma a Viaplana (1994) “la distinció entre elements predictibles i elements impredictibles en l'estructura lingüística permet la discriminació de fenòmens diferencials crucials en la variació dialectal que, en absència d'aquesta distinció, queden amalgamats en la simple distinció de les formes”. Així, per exemple, si analitzem les divergències que es produeixen en les formes pronominals de (3) des de l'òptica de la dialectologia tradicional o estructuralista, sembla que la distància lingüística entre les tres varietats sigui molt similar. Ontinyent i Sagunt difereixen únicament en la forma proclítica davant de verb començat en consonant; Sagunt i Elx coincideixen en totes les formes excepte en la que apareix darrere l'imperatiu; per la seva banda, Ontinyent i Elx es diferencien tant per la forma proclítica davant verb començat en consonant, com per l'enclítica darrere verb acabat en vocal.

(3)

	Ontinyent	Sagunt	Elx
a. em pense	[em.pén.se]	[me.pén.se]	[me.pén.se]
b. m'ocupe	[mo.kú.pe]	[mo.kú.pe]	[mo.kú.pe]
c. pensar-me	[pen.sá r.me]	[pen.sá r.me]	[pen.sá r.me]
d. porta'm	[pór.tam]	[pór.tam]	[pór.tame]

Des de l'òptica de la fonologia generativa, l'anàlisi de la variació entre aquestes formes és força diferent. Les varietats d'Ontinyent i Sagunt, coincidents en els exemples b, c i d, tenen una mateixa representació fonològica: /m/. El segment [e] que apareix en aquestes formes es pot justificar com una vocal epentètica, inserida per necessitats de sil·labificació (4). Les diferències que es produeixen en la forma proclítica davant de verb començat en consonant són degudes al tipus d'epèntesi. A Ontinyent aquest procés sempre és perifèric —és a dir, el segment epentètic sempre se situa a l'exterior del grup clític—; en canvi, a Sagunt la inserció de l'epèntesi té lloc sempre a la dreta del clític.

(4)

	Ontinyent/ Sagunt	Elx
a. em pense	*[m+pén.se]	[me+pén.se]
b. m'ocupe	[mo.kú.pe]	[m(e)+o.kú.pe]
c. pensar-me	*[pen.sá r+m]	[pen.sá r+me]
d. porta'm	[pór.ta+m]	[pór.ta+me]

El cas d'Elx és força diferent. A les varietats que presenten formes com [pór.ta+me] la representació subjacent d'aquest pronom no potser /m/, ja que l'estructura sil·làbica del grup clític integrat per l'imperatiu i el pronom no

requereix cap tipus de segment vocàlic. Ha de ser, per tant, /me/. Tenint en compte això, les divergències entre les varietats de Sagunt i Ontinyent, d'una banda, i Elx, de l'altra, s'expliquen, pel que fa a les formes, per la no-coincidència en la representació fonològica, /m/ en front de /me/, i per l'actuació de diversos processos fonològics: epèntesi, elisió de la vocal del clític (me+ocupe, a Elx), etc.

(5)

Divergències entre aquestes varietats:

Ontinyent: /m/ i epèntesi [e] perifèrica

Sagunt: /m/ i epèntesi [e] a la dreta

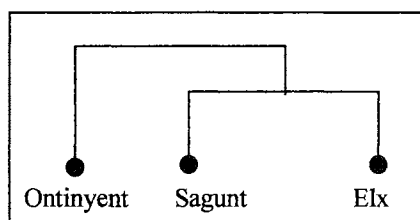
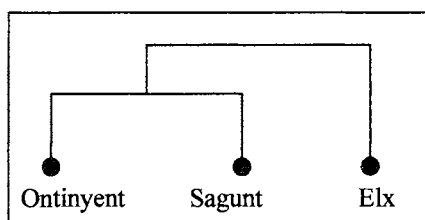
Elx: /me/ i elisió de [e] davant verb començat en vocal.

La distinció qualitativa que una anàlisi d'aquest tipus ens permet palesar té també implicacions considerables en un tractament quantitatiu com el que aquí es du a terme. Si en els casos de (3) definim la distància lingüística (el conjunt de divergències observables) entre aquestes varietats a partir de les dades fonètiques, arribarem a la conclusió que presenten una distància similar; ja que entre Ontinyent i Sagunt només hi ha una divergència: [et] en lloc de [te] (a.); entre Sagunt i Elx només hi ha també una divergència: [t] en lloc de [te] (d.); mentre que entre Ontinyent i Elx la distància seria una mica més gran, ja que presenten dues diferències: [et] en lloc de [te] (a.) i [t] en lloc de [te] (d.). D'acord amb això, podríem establir la següent matriu de diferències i la corresponent representació arbòria:⁴

⁴ Aquesta representació pot presentar dues formes perquè Sagunt es troba en una situació d'equidistància amb les altres dues varietats.

(6)

	Ontinyent	Sagunt	Elx
Ontinyent			
Sagunt	1		
Elx	2	1	



En canvi, si apliquem als casos de (3) l'anàlisi lingüística generativa i distingim entre diferències superficials i subjacents, el resultat és força diferent. Des d'aquesta perspectiva, com ja hem dit, la varietat d'Elx s'aparta clarament de les altres dues. Ara, doncs, les matrius de diferències i la representació arbòria corresponents són les següents:

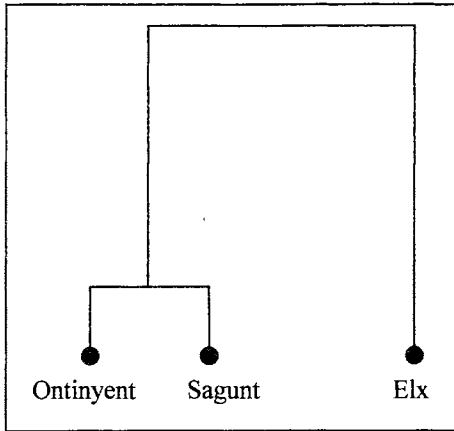
(7)

Diferències fonològiques

	Ontinyent	Sagunt	Elx
Ontinyent			
Sagunt	0		
Elx	4	4	

Diferències fonètiques

	Ontinyent	Sagunt	Elx
Ontinyent			
Sagunt	1		
Elx	3	3	



Una anàlisi com aquesta permet copsar en cada cas el tipus de variació que es produeix i determinar-ne el grau amb molta més precisió. Des d'aquesta perspectiva, les divergències de la varietat d'Elx respecte de les altres dues són de tipus morfològic i fonètic; són, per tant, considerablement més importants que les que separen les varietats de Sagunt i Ontinyent, que només tenen a veure amb la posició en què s'insereix el segment epentètic. D'acord amb aquests supòsits, l'anàlisi lingüística desenvolupada en aquest treball té com a finalitat la descripció explícita de les dades recollides en l'entrevista, mitjançant l'establiment de les formes subjacents i la determinació dels processos fonològics que relacionen aquestes amb les formes superficials.

4.2. Càlcul de la distància lingüística

Amb les formes subjacents i els processos fonològics definits s'estableixen les taules de comparació, que són la base per poder definir posteriorment la similitud i la distància lingüístiques. Les taules de

comparació posen en relació un conjunt de variables —en aquest cas, formes subjacents i processos fonològics— amb un conjunt d'individus —els informants— i presenten l'estructura següent:

(8)

Variables	X_1	X_2	X_3	...	...	...	X_p
Individus							
Informant 1	X_{11}	X_{12}	X_{13}	...	...	...	X_{1p}
Informant 2	X_{21}	X_{22}	X_{23}	...	...	...	X_{2p}
Informant 3	X_{31}	X_{32}	X_{33}	...	...	...	X_{3p}
...	...	...	...	...	...	...	...
Informant n	X_{n1}	X_{n2}	X_{n3}	...	...	...	X_{np}

A partir de les taules de comparació, i mitjançant el paquet informàtic SAS,⁵ es determinen les matrius de similitud, que constitueixen el següent pas del procés quantitatiu. Abans, però, cal establir la mesura de similitud a partir de la qual el programa informàtic comptabilitzarà les coincidències i establirà les similituds entre els informants.

L'elecció de la mesura o índex de similitud té una gran rellevància per al resultat final del procés, ja que segons el tipus de mesura establert els resultats poden variar considerablement; sobretot quan les matrius de comparació presenten caselles nul·les, a causa de respostes errònies en l'entrevista o a la manca de respostes, o quan les dades impliquen respostes múltiples. En aquests casos la mesura de similitud que se sol utilitzar és el percentatge de les coincidències, en relació amb el total d'elements comparats entre dues varietats. En aquest estudi, el primer d'aquests factors,

⁵ SASInstitute, INC (1990).

com ja hem assenyalat, era irrellevant, ja que no hi havia caselles buides; pel que fa a les respostes múltiples, la constitució de la matriu de comparació a partir dels informants, i no de les poblacions, permetia tenir en compte la possibilitat de respostes diferents dins d'una mateixa varietat; d'aquesta manera se simplificava molt la possibilitat de comptar amb aquest tipus de resposta, que implica una gran complexitat en el moment d'establir la mesura de similitud. Tenint en compte això, la mesura de similitud utilitzada és força simple i es basa en les coincidències, en relació amb les diferents variables, entre parells d'informants. Es pot definir de la manera següent:

(9)

$$s(i, j) = \sum_k \text{coin}_k(i, j)$$

On $\text{coin}_k(i, j)$ pren el valor 1 quan, pel que fa a la variable lingüística k , coincideixen els informants i i j , i pren el valor 0 en cas contrari.

Les matrius de similitud comptabilitzen les semblances o les divergències dels diferents individus en relació amb el conjunt de variables analitzades. Defineixen, per tant, les interdistàncies lingüístiques entre les varietats estudiades; i presenten l'aspecte que s'observa en la figura 2. En aquesta imatge, que representa un fragment de matriu de similituds adaptat de Clua (1999), es pot veure amb facilitat que les poblacions que presenten un nombre més alt de coincidències són Alcoi i Ontinyent; són, doncs, les poblacions amb una distància lingüística més petita. Mentre que les més allunyades lingüísticament són Alzira i Guardamar. Tanmateix, tret d'aquest tipus de consideracions entre parells de poblacions, les matrius de similituds no permeten una bona interpretació dels resultats ni una visualització

plausible de les interdistàncies, ni molt menys una classificació de les varietats a partir de la distància lingüística observada. Per aquesta raó és del tot necessari cercar una representació de l'estructura d'interdistàncies definida per la matriu que permeti veure amb claredat les relacions de proximitat o la llunyania entre les varietats analitzades. Es tracta d'aconseguir una representació en un espai de dimensions reduïdes (com per exemple, el pla) que, amb la mínima distorsió possible de l'estructura d'interdistàncies original (que és multidimensional), permeti una interpretació adequada de les dades.

	Guardamar	Elx	Novelda	Alacant	la Vila	Alcoi	Cocentaina	Orinyent	Dénia	Xàtiva	Gandia	Sueca	Alzira	València
Guardamar	0,00													
Elx	2123	0,00												
Novelda	2124	2233,17	0,00											
Alacant	2050	2156,42	2189,56	0,00										
la Vila	2054	2093,53	2149,57	2130,54	0,00									
Alcoi	2028	2028,13	2102,14	2090,47	2235,47	0,00								
Cocentaina	2053	2033	2107	2140	2256	2236	0							
Orinyent	2020	2051	2115	2140	2268	2284 2258	2258	0						
Dénia	2035	2057	2119	2174	2277	2245	2258	2273	0					
Xàtiva	1957	2045	2027	2074	2180	2180	2148	2199	2210	0				
Gandia	1907	1947	1974	1983	2120	2137	2108	2146	2152	2181	0			
Sueca	1982	1983	2022	2016	2161	2194	2160	2201	2202	2195	2166	0		
Alzira	1541 1541	1559	1592	1650	1711	1717	1701	1743	1718	1802	1874	1787	0	
València	1583	1576	1579	1650	1711	1710	1696 1696	1742	1739	1769 1769	1867	1811	2213	0

Figura 2.

4.3. Representació de la distància lingüística

Amb la finalitat de representar la distància lingüística, hom pot recórrer una vegada més a l'anàlisi de dades, ja que un dels objectius d'aquesta

disciplina se centra en la recerca, la representació i la interpretació de models que reflecteixin les analogies i diferències entre individus, poblacions o grups de poblacions. En aquest àmbit, hom sol agrupar les possibles representacions en dues classes segons el tipus d'espai geomètric utilitzat: representacions en models espacials o continus i representacions en models de xarxa o discrets. Amb els models continus o espacials es pretén visualitzar l'estructura d'interdistàncies entre les poblacions o els individus, i per això es representen aquests com a punts d'un espai —el pla—, de manera que la distància entre els punts reflecteixi la distància entre els individus o les poblacions representades. Entre les tècniques que utilitzen aquest model de representació geomètrica destaca el *Multidimensional Scaling*.

Quant als models discrets o de xarxa, la seva finalitat és posar de manifest les estructures de grups, i per això es representen els individus o les poblacions com a punts d'un espai en què tenim definits uns camins —xarxes— que permeten comunicar-los. Les longituds dels camins que uneixen les poblacions reflecteixen les distàncies entre les poblacions representades. Entre els models de xarxa, destaquen els arbres ultramètrics i els arbres additius.

Atès que un dels objectius bàsics d'aquest treball era determinar les possibles agrupacions entre varietats lingüístiques, a l'hora de representar les interdistàncies entre les varietats estudiades es va decidir utilitzar models discrets o de xarxa; concretament, es va optar per una representació mitjançant arbres ultramètrics (també anomenades “representacions dendrogràfiques”), que permeten una visualització òptima de l'estructura de grups. Aquest tipus de representació té com a finalitat la construcció de

grups, basant-se en les relacions de proximitat/llunyania observades entre els diferents elements a partir d'una mesura de similitud adequada.

Així doncs, a partir de les matrius de similituds i mitjançant l'algorisme de classificació, s'estableix una distància ultramètrica i una jerarquia indexada dels objectes que defineixen la matriu; finalment es representa en forma de gràfic —dendrograma— aquesta distància ultramètrica. L'element principal en aquest procés és l'algorisme de classificació, que permet dur a terme la necessària deformació de la dissimilitud inicial —la de la matriu— en una distància ultramètrica,⁶ a partir de la qual obtenim la jerarquia indexada dels diferents elements de la matriu.

La classificació duta a terme en aquest treball es basa en el mètode UPGMA (*Unweighted Pair-Group Method using Arithmetic Averages*), definit per Sokal & Michener (1958).⁷ Es tracta d'un mètode utilitzat àmpliament en les aplicacions pràctiques de la taxonomia numèrica en diverses disciplines, i, a més a més, sota certes condicions de regularitat optimitza el coeficient de correlació cofenètica, que és la mesura d'ajust entre la dissimilaritat inicial i la ultramètrica; posteriorment tornarem sobre aquest aspecte.

Pel que fa a la qualitat de la representació obtinguda en relació amb el grau de fidelitat del model ultramètric a les dades originals, en general es considera el coeficient de correlació cofenètica, mesura d'ajust entre les similituds de partida i les ultramètriques, que es caracteritza per: a) el coeficient de correlació cofenètica pren valors entre -1 i 1 ; b) el grau de distorsió del model considerat respecte de les dades originals es quantifica per la proximitat a 0 del coeficient de correlació. En aquest sentit, si el

⁶ Sobre les característiques de la distància ultramètrica vegeu Arcas & Cuadras (1987).

⁷ Per a una discussió detallada d'aquest tipus d'anàlisi de grups, vegeu Sneath & Sokal (1973: 228-240).

coeficient de correlació és molt pròxim a -1 o a 1 , el model considerat reflecteix bé l'estructura d'interdistàncies original; i si el coeficient de correlació és pròxim a 0 , llavors el model considerat distorsiona molt les dades originals. En aquest cas els coeficients de correlació cofenètica obtinguts en els diferents gràfics han sortit suficientment pròxims a 1 per assegurar que els models jeràrquics obtinguts reflecteixen amb fidelitat l'estructura d'interdistàncies original.

Tot i això, les representacions mitjançant arbres ultramètrics presenten algunes restriccions que cal tenir en compte. D'una banda, la distància ultramètrica entre elements de dues classes diferents sempre ha de ser la mateixa, encara que en la dissimilitud inicial no sigui així. Aquesta és una de les conseqüències de la deformació produïda en la dissimilitud inicial per l'algorisme de classificació, en convertir-la en ultramètrica. De l'altra, la distància entre elements d'una mateixa classe no pot superar la distància entre elements d'aquesta i elements de qualsevol altra classe. Ateses aquestes restriccions dels models ultramètrics, s'han tingut en compte també representacions geomètriques mitjançant arbres additius, amb la finalitat de comparar ambdues representacions. S'han utilitzat, concretament, models additius basats en l'algorisme de Tversky (1977). Les representacions mitjançant arbres additius no estan subjectes a les restriccions anteriors, la qual cosa proporciona una major flexibilitat de l'arbre i, per tant, una major fidelitat a l'estructura d'interdistàncies inicial, que permet incrementar el coeficient de correlació. En contrapartida, amb aquest tipus de representació es perd parcialment l'estructura de grups —vegeu, però, Salicrú & Arcas (1997). Els resultats d'aquestes representacions han servit per contrastar la validesa de les representacions ultramètriques.

El resultat del procés és un conjunt de representacions dendrogràfiques que forneixen la següent classificació de les varietats valencianes:

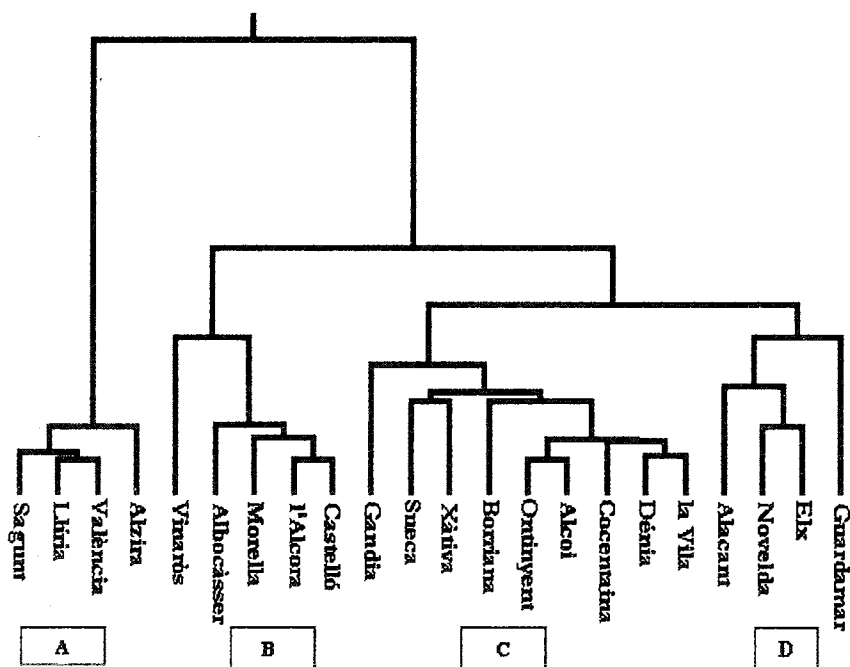


Figura 3.

Aquesta representació posa de manifest l'existència de quatre grups dialectals —separats per les línies discontinües— amb diversos graus de cohesió interna. En primer lloc, a l'esquerra del dendrograma, s'hi troben agrupades les varietats apitxades —grup A—, amb València, Llíria i Sagunt fortament cohesionades —és a dir, separades per una distància lingüística molt petita— i Alzira, que s'hi aplega a continuació, per tant, a una major distància lingüística.

A continuació, el grup B està constituït per les varietats més septentrionals; es tracta de totes les varietats castellonenques, tret de

Borriana. En aquest grup el primer conglomerat està integrat per les varietats de Castelló i l'Alcora, que presenten una distància lingüística mínima, semblant a la s'observava entre València i Llíria; tot seguit s'hi adjunten Morella i Albocàsser, a distàncies relativament petites. I, finalment, acaba incorporant-se en aquest grup la varietat de Vinaròs; en aquest cas, però, a un nivell de llunyania molt superior, la qual cosa trenca una mica la cohesió del grup.

El grup C, que és el més nombrós, està format per les varietats de València —tret de les apitxades—, és a dir: Sueca, Xàtiva, Gandia i Ontinyent; per les varietats del nord d'Alacant: la Vila Joiosa, Dénia, Alcoi i Cocentaina; i per Borriana.

Finalment, el grup D està format per les varietats alacantines situades al sud de la línia Biar-Busot: Alacant, Elx, Novelda i Guardamar. Aquest grup és el que exhibeix una major distància lingüística entre els seus integrants. En aquest sentit destaca la varietat de Guardamar com la més allunyada de totes.

Aquesta agrupació dialectal coincideix amb la delimitació del valencià apitxat de Sanchis Guarnier (1936). També està d'acord amb Colomina (1985) en la caracterització com a grup subdialectal de les varietats alacantines del sud de la línia Biar-Busot. En canvi, pel que fa a la resta de varietats, comporta diferències significatives amb les classificacions existents. D'una banda, el que aquí anomenem valencià septentrional no abasta totes les varietats castellonenques, ja que Borriana s'agrupa amb força claredat amb les varietats centrals. De l'altra, la nostra classificació implica definir un subdialecte, que anomenem central, integrat per les varietats compreses entre l'espai geogràfic delimitat per Borriana i la Vila Joiosa —amb l'excepció, de les varietats apitxades. Com que s'ha fet a

partir de l'anàlisi de les varietats dels caps de comarca del país valencià, tot i que no és geogràficament exhaustiva, es fonamenta en una perspectiva global de la variació geogràfica dels parlars valencians, la qual cosa li confereix rellevància en contraposició a altres delimitacions de varietats dialectals aïllades, fetes bàsicament amb una perspectiva més local.

Evidentment, es tracta d'una classificació centrada en les característiques de la morfologia flexiva; cosa que permet introduir consideracions qualitatives en el tractament quantitatiu. Alhora, però, aquest fet comporta que l'assumpció d'aquests resultats en termes de la globalitat dels sistemes lingüístics valencians s'hagi de fer amb certes reserves, ja que la consideració d'altres estrats lingüístics —com el sintàctic, per exemple— podria comportar modificacions en la classificació establerta. En aquest sentit, és remarcable que la consideració de les varietats alacantines del sud de la línia Biar-Busot com a grup dialectal diferenciat de les varietats veïnes del nord —el grup C de la figura 3— només té sentit si tenim en compte el conjunt de la morfologia flexiva, ja que, si considerem aïlladament la flexió verbal, la distància lingüística existent entre ambdós grups no permet establir-hi cap mena de divisió.

5. Cloenda

La dialectometria, resultat del contacte interdisciplinari entre la geolingüística i l'anàlisi de dades, ha esdevingut una eina cada cop més afinada per a l'anàlisi i el tractament de la variació lingüística. A partir del concepte de distància lingüística, les aplicacions dialectomètriques permeten determinar agrupacions de varietats lingüístiques o palesar

estructures i tendències en les xarxes dialectals. No hem de oblidar, però, que les aplicacions estadístiques (per molt que siguin cosa d'estat), més que varetes màgiques que ens permetin esbrinar l'estructura profunda de la variació, són eines argumentatives, això sí, fonamentades amb principis matemàtics contrastats, que el lingüista pot utilitzar amb finalitats heurístiques.

El desenvolupament de la disciplina i la proliferació d'aplicacions que han tingut lloc arreu en els darrers anys van lligats estretament a la possibilitat de disposar d'eines electròniques que faciliten el tractament de grans quantitats de dades i la realització de càlculs extensos. Ha estat, a més, un desenvolupament que s'ha produït quasi en paral·lel al de les aplicacions estadístiques en altres camps científics. En l'àmbit català, però, després d'unes temptatives molt primerenques i força contestades, els estudis dialectomètrics han estat pocs i esparsos. Tanmateix, aquesta situació sembla que hagi començat a redreçar-se en els darrers anys, si més no en l'esfera de les recerques que giren al voltant del COD. Esperem que aquesta línia es consolidi i que ben aviat puguem tenir aplicacions dialectomètriques a les diferents varietats del català; i per què, no? una aplicació a tot l'àmbit lingüístic, que ens doni una visió de les agrupacions dialectals del català a partir del tractament global de les dades del COD.

Referències bibliogràfiques

- ARCAS, A. & C. M. CUADRAS (1987): «Métodos geométricos de representación mediante modelos en árbol», *Publicacions de Bioestadística i Biomatemàtica*, 20.
- CLUA, E. (1999): *Variació i distància lingüística. Classificació dialectal del valencià a partir de la morfologia flexiva*, tesi doctoral inèdita, Universitat de Barcelona.

-
- _____ (2000): «Distància lingüística i classificació de varietats dialectals», *Caplletra*, 26, 11-26.
- CLUA, E., LLORET, M-R. & PEREA, M-P.(en premsa): «How to exploit a corpus: The experience of the Catalan *Corpus Oral Dialectal*», dins Agris Timuška (ed.), *Proceedings of the 4th International Congress of Dialectologists and Geolinguists. Riga, 2003*.
- COLOMINA, J. (1985b): *L'alacantí. Un estudi sobre la variació lingüística*, Alacant: Diputació provincial d'Alacant.
- COSTA, J. (1977): «Aproximació lingüística al català de Cerdanya», dins *Actes del 8è centenari de la fundació de Puigcerdà*, Puigcerdà.
- GOEBL, H (1981) «Éléments d'analyse dialectométrique (avec application à l' AIS)», *Revue de linguistique romane*, 45, 349-420.
- _____ (1987) «Points chauds de l'analyse dialectométrique: pondération et visualisation», *Revue de linguistique romane*, 51, 63-118.
- _____ (1989) «Problèmes et méthodes de la dialectométrie», dins Schouten, M.E.H. & P. Th. van REENEN (eds.), *New Methods in Dialectology. Proceedings of a workshop held at the Free University, Amsterdam, December 7-10, 1987*, Dordrecht: Foris Publications, 165-184.
- _____ (1992) «Problèmes et méthodes de la dialectométrie actuelle (avec application à l' AIS)», dins *Nazioarteko Dialektologia Biltzarra Euskaltzaindia. Bilbo 1991. X. 21/25*, 429-475.
- _____ (1993) «Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data», dins Köhler, R. & B.B. Rieger (eds.), *Contributions to Quantitative Linguistics*. Dordrecht, Boston, London: Kluwer, 277-315.
- _____ (1997) «Some Dendrographic Classifications of the Data of CLAE 1 and CLAE 2», dins *The Computer Developed Linguistic Atlas of England 2 (1997)* Tübingen: Max Niemeyer Verlag.
- GUITER, H. (1973): «Atlas et frontieres linguistiques», dins *Les dialectes romans de France à la lumière des atlas régionaux*, Paris: Centre National de Recherche Scientifique, 61-108.
- _____ (1978): «Panorama lingüístic des de Besalú», *Annals del Patronat d'Estudis Històrics d'Olot i Comarca 1978*, Olot, 35-48.

- ORTEGA, J. C. (1998): «Aplicació de tècniques de socioestadística avançada en l'anàlisi de dades dialectals: classificació dels parlars de la comarca de la Marina», *Sarrià*, 0, Callosa d'en Sarrià, La Marina.
- POLANCO, L. (1992): «Llengua i dialecte: una aplicació dialectomètrica a la llengua catalana», dins *Miscel·lània Sanchis Guarner*, III, Barcelona: Publicacions de l'Abadia de Montserrat, 5-28
- SALICRÚ, M. (1984): «Consideraciones sobre desemejanzas y clasificaciones asociadas», *Publicacions de Bioestadística i Biomatemàtica*, 14.
- _____ (1992): «Pseudocontinuidad en la clasificación jerárquica asociada al método UPGMA», *Estadística Española*, 34, 131, 383-394.
- _____ (1997): «Un algoritmo para la obtención de árboles aditivos basado en la conservación de la estructura de clusters» [manuscrit].
- SANCHIS GUARNER, M. (1936): «Extensión y vitalidad del dialecto valenciano "apitxat"», *Revista de Filología Española*, 23, 45-62.
- SARDÀ, A. & GUITER, H. (1975): «L'Atlas Lingüístic de la Catalunya i la fragmentació dialectal del català», *Miscellanea Barcinonensis*, 40, 93-112.
- SAS INSTITUTE, INC. (1990): *SAS User's Guide: Statistics. Edition 6.04*. Cary, North Carolina.
- SATTATH, S. & A. TVERSKY (1977): «Additive similarity trees», *Psychometrika*, 42, 3, 319-345.
- VENY, J. (1986): *Introducció a la dialectologia catalana*, Barcelona: Enciclopèdia Catalana.
- _____ (1992) «Fronteras y areas dialectales», dins *Nazioarteko Dialektologia Biltzarra Euskaltzaindia. Bilbo 1991. X. 21/25*, 197-245.
- VIAPLANA, J. (1994) «Dialectologia generativa en l'àmbit morfofonològic: Una perspectiva en els estudis dialectals», dins *Estudis de llengua i literatura catalanes XXIX. Miscel·lània Germà Colón*, Barcelona: Publicacions de l'Abadia de Montserrat, 311-332
- _____ (1999): *Entre la dialectologia i la lingüística: Una anàlisi dialectal aplicada al nord-occidental*, Barcelona: Publicacions de l'Abadia de Montserrat.
- VIEREK, W. (1987): «Lowman's southern English dialectal data and dialectometry», *English Wordl-Wide*, 8, 11-23.

**Presentació de l'atles sonor en el marc de l'«Atles
lingüístic del ladí dolomític i dels dialectes veïns,
primera part» (ALD-I)**

Hans Goebel
Universität Salzburg

1. Introducció

La raó que m'ha motivat a crear l'atles sonor es basa en una situació de completa absurditat. Qualsevol atles lingüístic —entès com a documentació d'una realitat parlada multiforme i, naturalment, situada sota la responsabilitat física de la boca i de l'orella— es presenta, des de fa més de cent anys, com a un mitjà documental accessible només a les mans i als ulls. Aquest estrany metabolisme sensual m'ha intrigat sempre enormement. Per aquesta raó he decidit d'obrir —sempre en la mesura de les possibilitats tècniques disponibles— una part dels materials geolingüístics que tinc al meu abast també al seu estudi sistemàtic per mitjà de l'orella. A la fi dels anys vuitanta del segle passat les possibilitats tècniques em semblaven prou madures per a la realització d'aquesta reorientació empírica.¹

¹ Cf. sobretot la nostra relació de treball de 1990 (Bauer *et alii*).

2. Què és l'atles lingüístic del ladí (ALD)?

Voldria presentar, en primer lloc, el nostre projecte atlantogràfic ALD la primera part del qual es va publicar el 1998 mitjançant quatre volums grossos en format *in-folio* i tres volums, més petits, d'índexs (alfabètic progressiu, alfabètic regressiu i etimològic).

En els quatre volums atlantogràfics hi ha 884 mapes lingüístics. La respectiva documentació electrònica ha estat dipositada —poc temps després— en tres CD-ROM i, el 2002, en un DVD únic.²

La xarxa de l'ALD comprèn 217 localitats situades en un extens territori de 25.000 quilòmetres quadrats aproximadament. Aquest territori està situat, des del punt de vista polític, al nord-est de l'Itàlia septentrional i en la part sud-oriental de Suïssa.

Des del punt de vista geolingüístic, la nostra zona engloba, a més de l'engadinès, del ladí dolomític i del furlà occidental, també els dialectes de la Llombardia oriental, del Trentí sencer i dos terços del domini lingüístic del vénèt (venecià). Es tracta, doncs, d'un atlas de tipus interregional destinat a documentar les múltiples intercorrelacions d'una zona dotada de contactes lingüístics molt complexos.

Per motius econòmics i logístics la realització del projecte ALD ha estat concebuda en dues parts. La primera part (ALD-I) —que ha estat desenvolupada entre 1985 i 1998, data de la publicació dels set volums esmentats més amunt— està dedicada a l'estudi de la *fonètica* i de la *morfologia elemental*, mentre que les preguntes del qüestionari de la segona part de l'ALD (ALD-II) se centren sobretot en el *lèxic*, en la *morfologia elaborada* i també en la *sintaxi*. Actualment, la segona part es troba en fase

² El DVD s'ha incorporat en forma d'annex en aquest volum.

de realització.³ Confio que reeixeré a acabar feliçment l'ALD-II abans del meu emeritatge a la fi del any 2012.

3. Les etapes de realització de l'atles sonor

Quan, a finals dels anys vuitanta del segle passat, es perfilava lentament la possibilitat d'un matrimoni entre les noves possibilitats tècniques i la nostra idea d'una documentació atlantogràfica accessible també a l'oïda, vam cercar la cooperació d'una empresa informàtica alemanya (RST, a Essen), dirigida per un professor de lingüística computacional (Reinhard Köhler).⁴ Aquesta empresa ha desenvolupat —abans del període d'ús del programa *Sound wave*, avui àmpliament disponible— un dispositiu electrònic per a la segmentació del so i per a la seva integració consecutiva en un banc de dades. Tot això era, des del punt de vista actual, encara molt pioner i elemental.

Ja llavors era evident que la massa enorme de treball necessari per a la segmentació correcta del so permetia tan sols el tractament d'una part molt reduïda dels materials integrals que teníem a la nostra disposició. Així nasqué la decisió de considerar, per a l'elaboració de l'atles sonor, només els dialectes de la zona central de l'ALD; és a dir, les 21 localitats de la Ladínia dolomítica.

El primer producte de la nostra col·laboració amb l'empresa alemanya citada anteriorment era un CD-àudio normal,⁵ que incloïa 98 localitats i només tres preguntes (*la catena, egli chiama, la chiesa*: “la cadena, ell

³ Cf. les nostres relacions de treball de 2002 i 2004.

⁴ Vegeu, per a una presentació detallada, Bauer (1996a i 1996b).

⁵ Cf. Bauer (1991).

crida, l'esglèsia") i que permetia, en l'àmbit de la tecnologia del disc compacte clàssic, l'audició, de moltes maneres diferents, de 294 (= 98 per 3) fragments auditius, tant en la forma d'una "via auditiva" coherent com mitjançant consultes fòniques individuals.⁶ És clar que paral·lelament a la nova disponibilitat individualitzada del so, es va tornar a proposar —tot seguit i ineluctablement— l'inevitable problema de la correcció de les transcripcions fonètiques respectives.

El procés ulterior dels nostres treballs ha estat, però, determinat no tant per problemes tècnics sinó per un espinós problema audiopsicològic. Es tracta que els nostres enregistraments, que es van fer segons els estàndards normals d'aquell període (és a dir, mitjançant magnetòfons de casset, cassetts magnètics C-90 i micròfons dotats de només una discreta característica direccional) contenien, malgrat totes les nostres precaucions, un percentatge massa elevat de sorolls paràsits. En l'audició comparativa i reiterada de petits fragments auditius aquest fenomen destorbava molt pel que fa a la precisió de la percepció auditiva. Efectivament, l'orella és un òrgan molt delicat. D'aquesta manera, esdevingué clar que un veritable atlas sonor només es podia fer amb un material sonor d'elevada qualitat tècnica i perceptiva. Per aquesta raó vam decidir, el 1993, d'explorar, una altra vegada, totes les 21 localitats de la Ladínia dolomítica amb l'ajut d'un material audioelectrònic d'alta qualitat: un excel·lent micròfon direccional per a la captació del so i les cintes DAT (*Digital Audio Tape*) per a l'enregistrament digitalitzat del so.

Vam encomanar les noves exploracions a la nostra millor entrevistadora de l'ALD-I,⁷ la tasca de la qual era dirigir les entrevistes en

⁶ Aquest CD-àudio normal es va adjuntar a la publicació de Bauer *et alii* (1990), com el DVD està annexat a aquesta contribució.

⁷ Es tracta de Helga Böhmer de Bruck/Mur (Estíria, Àustria).

un clima de perfecta tranquil·litat. Els resultats resultaren excel·lents: cap soroll paràsit, cap resposta feta en veu baixa, cap toc de campanes o de rellotges.

Després vam començar, amb tot aquest nou material fònic, la tasca enorme de segmentació integral d'aproximadament 19.000 fragments fònics mitjançant els nous estàndards del programa *Soundwave* i una versió renovada del banc de dades concebut, alguns anys enrere, per la nostra empresa informàtica alemanya. Els treballs de segmentació van durar tres anys i es van prolongar des del 1995 fins al 1998. La publicació de tres CD-ROM amb el resultat final va tenir lloc el 1999.

4. Demostració de la instal·lació pràctica de l'atles sonor

A continuació indico com fer la demostració pràctica de la instal·lació del DVD del nostre atles sonor mitjançant 13 imatges que es troben en forma d'annex a la fi de aquest article. Les dues primeres imatges serveixen per a la instal·lació del DVD; la resta, per a la il·lustració d'algunes possibilitats heurístiques de l'atlas sonor.

4.1. Instal·lació de l'atlas sonor

Vegeu els gràfics 1 i 2. Després de la inserció del DVD en el lector de l'ordinador⁸ i després de l'activació del DVD mitjançant el programa estàndard «Explorer» es presenta generalment la imatge del gràfic 1. Cal, doncs, procedir com s'indica en aquest gràfic.

⁸ L'ordinador ha de tenir el sistema operatiu Windows (versions 98, 2000, NT o XP).

La instal·lació pròpiament dita s'ha de realitzar segons el gràfic 2. A la fi de la instal·lació es presenta una finestra semblant a la que apareix en el gràfic 3. S'hi veu la zona de la Ladínia dolomítica i, a l'esquerra, en la part superior, una finestra bipartida que tapa l'extrem necessari per continuar la consulta de l'atlas sonor. Per aquesta raó cal procedir com s'indica en el gràfic 3.

Després es pot escollir la llengua d'ús de l'atlas sonor. Hi ha tres possibilitats: alemany, italià i francès. Proposo d'escollir l'italià (vegeu el gràfic 4).

4.2. L'audició segons les localitats

Voldria presentar a continuació una de les modalitats estàndard d'audició, és a dir, l'audició segons les localitats. Aquesta audició correspon a la consulta acústica d'un diccionari dialectal normal en ordre alfabètic (vegeu els gràfics 6 i 7).

Després d'haver activat el gràfic 7 podem procedir a la primera audició. Però atenció: la localitat parlant es presentarà sempre en groc, la localitat muda en blau. Les transcripcions respectives apareixeran en un quadre especial que es pot desplaçar fàcilment amb el ratolí. Els símbols fonètics equivalen als de l'atles italo-suís (AIS) de Karl Jaberg i Jakob Jud. Cal fer una darrera observació molt important: les transcripcions que apareixen durant les audicions han estat fetes en el curs de les primeres entrevistes a la Ladínia i no es corresponen, per aquest motiu, de manera perfecta al so que es pot escoltar. L'espai de temps transcorregut entre la redacció de les transcripcions (1985-86) i l'enregistrament del so de qualitat (que es podrà escoltar ben aviat) és de set o vuit anys.

La finestra amb les transcripcions es pot desplaçar mitjançant el ratolí perquè no amagui cap de les localitats de la xarxa integral.

Cal esmentar encara una eina de treball molt útil per al tractament de les possibles diferències entre el so de l'atles ladí *sonor* i les transcripcions de l'atles ladí *imprès*: és a dir, el que s'anomena (en italià) “trascrizzionario”. El “trascrizzionario” (que es troba disponible a internet)⁹ comprèn dues sèries de llargues llistes de transcripcions ladines de l'atles imprès i, al costat, espais en blanc. Les llistes s'ordenen, d'una banda, segons les localitats i, d'una altra, segons les preguntes. Aquesta disposició documental i la presència paral·lela dels espais en blanc permet, a l'usuari de l'atles sonor, de controlar i de corregir les nostres transcripcions segons el seu parer.

4.3. L'audició segons estímuls/respostes

Vegeu el gràfic 8. Una altra manera de fer l'audició és escoltar “secondo Stimuli” (“segons estímuls”), que correspon a la consulta acústica d'un mapa d'un atles lingüístic. La demostració s'inicia amb el primer ítem de l'atles, és a dir, *l'aceto* (“el vinagre”). Es poden sentir ara els diversos matisos de les denominacions ladines del vinagre, etc.

4.4. L'audició amb modalitats acústiques especials

Vegeu el gràfic 9. Com que l'orella és un òrgan molt delicat, cal que la programació de l'atles sonor disposi de diverses possibilitats de repetició i també d'intercalació de pauses de durada variable. Cal activar per a això el gràfic 9.

⁹ Es llegeix l'adreça electrònica següent: http://www.sbg.ac.at/rom/people/proj/ald/ald_home.htm.

4.5. L'audició segons modalitats científiques particulars

Vegeu els gràfics 10-13. Amb tot, la veritable capacitat documental de l'atles sonor es manifesta mitjançant les audicions *selectives*, *contrastives*, *reiterades* i *creuades*. Prèviament cal programar les audicions desitjades, la qual cosa es duu a terme mitjançant les llistes de les localitats (*luoghi disponibili*) i les llistes de les respostes disponibles (*stimuli disponibili*).

Per començar proposo, des d'un punt de vista geogràfic, l'audició creuada amb una repetició simple de l'estímul *capra* ("cabra") en dues localitats ladines situades en dues valls ladines veïnes (Val Badia: P. 81-La Pli; Val Gardena: P. 86-Bula). Per a això cal activar les imatges 10 i 11.

És possible també fer audicions amb finalitats científiques precises; per exemple amb un interès únicament lèxic. Es proposa, amb aquesta finalitat, utilitzar la xarxa integral i fer l'audició de les diverses denominacions ladines del concepte "il capriolo" ("el cabirol"),¹⁰ fent una repetició per a cadascuna localitat i fent pauses breus entre les respostes (vegeu el gràfic 12).

El nostre últim exemple fa referència a dos problemes fonològics clàssics en l'àmbit de la ladinística; és a dir, la palatalització del nexa llatí C + A en posició sil·làbica inicial i la realització de la A llatina accentuada en síl·laba tancada (CÁR|RU) i en síl·laba oberta (CÁ|RU). Escollim cinc localitats (en cinc valls ladines diferents) i les respostes per a *il carro* ("el carro") i per a l'adjectiu *caro* ("car"). Les respostes respectives se sentiran tres vegades i es poden presentar de dues maneres: *localitat rere localitat* (amb combinació dels estímuls respectius) o *resposta rere resposta* (amb encreuament geogràfic) —vegeu el gràfic 13.

¹⁰ La forma lèxica *rexl* és d'origen tirolés mentre que *kapriol*, per exemple, és d'origen italià.

5. Conclusió

No és possible esmentar en aquesta petita contribució totes les possibilitats heurístiques ofertes per l'atles sonor. Em limito sols a mencionar encara la lliure combinació de sintagmes o de frases sencers que, des del punt de vista gramatical, són correctes, però que no han mai estat pronunciats en la realitat; per exemple, *la fiamma rossa, il fiore verde, il cane grasso* o *voglio mangiare l'oca grassa/il gallo grasso/le castagne verdi, voleva vedere le donne nere, io sono bello – tu sei brutta*, etc.¹¹

Una aplicació molt valuosa de l'atles sonor es fa constar en la documentació general (i no tan sols lingüística) del patrimoni sonor d'una comunitat lingüística en molts àmbits extracientífics com ara en els de la pedagogia o la didàctica, de l'etnografia, de la història o de la geografia regional.

¹¹ Per a l'audició d'aquestes frases hi ha novament dues possibilitats: d'una banda, l'audició paradigmàtica amb la presentació separada dels matisos geolingüístics dels components particulars; i, d'una altra, l'audició sintagmàtica amb la presentació combinada dels matisos geolingüístics de la seqüència sencera de les frases respectives.

Bibliografia relativa a l'atles sonor i a l'ALD-II

- AIS (1928-1940): Jaberg, Karl / Jud, Jakob, *Sprach- und Sachatlas Italiens und der Südschweiz*, Zofingen (Ringier), 8 vol. (reimpressió: Nendeln, Kraus, 1971).
- ALD-I (1998-2002): Goebel, Hans / Bauer, Roland / Haimlerl, Edgar (eds.): *Atlant linguistisch dl ladin dolomitich y di dialec vejins*. 1ª parte. Atlante linguistico del ladino dolomitico e dei dialetti limitrofi. 2ª parte. Sprachatlas des Dolomitenladinischen und angrenzender Dialekte. 1. Teil, Wiesbaden (Dr. L. Reichert) 1998, 7 vol., 3 CD-ROM (Salzburg, Institut für Romanistik) 1999, 1 DVD (Salzburg, Institut für Romanistik) 2002.
- BAUER, Roland (1991): *ALD-I - CD: 98 Tonproben zum ALD-I / CD - ALD-I: 98 campioni fonici per l'ALD-I*, Salzburg / Salisburgo, Compact Disk mit zweisprachigem Booklet (dt./ital.) / disco compatto audio con booklet bilingue (ital./ted.), 44 Seiten / pp.
- _____ (1992): «L'informatizzazione dell'atlante linguistico sonoro ALD-I (Atlante linguistico del ladino centrale e dialetti limitrofi I)», dins *Linguistica* XXXII, 197-212.
- _____ (1993a): «Ein Sprach-Atlas beim Wort genommen: ALD-I, der "Sprechende"», dins Winkelmann, Otto (ed.), *Stand und Perspektiven der romanischen Sprachgeographie*, Wilhelmsfeld (Egert), (= pro lingua, 15), 283-306.
- _____ (1993b): «Neue Perspektiven der LDV im Forschungsprojekt ALD-I - Der sprechende Sprachatlas», dins Viereck, Wolfgang (ed.), *Verhandlungen des Internationalen Dialektologenkongresses Bamberg 1990*, vol. 1: Plenarvorträge, Computative Datenverarbeitung, Dialektgliederung und Dialektklassifikation, Stuttgart (Steiner), (= ZDL-Beiheft, 74), 124-146.
- _____ (1993c): «Linguistik und Multimedia. Neue Wege der Mundartforschung, aufgezeigt am Beispiel des dolomitenladinischen Sprachatlases (ALD-I)», dins *Moderne Sprachen* 37/2, 66-80.
- _____ (1994): «Il ladino letto e ascoltato», dins *Italiano & oltre* IX/2, 76-79.
- _____ (1995): «Dialettologia computazionale ed atlanti linguistici: la dimensione sonora», dins Maria Teresa Romanello / Immacolata Tempesta (eds.), *Dialetti e*

lingue nazionali, Atti del XXVII Congresso internazionale della SLI, Roma (Bulzoni), (= Pubblicazioni della SLI, 35), 155-167.

____ (1996a): «Die romanische Geolinguistik im Spannungsfeld zwischen Wirtschaft und Wissenschaft: Kooperationsmodelle im Projekt ALD-I», dins Edgar Radtke / Harald Thun (eds.), *Neue Wege der romanischen Geolinguistik. Akten der gleichnamigen Tagung*, Kiel (Westensee), (= Dialectologia Pluridimensionalis Romanica, 1), 430-444.

____ (1996b): «Le système de gestion de base de données de l'Atlas Linguistique du Ladin Central (ALD-I)», dins Georges Moracchini (ed.), *Bases de données linguistiques: conceptions, réalisations, exploitations. Actes du Colloque International*, Corte (Université de Corse), 195-211.

____ (1997): «Zur Rolle des Computers beim dolomitenladinischen Sprachatlas (ALD-I)», dins *Ladinia* XXI, 217-223.

____ (2004): «Analyse- und Visualisierungstechniken sprachgeographischer Daten am Ende des 20. Jahrhunderts», dins Wolfgang Dahmen *et al.* (eds.), *Romanistik und neue Medien. Romanistisches Kolloquium XVI*, Tübingen (Narr), (= Tübinger Beiträge zur Linguistik, 455), 189-208.

BAUER, Roland; Helga BÖHMER; Silvio GISLIMBERTI; Hans GOEBL; Reinhard KÖHLER; Martin SCHLEUSSER; Tino SZEKELY; Hans TYROLLER (1990): «Arbeitsbericht 5 zum ALD-I - Relazione di lavoro 5 per l'ALD-I», *Ladinia*, XIV, 259-304.

BAUER, Roland; Helga BÖHMER; Hans GOEBL; Edgar HAIMERL; Guillaume SCHILTZ; Hans TYROLLER (1991a): «Arbeitsbericht 6 zum ALD-I», *Ladinia*, XV, 203-254.

BAUER, Roland; Hans GOEBL (1991b): «Presentazione di ALD-I - Atlante linguistico del ladino dolomitico e dialetti limitrofi, parte prima», dins *Per Padre Frumenzio Ghetta, O.F.M., Scritti di storia e cultura ladina, trentina, tirolese e nota bibliografica in occasione del settantesimo compleanno*, Trento (Biblioteca Comunale di Trento / Istitut Cultural Ladin "Majon di Fashegn"), 73-99.

____ (1992): «Arbeitsbericht 7 zum ALD-I», *Ladinia*, XVI, 169-184.

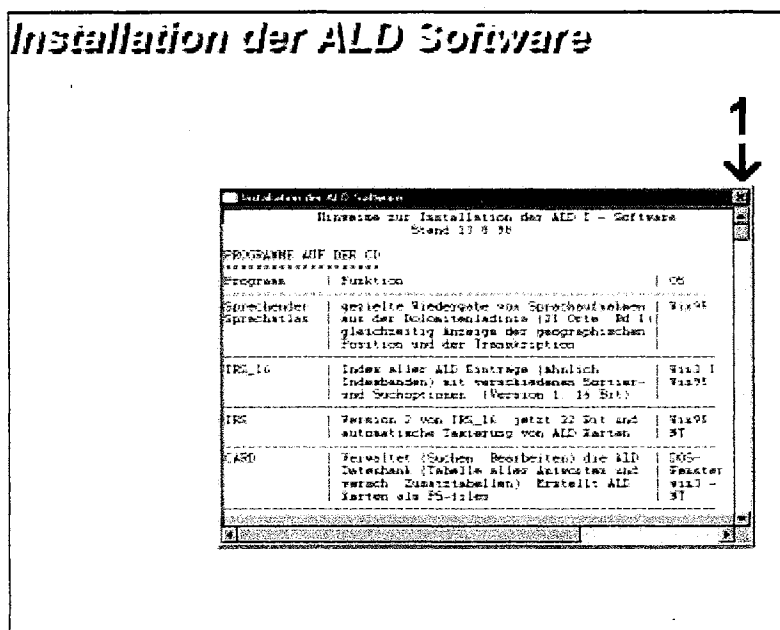
BAUER, Roland; Hans GOEBL (2000): «Utilisation nouvelle de l'informatique dans les Atlas linguistiques en Europe (1980-2000)», *Verbum* XXII/2, 169-185 («Géolinguistique en Europe»).

BAUER, Roland; Hans GOEBL; Edgar HAIMERL (1993): «Arbeitsbericht 8 zum ALD-I», *Ladinia*, XVII, 125-152.

- BAUER, Roland; Hans GOEBL; Edgar HAIMERL; Ulrike HOFMANN; Heidi PAMMINGER (1995): «Arbeitsbericht 9 zum ALD-I», *Ladinia*, XIX, 307-330.
- BAUER, Roland; Hans GOEBL; Edgar HAIMERL (1996): «Arbeitsbericht 10 zum ALD-I», *Ladinia*, XX, 191-221.
- GOEBL, Hans (1992): «L'atlas parlant dans le cadre de l'Atlas linguistique du ladin central et des dialectes limitrophes (ALD)», dins *Actes du congrès international de dialectologie / Nazioarteko dialektologia biltzarra*, Agiriak Bilbao / Bilbo [Bilbao] (Académie de la langue basque / Euskaltzaindia), 397-441.
- _____ (1994): «L'Atlas linguistique du ladin central et des dialectes limitrophes (première partie, ALD-I)», dins Garcia Mouton, Pilar (ed.), *Geolingüística. Trabajos europeos*, Madrid (Consejo Superior de Investigaciones Científicas), 155-168.
- _____ (1995): «Un nouvel atlas linguistique des Alpes centrales: l'Atlas linguistique du ladin des Dolomites (ALD-I)», dins *Estudis de lingüística i filologia oferts a Antoni M. Badia i Margarit*, volum 1, Barcelona: Departament de Filologia Catalana de la UB, Publicacions de l'Abadia de Montserrat, 295-312.
- _____ (1996/1997): «L'Atlas linguistique du ladin central et des dialectes limitrophes (première partie, ALD-I)», *Cahiers de l'Institut de Linguistique de Louvain*, 22/3-4; 23/1-2, 276-279.
- _____ (2001): «Der ALD-I am Ziel. Ein Rückblick auf die zweite Halbzeit», dins Iliescu, Maria / Plangg, Guntram A. / Videsott, Paul (eds.), *Die vielfältige Romania. Dialekt - Sprache - Überdachungssprache. Gedenkschrift für Heinrich Schmid (1921-1999)*, Vich (Istitut Cultural Ladin "majon di fascegn") / San Martin de Tor (Istitut Cultural Ladin "Micurà de Rü") / Innsbruck (Institut für Romanistik), 171-187.
- _____ (2002): «Stato attuale dei lavori all'ALD-II (2° parte dell'Atlante linguistico del ladino dolomitico e dei dialetti limitrofi)», *Bollettino dell'Atlante Linguistico Italiano*, III serie, 26, 185-186.
- _____ (2003): «Breve presentazione dell'ALD-I: Atlante Linguistico del Ladino Dolomitico e dei dialetti limitrofi. Prima parte», dins Raffaella Bombi / Fabiana Fusco (eds.), *Parallela 10. Sguardi reciproci. Vicende linguistiche e culturali dell'area italoфона e germanoфона. Atti del Decimo Incontro italo-austriaco dei linguisti*, Udine (Forum), 87-92.

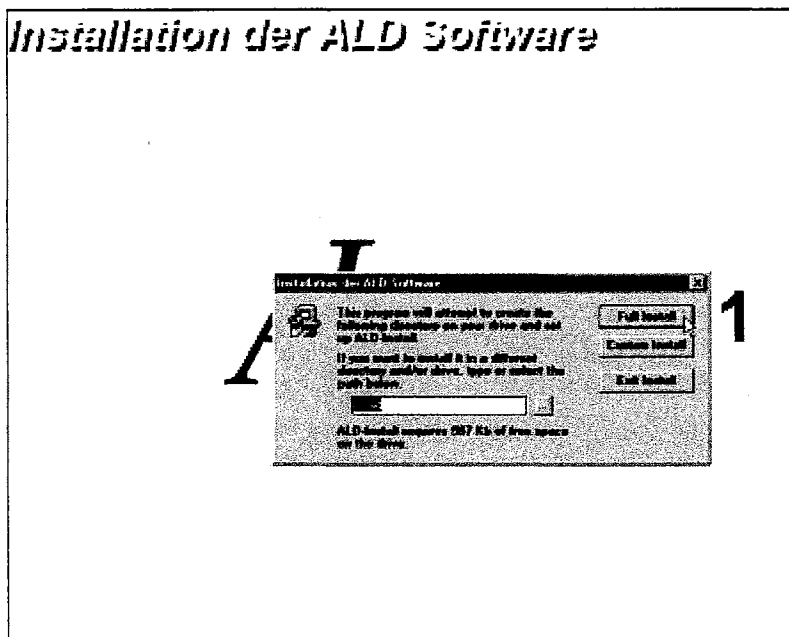
-
- _____ (2004) «Brève présentation de l'Atlas linguistique ladin (ALD: 1^{ère} et 2^{ième} partie)», *Bretagne linguistique*, 13, 375-381.
- GOEBL, Hans; Roland BAUER (1992): «L'atlante linguistico del ladino centrale e dialetti limitrofi (ALD-I): Stato attuale dei lavori», dins Ruffino, Giovanni (ed.), *Atlanti linguistici italiani e romanzi, Esperienze a confronto*, Palermo (CSFLS), 331-341.
- HAIMERL, Edgar (1997): «ALD-I - a Linguistic Atlas Published on Multiple Media», dins Thomas, Alan R. (ed.), *Issues and Methods in Dialectology*, Wales Bangor (Department of Linguistics), 200-210.
- KATTENBUSCH, Dieter (2003): «ALD-I und VIVALDI und die Segnungen der akustischen Sprachgeographie», *Quo vadis Romania?*, 22, 22-30.

Annex: INSTAL·LACIÓ DEL DVD *ALDV-I*



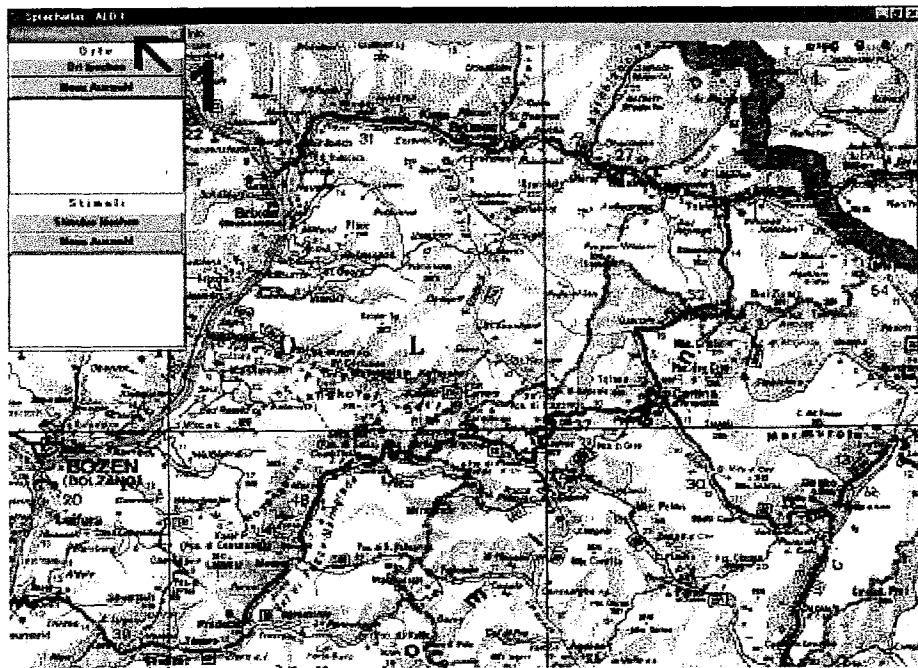
Gràfic 1: Instal·lació A

- 1 Feu un clic per eliminar la finestra oberta i per poder així continuar la instal·lació.



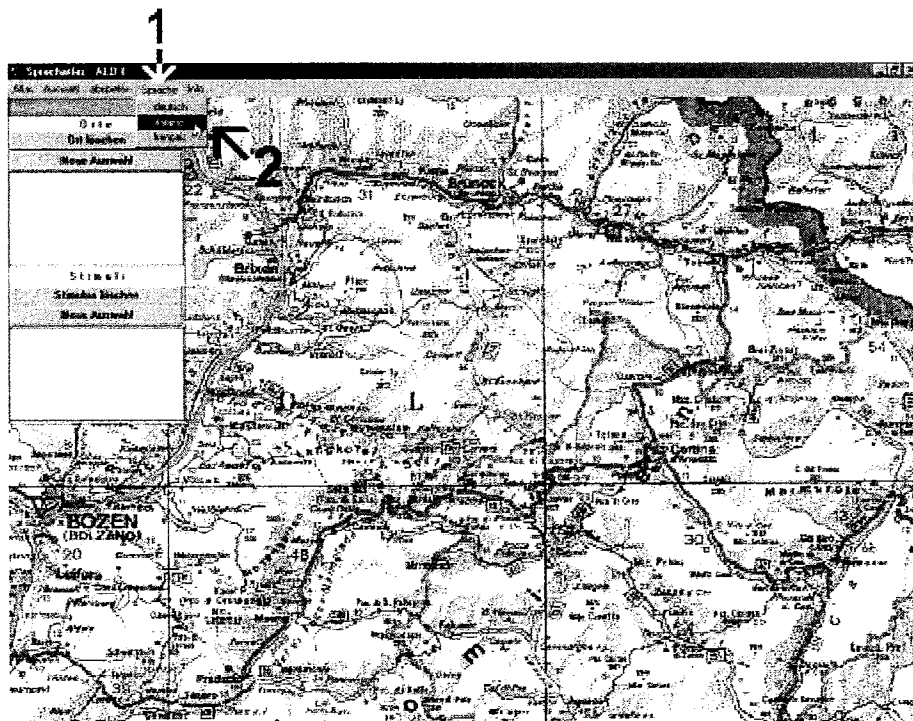
Gràfic 2: Instal·lació B

- 1 Pitgeu el botó “Full Install” per iniciar la instal·lació i per completar-la successivament fins arribar al final.



Gràfic 3: Preparació A

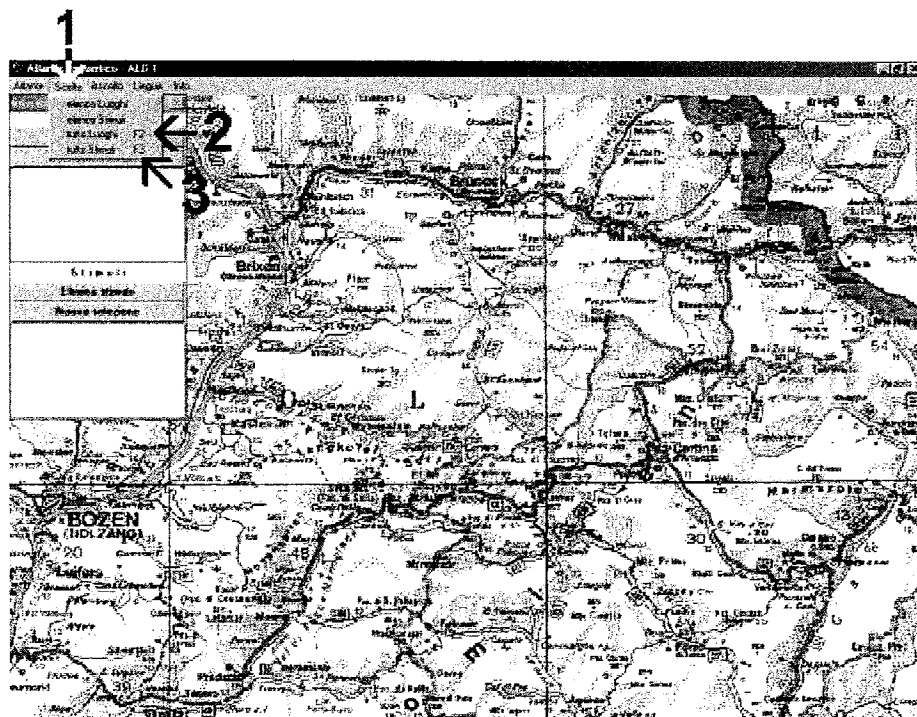
1 Abaixeu la finestra mitjançant la barra blava.



Gràfic 4: Preparació B

1 Feu clic en el botó “Sprache”.

2 Feu clic en el botó “italiano”.



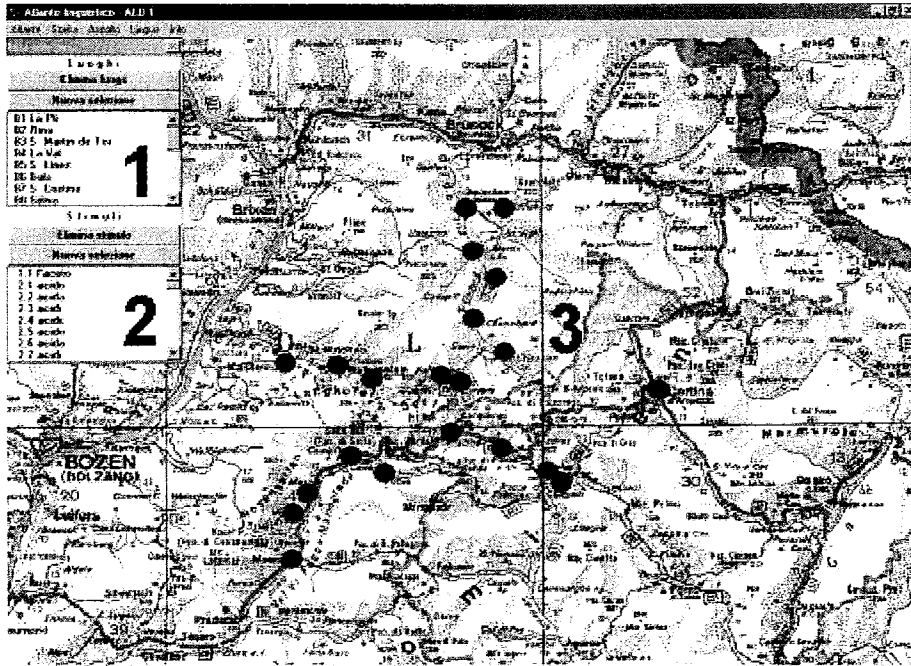
Gràfic 5: Preparació C

Abaixeu la finestra mitjançant la barra blava.

1 Feu clic en el botó “Scelta”; i després:

2 feu clic en el botó “tutto Luoghi F 2”; repetiu el punt 1;

3 feu clic en el botó “tutto Stimuli F 3”.

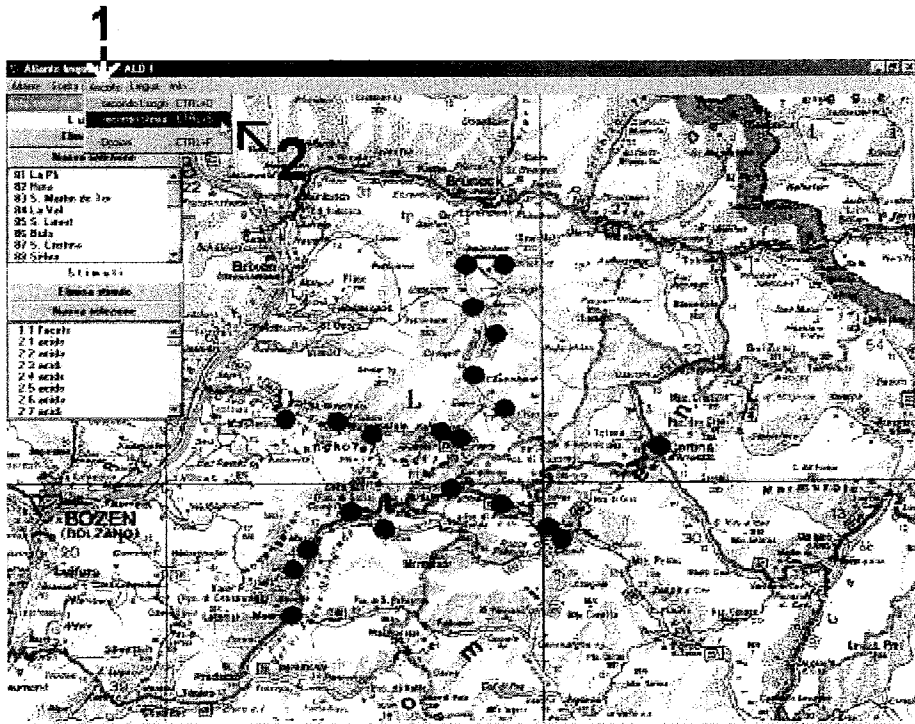


Gràfic 6: Preparació D

1 Apareix la llista de les 21 localitats de la Ladínia dolomítica, la numeració de les quals és la que s'ha publicat en l'ALD-I.

2 Apareix la llista de tots els "estímuls" (preguntes) del qüestionari original de l'ALD-I (en ordre alfabètic). La numeració fa referència només al qüestionari original (no publicat) de l'ALD-I; no correspon, per tant, al de l'ALD-I (publicat).

3 La xarxa de les 21 localitats de la Ladínia dolomítica (en blau: localitat "muda").

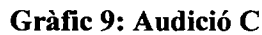


Gràfic 8: Audició B

1 Feu clic en el botó “Ascolto” (cf. els gràfics/escenaris 7 i 8, etapa 2); i després:

2 feu clic en el botó “secondo Stimuli CTRL + S”. Resultat: s’inicia l’audició d’un mapa lingüístic en l’ordre establert per a les 21 localitats. Durant l’audició, les localitats i els estímuls activats apareixen, a dins de la finestra (a l’esquerra, a dalt), sempre en blau.

Atenció: la finestra de les transcripcions es pot desplaçar lliurement!

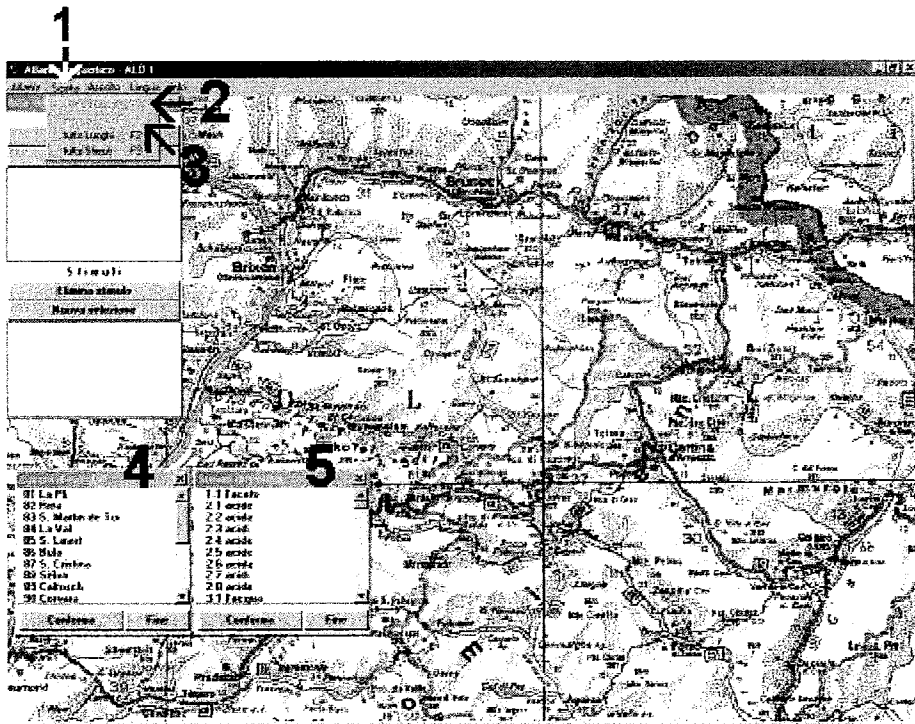


feu clic en el botó “Opzioni CTRL + P” (cf. els gràfics 7 i 8, punt 2).

1 Definiu el nombre de les repeticions.

3 Definiu la durada de la pausa entre les *preguntes* (recomanació: 500 msec).

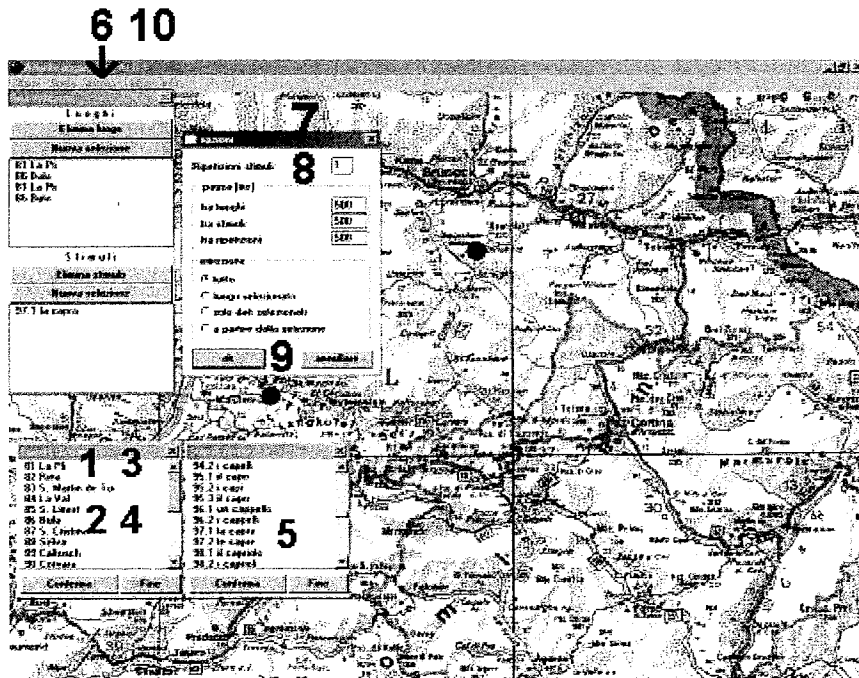
5 Feu clic en el botó “ok” per confirmar totes les seleccions efectuades.



Gràfic 10: Audició D

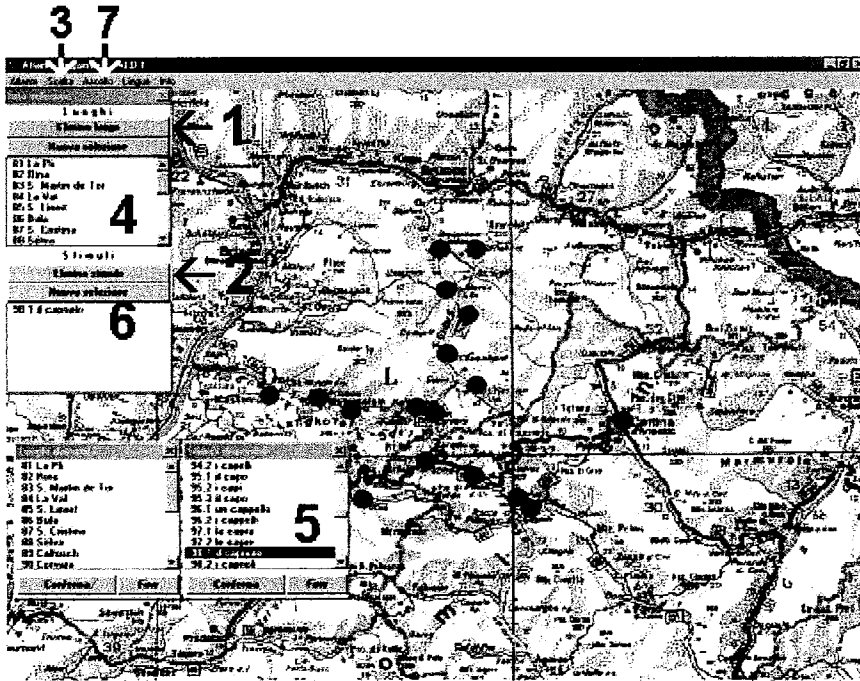
Seguiu les indicacions del gràfic 3 i creeu una finestra “de comandament” buida. (Hi ha una altra possibilitat: feu clic en els dos botons “Nuova selezione” en la finestra “de comandament”. Resultat: la finestra es buida completament; cf. el gràfic 13, punts 1 i 2.); i després:

- 1 feu clic en el botó “Scelta”.
- 2 feu clic en el botó “elenco Luoghi”. Resultat: apareix la finestra 4 (“Luoghi disponibili”).
- 3 feu clic en el botó “elenco Stimuli”. Resultat: apareix la finestra 5 (“Stimuli disponibili”).



Gràfic 11: Audició E

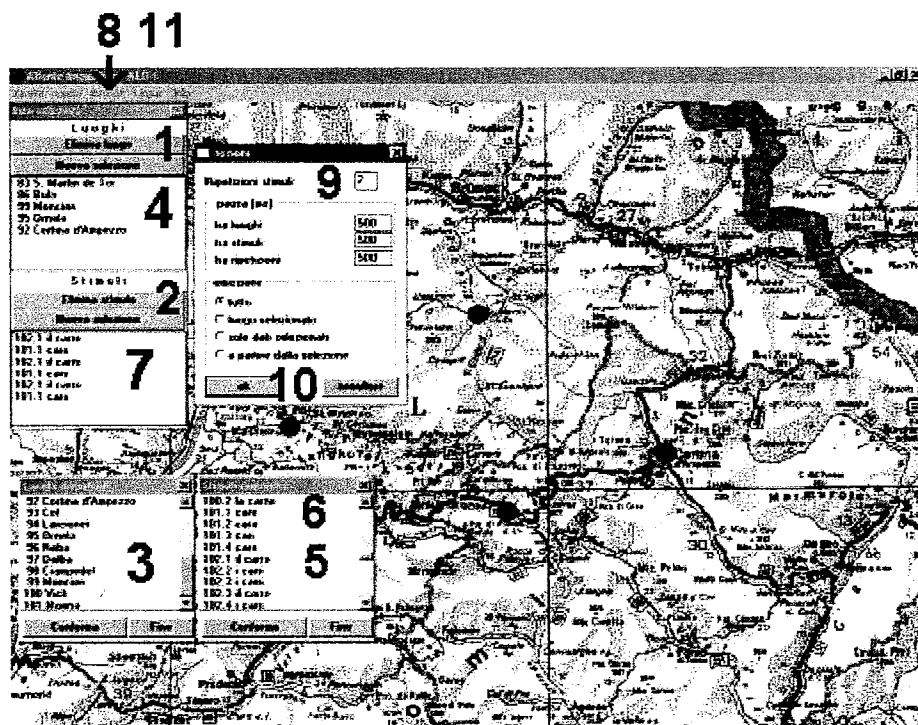
- 1 Feu clic dues vegades a “81 La Pl”.
 - 2 Feu clic dues vegades a “86 Bula”.
 - 3 Repetiu el punt 1.
 - 4 Repetiu el punt 2.
 - 5 Feu clic dues vegades en “97.1 la capra”.
 - 6 Feu clic (una vegada) en el botó “Opzioni”.
 - 7 Resultat: la finestra “Opzioni” s’obre (cf. el gràfic 9).
 - 8 Definiu una sola repetició (1).
 - 9 Feu clic en el botó “OK” per confirmar aquesta definició.
 - 10 Feu clic en el botó “Ascolto”. I després: feu clic en els botons “secondo Luoghi” o “secondo Stimuli”. Resultat: l’audició de les diverses denominacions ladines del mot “cabra” s’inicia.
- Atenció: Totes les finestres (de les transcripcions, per exemple) es poden desplaçar lliurement per deixar al descobert les localitats seleccionades!



Gràfic 12: Audició F

- 1 Feu clic una vegada en el botó “Nuova selezione”. Resultat: la meitat superior de la finestra es buida.
- 2 Feu clic una vegada en el botó “Nuova selezione”. Resultat: la meitat inferior de la finestra es buida.
- 3 Feu clic una vegada en el botó “Scelta”. I després: feu clic una vegada en el botó “tutto Luoghi F 2”.
- 4 Resultat: a) la meitat superior de la finestra s’omple amb els noms de les 21 localitats, b) les 21 localitats de la xarxa apareixen en blau.
- 5 Feu clic dues vegades en “98.1 il capriolo”.
- 6 Resultat: en la meitat inferior de la finestra es presenta l’estímul “98.1 il capriolo”.
- 8 Feu clic una vegada en el botó “Ascolto”. I després: feu clic una vegada en el botó “secondo Luoghi CTRL + 0”. Resultat: l’audició s’inicia segons la numeració establerta en les 21 localitats.

Atenció: La finestra de les transcripcions es pot desplaçar lliurement!



Gràfic 13: Audició G

- 1 Feu clic una vegada en el botó “Nuova selezione”. Resultat: la meitat superior de la finestra es buida.
- 2 Feu clic una vegada en el botó “Nuova selezione”. Resultat: la meitat inferior de la finestra es buida.
- 3 Feu clic (dues vegades) successivament en les localitats següents: 83 S. Martin de Tor, 86 Bula, 99 Moncion, 95 Ornella, 92 Cortina d’Ampezzo (abaixeu, si fos necessari, la llista de les localitats mitjançant la barra de la dreta).
- 4 Resultat: la meitat superior de la finestra s’omple amb els noms de aquestes localitats.

-
- 5 Feu clic dues vegades en “102.1 il carro”.
 - 6 Feu clic dues vegades en “101.1 caro”; i després: repetiu successivament els punts 5 i 6 dues vegades.
 - 7 Resultat: la meitat inferior de la finestra s’omple amb tres combinacions dels estímuls “102.1 il carro” i “101.1 caro”.
 - 8 Feu clic una vegada en el botó “Ascolto”. I després: feu clic una vegada el botó “Opzioni CTRL + P”.
 - 9 Resultat: la finestra “Opzioni” s’obre. I després: definiu el nombre de les repeticions: 0.
 - 10 Feu clic en el botó “OK” per confirmar aquesta selecció.
 - 11 Feu clic una vegada en el botó “Ascolto”. I després:
 - a) per a una audició [sintagmàtica] *localitat rere localitat* (d’interès més aviat *fonològic*): feu clic una vegada en el botó “secondo Luoghi CTRL + 0”. Resultat: l’audició s’inicia.
 - b) per a una audició [paradigmàtica] *resposta rere risposta* (d’interès més aviat *fonètic*): feu clic una vegada en el botó “secondo Stimuli CTRL + S”. Resultat: l’audició s’inicia.
- Atenció: Totes les finestres es poden desplaçar lliurement per deixar al descobert la xarxa seleccionada!

L'«Arxiu audiovisual dels dialectes catalans de les Illes Balears»

Jaume Corbera Pou

Universitat de les Illes Balears

L'any 1996 el projecte titulat *Arxiu audiovisual dels dialectes catalans insulars* va obtenir l'ajuda a la recerca “Aula Francesc de B. Moll” de l'Estudi General Lul·lià de Mallorca. Va ser la fita inicial que ens va permetre als promotors, Brauli Montoya Abat i qui firma aquestes línies, Jaume Corbera Pou, com a director, aleshores tots dos professors titulars de Filologia Catalana de la Universitat de les Illes Balears, fer efectiva la idea que arrossegàvem des de feia uns anys: fer un arxiu en vídeo dels parlars tradicionals de les illes Balears, amb entrevistes a persones que mantenguessin encara amb tota l'autenticitat la manera de parlar característica forjada pels set segles d'història lingüística catalana de les illes. La intenció era aplegar una col·lecció documental bastant extensa perquè fos una bona mostra d'allò que la dialectologia catalana catalogava com a “dialecte balear”, en què hi intervenguessin parlants majoritàriament d'edat prou avançada com poder-los considerar representants de la llengua tradicional, la llengua mantenidora dels trets històrics propis dels parlars de les illes, amb les interferències limitades només al resultat d'uns segles de

domini cultural de la llengua castellana, la qual, però, no havia estat gens usual dins els sectors populars dels habitants illencs, encara unilingües fins a les generacions nades a principis del segle XX. L'arxiu no volia ser sociolingüístic, sinó geolingüístic, per la qual cosa havíem de cercar entrevistats que conservassin bé el parlar del lloc representat, i aquests, avui en dia, són sobretot persones per damunt els 60 anys, per no dir 70; així i tot, n'hi podia haver qualcuna de més jove, considerant que a vegades trobes gent jove que, gràcies a circumstàncies diverses, encara conserven aquesta llengua autèntica no contaminada. La idea de fer un arxiu en vídeo, i no en casset com s'havien fet fins al moment les arreplegues documentals lingüístiques més modernes, era per poder guardar també, a més a més del so, les expressions i els gests corporals (amb les mans, amb la cara, de vegades amb tot el cos), i les imatges dels objectes o les maneres a les quals els informadors poguessin fer referència. Aixímateix, volíem que els enregistraments fossin de la millor qualitat possible (possible segons els nostres recursos, és clar), perquè fossin ben aprofitables per a la recerca científica; que no fos un material purament testimonial, sinó també pràctic i útil, amb el qual els investigadors que volguessin tenir una mostra del llenguatge de les Balears per poder analitzar hi poguessin comptar amb garantia.

Amb aqueixa ajuda de l'Estudi General Lul·lià, doncs, vàrem començar el projecte, que va poder tenir continuïtat gràcies, primer, al fet que la Direcció General de Cultura i Política Lingüística del Govern Balear va doblar l'import inicial que aquella institució acadèmica ens havia concedit, i llavors a haver estat un dels projectes inclosos al Programa Mossèn Alcover que el Consorci per al Foment de l'Ús del Català va impulsar el 1998.

El pla del projecte era fer entrevistes a grups de persones, amigues o familiars perquè hi hagués confiança entre elles, que establissin una conversa guiada per nosaltres, que els estiràriem cap a uns centres d'atenció, de manera

que totes les entrevistes serien més o menys semblants. Prevèiem també la possibilitat de fer entrevistes a persones soles, les quals serien interessants per la formació cultural especial que tendrien, la qual podria suposar uns usos lingüístics també especials. Ens constituïrem com a equip entrevistador els següents: Joan-Francesc López Casasnovas a Menorca, Marià Torres Torres a Eivissa i Formentera, i Brauli Montoya Abat i Jaume Corbera Pou, amb la funció d'investigador principal, director o responsable —com li vulgueu dir—, a Mallorca. El 2000, però, Brauli Montoya es va incorporar com a catedràtic a la Universitat d'Alacant i va haver de deixar el projecte; així que a Mallorca hi vaig quedar solament el sotassignat. Les localitats on havíem de fer les entrevistes, davant la impossibilitat que fossin totes, varen ser triades segons els següents criteris: *a)* totes les capitals insulars; *b)* totes les ciutats de més de 20.000 habitants autòctons (és a dir, no comptant-hi els immigrants, majoritàriament castellanoparlants); *c)* totes les localitats o àrees geogràfiques amb característiques dialectals específiques i singulars respecte al parlar general de l'illa. D'acord amb aquests criteris, s'han fet entrevistes a les poblacions o àrees següents marcades al mapa adjunt (un mínim de 2 entrevistes a cada una) —vegeu figura 1.

Al final (ja que el pressupost ens bastava) vaig decidir que també hi inclouríem la vila valenciana de Tàrbena, a la Marina, els habitants de la qual són d'ascendència mallorquina. Podia ser interessant tenir-ne una mostra lingüística per contrastar amb el mallorquí "autèntic", i va ser l'oportunitat perquè Brauli Montoya tornàs col·laborar en el projecte.

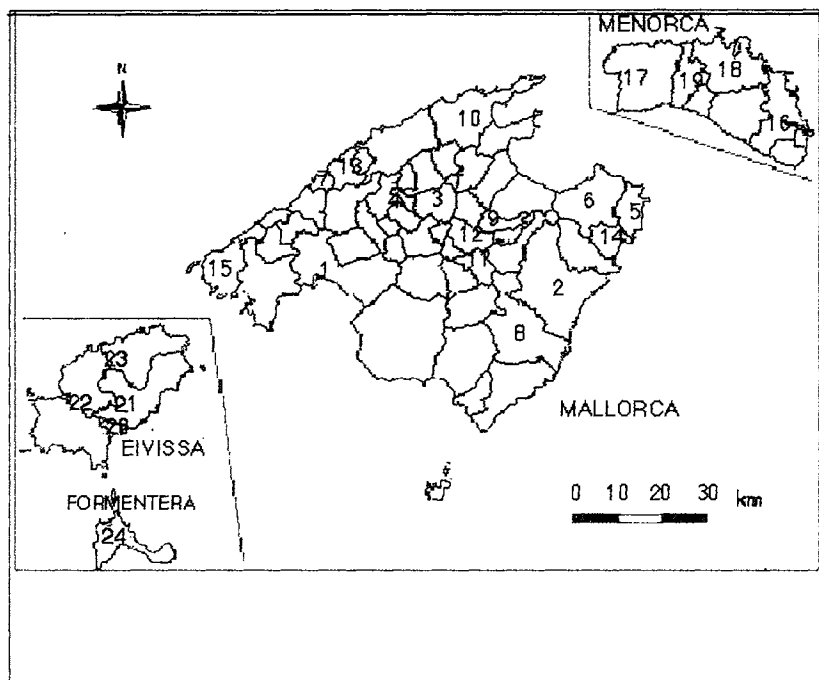
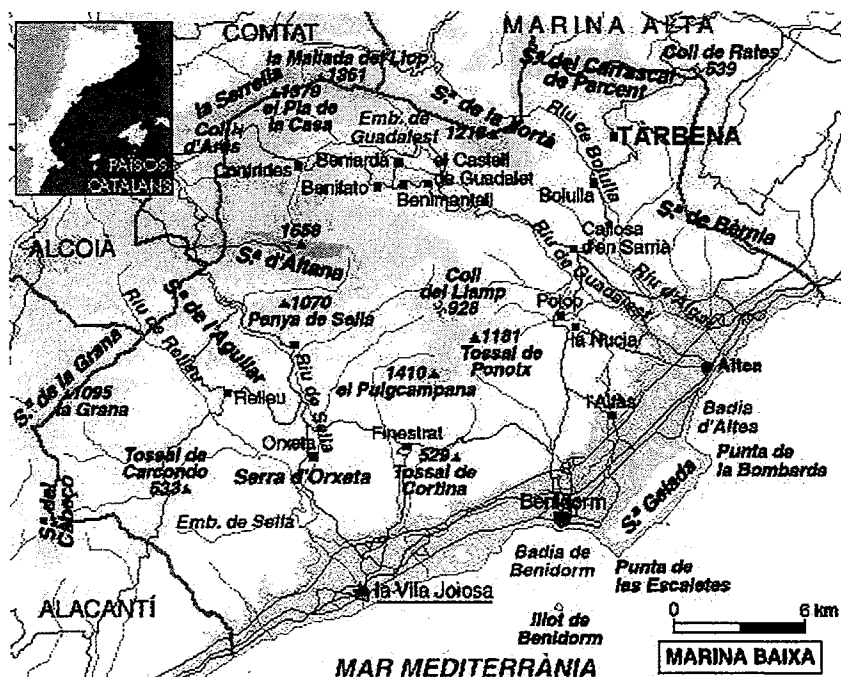


Figura 1.

1: Palma (9). 2: Manacor (4). 3: Inca (4). 4: Lloseta/Binissalem (2). 5: Capdepera / Cala Ratjada (2). 6: Artà (3). 7: Deià (2). 8: Felanitx (2). 9: Maria de la Salut (2). 10: Pollença (2). 11: Sant Joan (2). 12: Sineu (2). 13: Sóller/Fornalutx (2). 14: Son Servera (2). 15: Andratx/S'Arracó (2). 16: Maó/el Castell/Alaior/Sant Lluís (4). 17: Ciutadella (2). 18: el Mercadal/Fornells (2). 19: Ferreries/el Migjorn Gran (2). 20-21: Eivissa (3)/Sant Rafel (1). 22: Sant Josep/Sant Antoni (2). 23: Santa Agnès/Sant Mateu/Sant Joan (2). 24: Formentera (2). 25: Porreres (2). 26: Santanyí (2). A una de les enquestes fetes a Palma, les informadores són dels pobles de Petra i Ariany.

Amb les 2 entrevistes de Tàrbena, doncs, el total de vídeos enregistrats és de 67, que representen aproximadament 100 hores d'entrevistes.



Situació de Tàrbena dins la comarca valenciana de la Marina Baixa.

De les persones entrevistades, 111 són homes i 73 dones. Per professions, tenim gent del camp (la més abundant: senyors de possessió, pagesos, carboners, margers, caçadors, collidors d'oliva, un guardabosc, etc.), gent de mar (mariners, pescadors, peixaters), madones de casa, ferrers, fusters, barbers, forners, sabaters, picapedrers, miners, mestres escudellers, artesans diversos, empleats públics, tècnics superiors diversos, comerciants, polítics, ensenyants, capellans, monges, músics, una carnissera, una actriu, un impressor. A totes aquestes persones (algunes de les quals desgraciadament ja han traspassat en el moment d'aparèixer aquesta publicació i reposen ara en la pau eterna) els vull agrair des d'aquestes línies la seva magnífica i desinteressada col·laboració a l'arxiu.

En el DVD que aquí es presenta¹ oferim una antologia de totes les entrevistes. Hem seleccionat fragments de bona qualitat auditiva i d'interès d'imatge o de contingut, sense que aqueixa tria vulgui dir que no n'hi ha molts d'altres, de fragments bons i interessants, a totes les converses. Aquests fragments són presentats amb la transcripció ortogràfica i amb la transcripció fonètica, de manera que són al màxim d'aprofitables per als lingüistes interessats. (No hi ha, però, la transcripció fonètica d'algunes gloses cantades amb acompanyament instrumental, perquè en aquests casos la pronúncia o bé és distorsionada per les inflexions del cant o bé és parcialment amagada pel so de l'instrument.) La transcripció ortogràfica ho és, ortogràfica, al màxim; és a dir, mantenim les formes escrites segons els principis de l'ortografia catalana, a la qual s'adapten totes les variants fonètiques (*aigua*, no "aigo" ni "èuia"; *posar*, *rosegar*, no "poar" ni "roegar" o "rovegar"; *orelles*, no "oreies" ni "orees"; *festejar*, no "festetjar"; *eixugar*, no "aixuar"; etc). Hi ha qualche excepció, com és el cas del tarbener *m'agüelo*, que no hi ha manera d'escriure'l correctament; o la preposició *an* davant l'article *es* o davant el demostratiu; o *acuaria* / *acuu* (per *acusaria* / *acusi*), dins uns versos d'un romanç; o formes amb afèresi que condicionen el context, com *(ha)via*, dins un vers d'una glosa, en què la síl·laba *ha* no es pot restituir per mor de la mètrica, o quan és precedida d'un

1 Hem presentat el projecte de l'*Arxiu* a diversos congressos internacionals: al Segon Congrés de la Societat Internacional de Dialectologia (Amsterdam, 28 juliol-1 agost, 1997); al XXVII Simposi de la Societat Espanyola de Lingüística (Palma, 15-19 de desembre, 1997); al XV Col·loqui Germano-Català (Friburg de Brisgòvia, 2-4 d'octubre, 1998; publicat: Claus D. Pusch (Hrsg.): *Katalanisch in Geschichte und Gegenwart. Sprachwissenschaftliche Beiträge / Estudis de lingüística*. Tübingen, Stauffenburg Verlag, 2001); al Primer Encontre Friburguès de Corpus Lingüístics (Friburg de Brisgòvia, 6-8 d'octubre, 2000; publicat per Wolfgang Raible, Claus D. Pusch (Hrsg.): *Romanistische Korpuslinguistik / Romance Corpus Linguistics*. Tübingen, Gunter Narr Verlag, 2002); al 4th International Congress of Dialectologists and Geolinguists (Riga, 28 juliol-2 agost, 2003); al XIIIè Col·loqui Internacional de Llengua i Literatura Catalana" (Girona, 8-13, setembre, 2003); i finalment a l'Onzè Col·loqui Lingüístic de la Universitat de Barcelona.

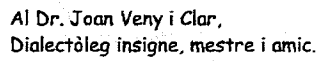
pronom feble *el* (*el (ha)via*); o infinitius plans amb *-r* analògica seguits d'una forma plena de pronom feble, com *escriurer-lo*; o (*a*)*cer* precedit de l'article masculí *es*: *es (a)cer*. A part de superar la dificultat que representa trobar una grafia no normativa adequada a la pronúncia dialectal, l'opció ortogràfica facilita la lectura del text a qualsevol persona, sense necessitat que sàpiga prèviament les realitzacions dialectals pròpies del lloc. Per saber com es pronuncia cada paraula hi ha, precisament, la transcripció fonètica. Una altra cosa són les variants morfològiques, els mots lèxics (els castellanismes) i les construccions, que es respecten absolutament (*llavonses, noltros, sebre, diguent, estassin, etc.; sigle, panaderia, més o menos, etc.*), igual que determinades expressions dialectals com *mem* o *idò*.

L'*Arxiu audiovisual dels dialectes catalans de les Illes Balears* queda conclòs amb aquesta mostra recollida al DVD. Hauria pogut ser molt més extens, i en lloc de 67 vídeos n'hauríem pogut fer 132, i en lloc de 184 persones n'hauríem pogut entrevistar 368... De gent idònia i interessant per conversar-hi encara n'hi ha molta, per sort, de moment. Però ens hem hagut de limitar per una pura i simple qüestió econòmica: cada enregistrament és una despesa important i no podíem continuar fins a l'infinit; ens havíem d'aturar un moment o l'altre. Pensam, tanmateix, que el producte final —tot i haver-hi una certa desigualtat entre les converses— és ben interessant i suficient per tenir una bona representació del català parlat a les Balears per les generacions majors hereves de la llengua tradicional. Els enregistraments complets en vídeo VHF són al nostre arxiu i estan a disposició de tots aquells investigadors a qui els interressi; els drets d'imatge dels entrevistats, però, ens impedeixen fer-ne una difusió ampla. D'altra banda, per limitacions en la capacitat dels DVD, no hi podem incloure els arxius de so separats en format WAV, com era el nostre desig inicial, i fer-ne un volum a

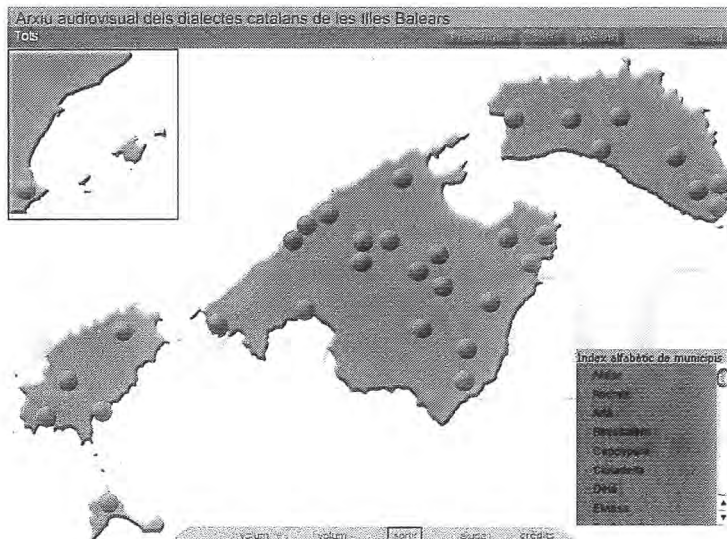
part ens encareix el producte final i supera el pressupost d'aquesta antologia; per això, l'únic que podem fer és oferir aquests arxius a tothom que hi tengui interès. Tant les còpies dels vídeos com les dels arxius de so les podeu demanar, al preu del cost que impliqui fer-les per part dels serveis tècnics corresponents, a aquesta adreça electrònica: jaume.corbera@uib.es.

No vull acabar aquesta presentació sense remarcar que aquest *Arxiu* de fet no és més que una realització incompleta d'un projecte molt més ambiciós: *l'Arxiu audiovisual dels dialectes catalans*, sense més precisions. Quan ja fa anys vaig llançar la idea de fer aquest gran arxiu de tots els nostres parlars varen ser molts els qui s'hi entusiasmaren, de tots els indrets dels Països Catalans. Aquell entusiasme inicial, però, va quedar majoritàriament sols en això, en entusiasme, i ningú més s'hi va posar de veres, a fer-lo real. Crec que la culminació de l'arxiu balear pot ser l'espira que encengui la il·lusió una altra vegada a qualque estudiós apassionat de l'extraordinària riquesa dialectal catalana, que compregui que aquesta riquesa és avui en dia en perill imminent d'extinció (hi duu tant la inèrcia uniformadora del món modern, com la política lingüística assimilacionista encara practicada pels estats espanyol, francès i italià, com la desgraciada política regional de les comunitats catalanes, encabotades a copiar la manera de fer dels estats opressors) i que cal fer qualque cosa perquè, almenys, en quedi constància per a les generacions futures i perquè els dialectòlegs, tant els presents com els futurs, la puguin estudiar amb el màxim de facilitat possible. Amb el present resultat a les mans, doncs, us encoratj, a tots els qui compartiu la meva mateixa inquietud i la meva mateixa passió lingüística, a esforçar-vos perquè dins un període de temps no gaire llarg puguem comptar amb un grandíol arxiu audiovisual de tots els dialectes catalans, més ben fet i aprofitat globalment que el que representa la parcial aportació insular.

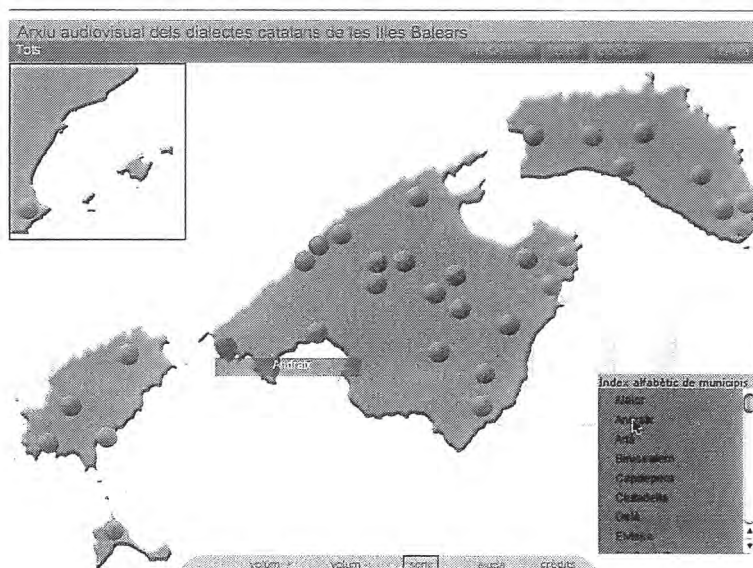
Arxiu audiovisual dels dialectes catalans de les Illes Balears



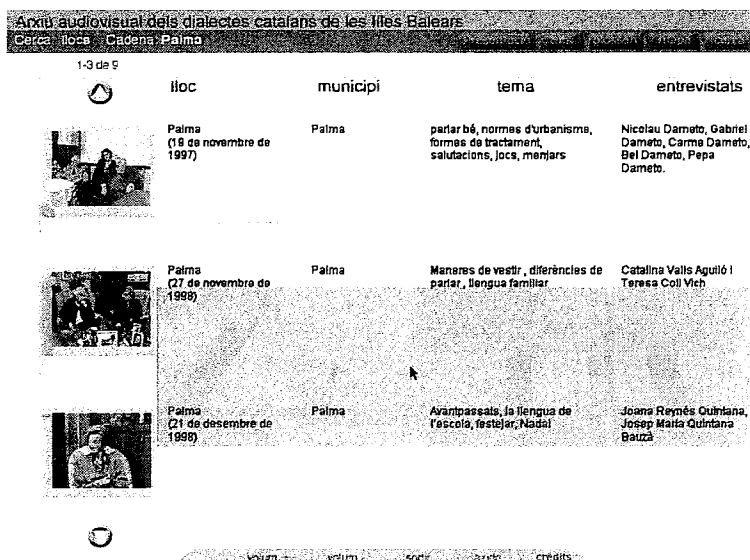
Imatge 1. Presentació.



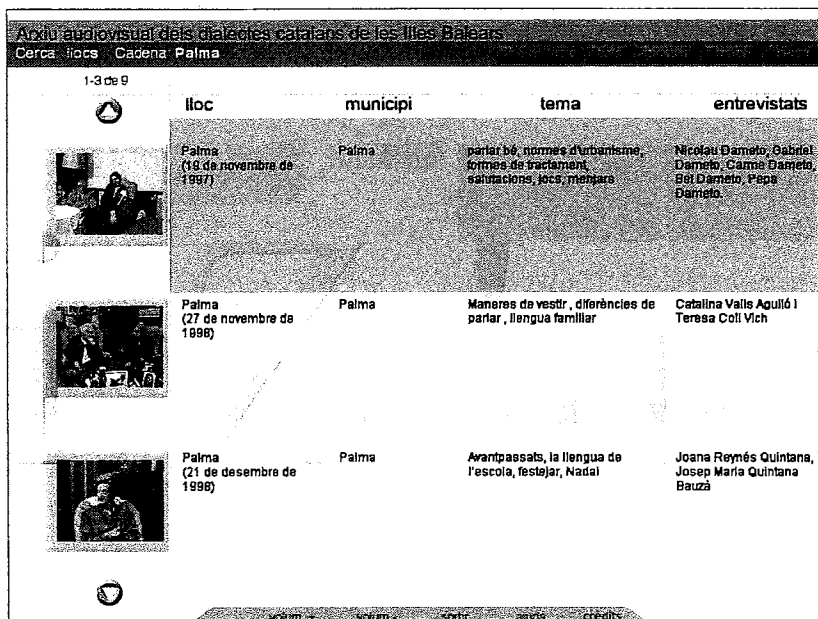
Imatge 2. Mapa de les Illes (amb l'apèndix de Tàrbena) i els menús de maneig del programa.



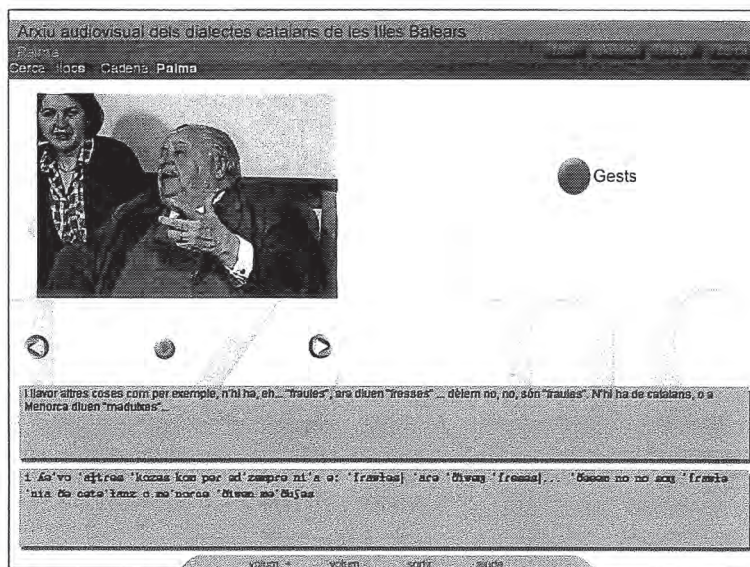
Imatge 3. Accés a la presentació, als menús i la visualització de qualsevol dels vídeos de les localitats corresponents. Passant el cursor per damunt els punts sabem de quina vila es tracta, a la qual també podem accedir des de la llista de la dreta.



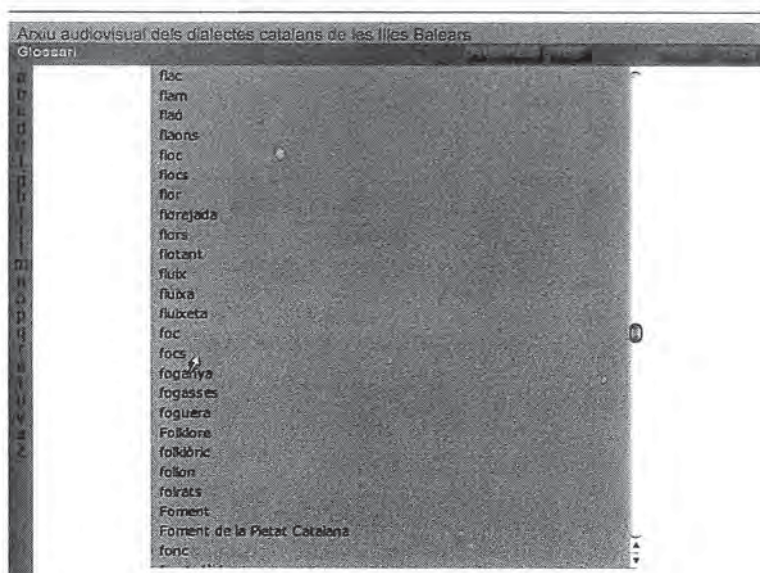
Imatge 4. Si anam, per exemple, a Palma, veiem que hi ha nou vídeos disponibles, presentats per la imatge inicial i les dades del lloc, el municipi a què pertany, els temes de què parlen i els noms de les persones que hi surten.



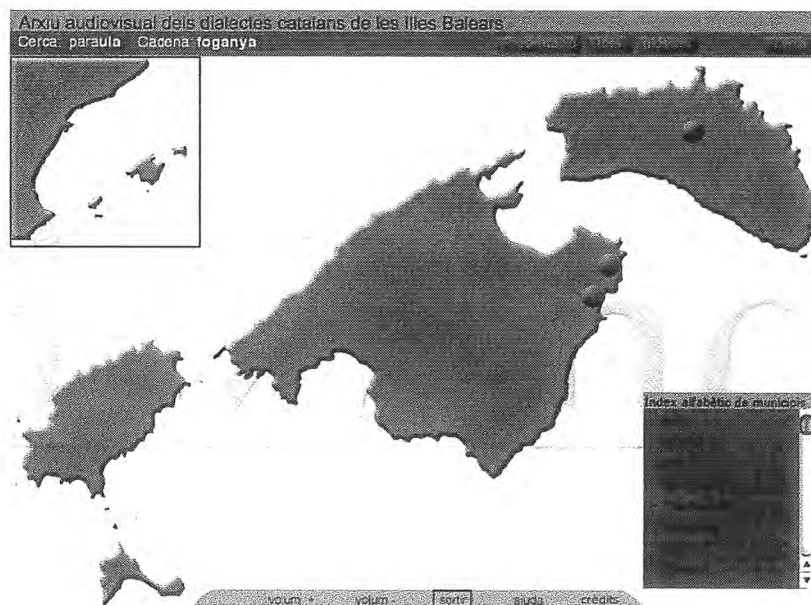
Imatge 5. Si activem el primer, veurem i sentirem la família de Can Dameto de Sa Quartera.



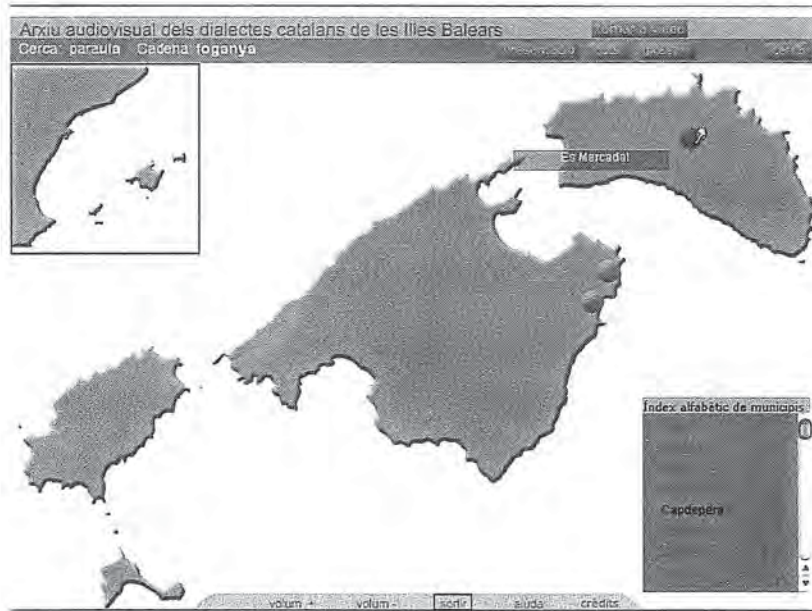
Imatge 6. A l'esquerra tenim la imatge (amb uns botons de control), a la dreta un botó que ens dona entrada a l'aplicació dels gests (en parlarem més endavant) i a baix el text, transcrit en ortografia i en grafia fonètica, del que diuen els parlants.



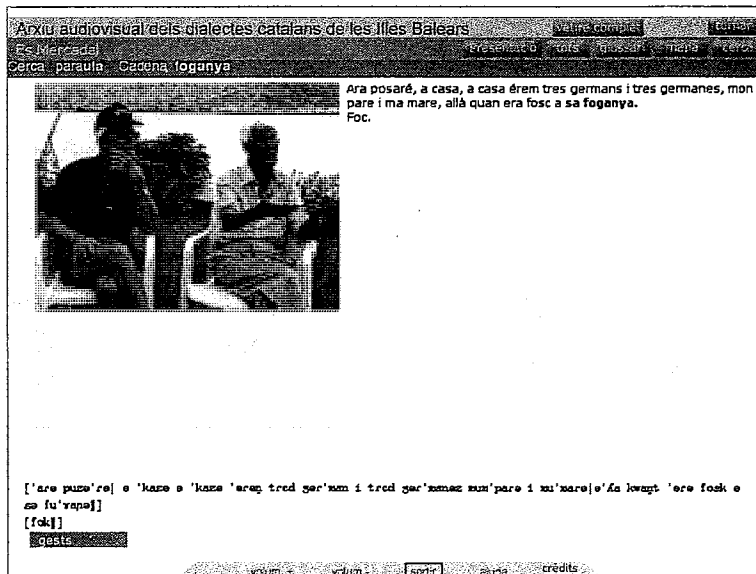
Imatge 7. El menú “Glossari”, accessible en qualsevol moment, ens permet conèixer totes les paraules que contenen tots els vídeos.



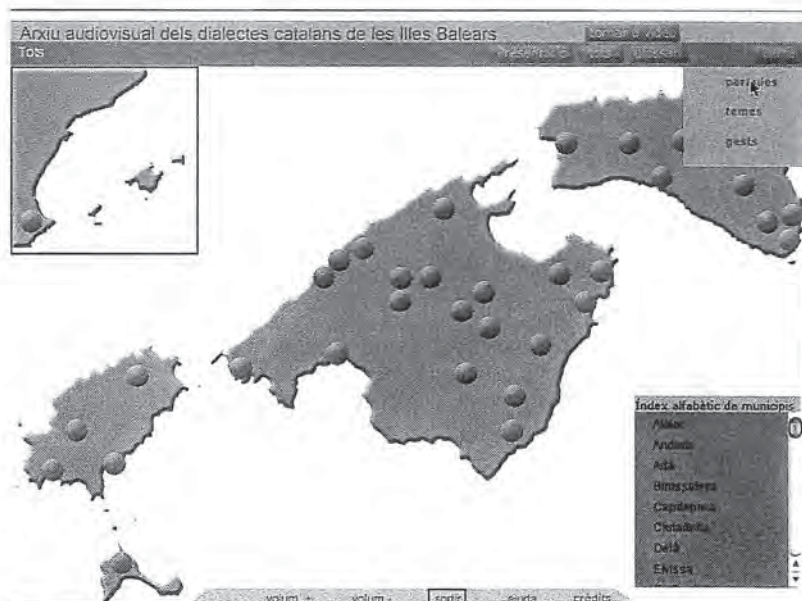
Imatge 8. Si, per exemple, activam “foganya”, trobarem al mapa a quins llocs apareix aquesta paraula.



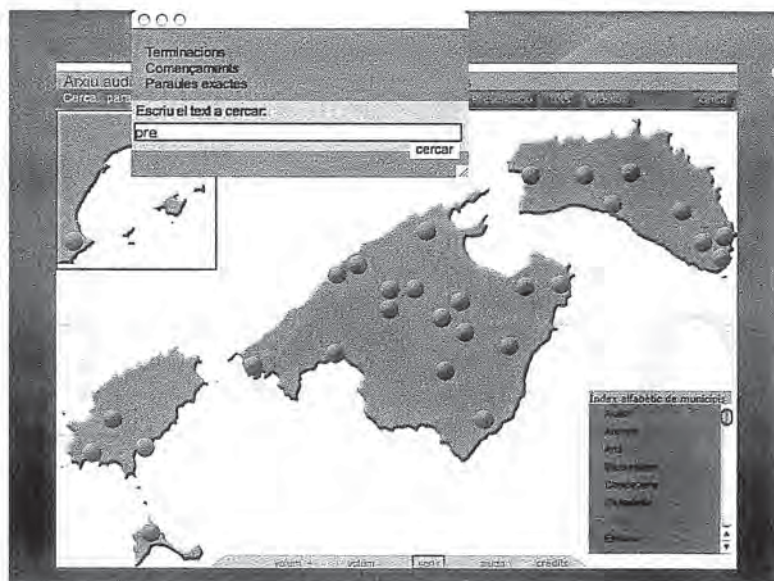
Imatge 9. Seleccionam un dels llocs i ens apareix la presentació del vídeo des Mercadal.



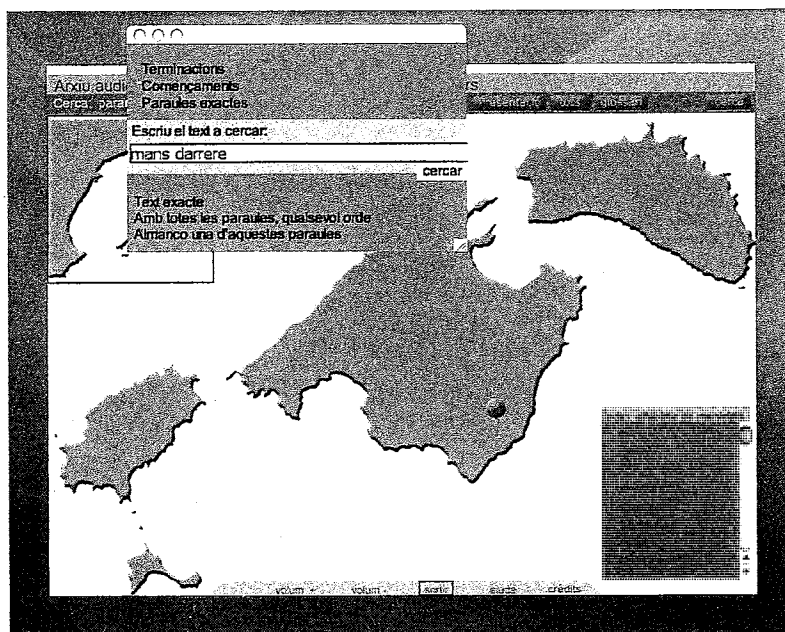
Imatge 10. Activant el vídeo, ens presenta el paràgraf en què surt el mot que ens interessa, sempre amb les transcripcions ortogràfica i fonètica.



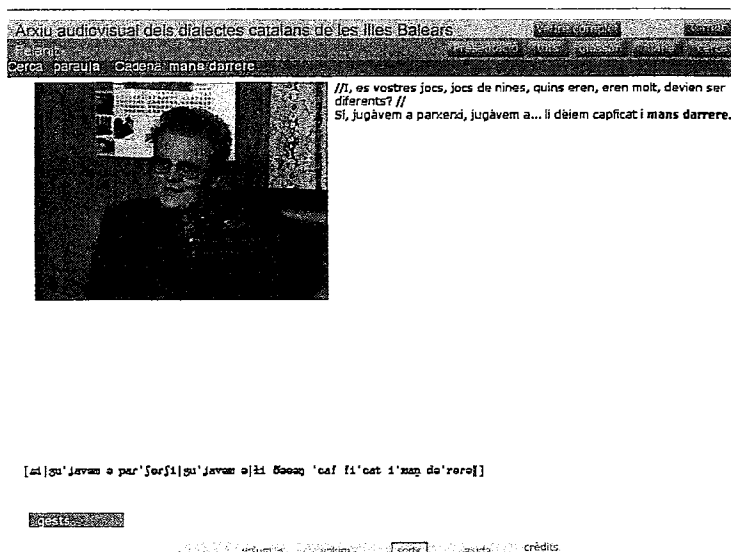
Imatge 11. El menú “mapa” ens torna al mapa que inclou la selecció que hem fet, i el menú “cerca” ens permet fer tres classes de cerques: de formes lingüístiques, de temes i de gests.



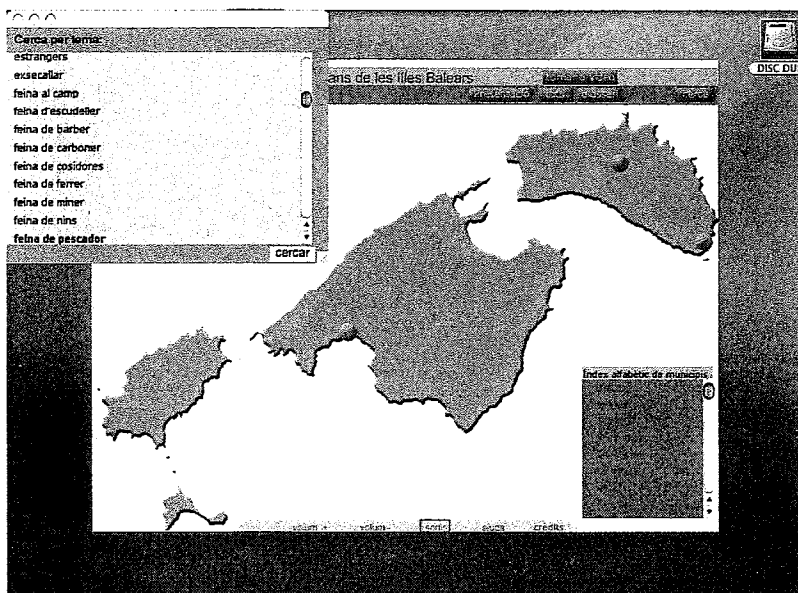
Imatge 12. Amb la primera pantalla podem trobar començaments de paraula.



Imatge 13. També podem trobar terminacions, paraules senceres o més d'una paraula alhora. En tots els casos, el resultat de la cerca se'ns mostra en els punts que ens indiquen a quines localitats trobarem la forma cercada. Si no hi ha la paraula que cercam, ens apareix l'avís de "No trobat".



Imatge 14. El resultat d'aquesta cerca ens duu a Cas Concos (Felanitx).



Imatge 15. La segona classe de cerca que el programa ens ofereix és la del tema.

Arxíu audiovisual dels dialectes catalans de les Illes Balears

Palma

Cerca tema: Cadena feina de pescador

Jo vaig començar amb mon pare, de mon pare, me vaig casar, me vaig voler establir per jo, me vaig fer una barca per jo, o sigui, de d'una barca vella, la vaig renovar i la'm vaig fer per jo, i érem cinc germans a ca nostra,

//A Cabrera.//
A Cabrera
Però a Cabrera mataba.
Aquí n'hi havia molt de peix: però a Cabrera era una cosa grossa.

[go 'vaj koman'as a mon'para] de mon'para na vaj ca'za na vaj vo'le stab'bil par go na vaj 'fe wna 'barca per 'go] o 'aiji de de 'una 'barca 'vee to vaj ramo'va i iag vaj 'fe per 'go] i 'eran sig ger'mans a ca 'nostre]

pp16

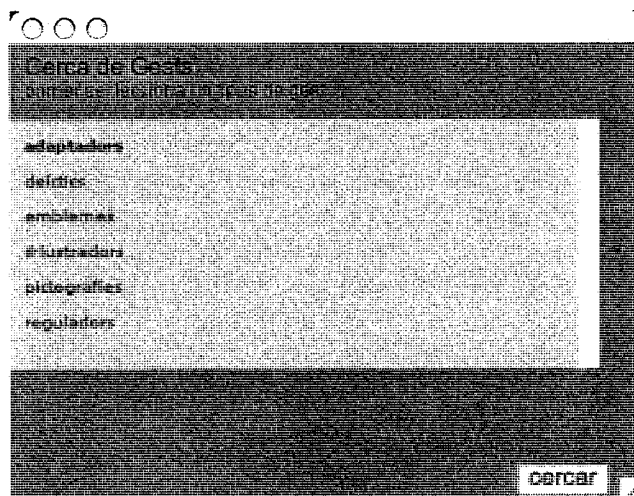
volum 1 volum 2

sortir ajuda crèdits

Imatge 16. La “feina de pescador” que hem demanat la trobam a Palma, a Fornells (el Mercadal) i al Castell. Si s’activa Palma i el vídeo corresponent apareix la imatge on la persona parla d’aquesta activitat. Com veiem, hi ha dos paràgrafs diferents de la conversa que tracten el tema.

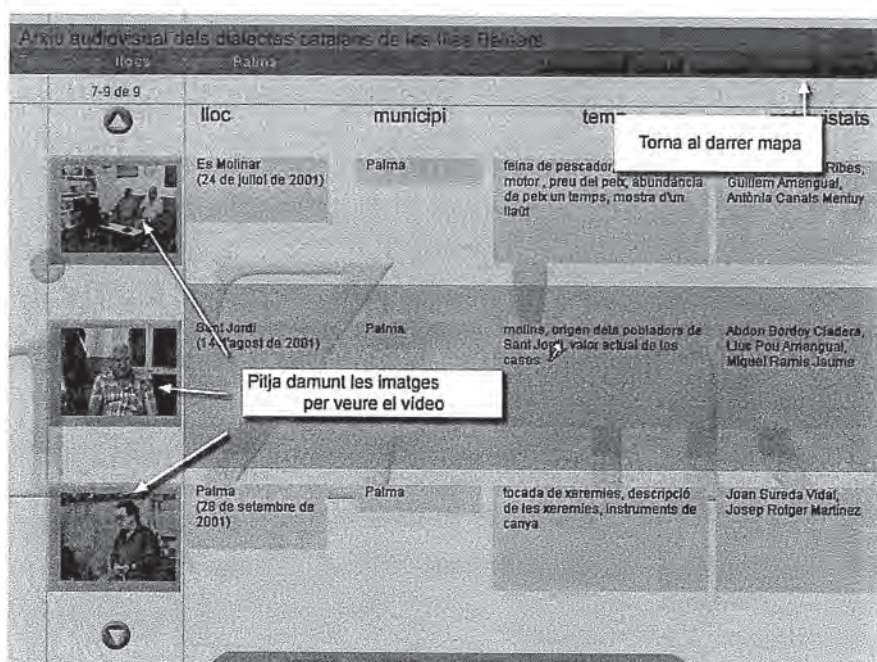


Imatge 17. Una particularitat d'aquest DVD especialment interessant és la possibilitat de crear una base de dades dels gests que fan els entrevistats. Pitjant el botó GESTS tenim les possibilitats que apareixen en la pantalla.



Imatge 19. En pitjar "afegir un gest" l'aplicació guardarà, d'acord amb una tipologia que ella mateixa oferirà i seguint la ruta que l'usuari li indiqui, els paràmetres que permetran localitzar aquell gest des de dins el vídeo o des de fora. Quan aquesta base de dades tenguí ja un mínim contingut, aleshores serà operativa l'opció "cercar gests".

Aquest selector suposa que l'usuari ja té guardats gests que corresponen a cada una d'aquestes categories significatives. Si hi manca la categoria de les "ideografies" és perquè no n'hi ha cap d'enregistrada.



Imatge 20. En totes les passes que fem podem comptar amb l'"Ajuda" que se'n ofereix en la part inferior de la pantalla.

Mapes electrònics i mapes sonors*

Maria Pilar Perea

Universitat de Barcelona

1. Introducció

La geografia lingüística pretén d'estudiar la diversitat dialectal segons l'espai, i un dels instruments tradicionals per representar les dades obtingudes mitjançant les enquestes —de tipus fonètic, morfològic, lèxic i eventualment sintàctic—, han estat els mapes lingüístics, el conjunt dels quals configura un atlas dialectal.

El terme “mapa electrònic” és un terme d'ampli contingut semàntic que actualment s'aplica en els estudis de caràcter geogràfic i climatològic, i també es fa servir en les guies de viatge. Aquest mecanisme s'ha desenvolupat notablement en els darrers temps mitjançant l'ús de les actuals plataformes de sortida: la xarxa internet o els formats en CD-ROM o en DVD.

* Aquest treball s'ha redactat en el marc del projecte VALDIC (“Anàlisi i interpretació de la variació lingüística dialectal a partir de l'explotació d'un corpus oral”), finançat pel MCYT (BFF2001-3798).

La cartografia lingüística ha adoptat també els recursos electrònics¹ en la confecció de mapes i d'atles dialectals amb dues finalitats bàsiques: *a)* fer més accessible la consulta de les dades i *b)* economitjar despeses en l'elaboració d'aquests instruments cartogràfics. És ben conegut de tots els que s'hi dediquen que la tasca de preparació de mapes dialectals (o d'atles) per ser cartografiats en paper —el format de sortida tradicional— és molt difícil, i el cost material de la publicació —temps, recursos humans, recursos tècnics, recursos econòmics, extensió del format i preu d'adquisició del producte resultant— és extremadament elevat. Aquests condicionaments expliquen el retard considerable que existeix entre la recollida de les dades i la seva publicació, i el fet que molt sovint els resultats no s'arribin mai a editar. Amb tot, els atles lingüístics es continuen elaborant com a forma més desitjable de publicació de les dades d'una enquesta dialectal —atesa la facilitat de visualització de la distribució geogràfica de les respostes—, malgrat els costos de tota mena i el volum de les dades, amb la contribució actualment de les eines informàtiques.

El desenvolupament dels procediments d'anàlisi de les dades i dels recursos de cartografia dialectal ha permès també de generar els "atles parlants",² que contenen dades sonores, en els quals s'ha incorporat no sols la transcripció fonètica de les dades enquestades en una determinada

¹ En una disciplina com ara la dialectologia, on l'aplec i la conservació dels materials i el processament i l'anàlisi d'unes dades força nombroses són unes tasques bàsiques, la utilització dels recursos informàtics és —o hauria de ser— un fet gairebé natural. Viereck (1988: 524) situa a l'inici dels anys seixanta el punt de partida de l'ús primerenc de l'ordinador en el tractament efectiu de les dades dialectals de l'anglès. En l'àmbit de la dialectologia catalana cal retardar més aquesta data.

² Vegeu, per exemple, Goebel (1992: 397-412 i 1994: 158-168). A causa de la finalitat descriptiva d'aquest article, no s'aprofundirà en les diverses iniciatives que actualment tenen com a objectiu d'elaboració d'atles sonors ni tampoc es farà referència a nous projectes, com ara AMPER, que pretén d'elaborar un atlas multimèdia de la prosòdia en el domini romànic. Vegeu-ne les referències en l'article de Pradilla, Carrera i Fernández, inclòs en aquest volum.

localitat sinó també el corresponent enregistrament sonor, molt sovint acompanyat de la representació espectrogràfica.

Aquest treball té l'objectiu de presentar un prototipus d'atles electrònic i d'atles sonor —entès com a conjunt de mapes informatitzats— que actualment desenvolupa el *Grup de Variació Dialectal*, format per alguns membres del Departament de Filologia Catalana, juntament amb professors d'altres universitats. Es tracta d'un prototipus que encara està en fase de desenvolupament, però que ja pot donar una idea dels objectius que es planteja i dels resultats que vol donar a conèixer.

2. Els materials dialectals

Malgrat que ja s'ha fet referència en altres treballs —cf. Lloret i Perea (2002) i Alturo, Boix i Perea (2002)—, cal fer, per entendre el funcionament de les diverses aplicacions que es descriuran a continuació, una breu presentació dels materials que s'han utilitzat en la confecció dels mapes sonors. Les dades que componen el corpus van ser obtingudes mitjançant enquestes efectuades en tots els caps de comarca —o localitats equivalents— del domini lingüístic català. El nombre d'indrets visitats fou de 86. Les enquestes consistien a elicitat les respostes d'un qüestionari de tipus foneticomorfològic, i la producció d'un text lliure, de caràcter espontani.

En la fase de processament dels materials, els textos, que tenen una durada que oscil·la entre els 5 i 10 minuts, s'han transcrit fonèticament i ortogràficament, aplicant-hi unes determinades convencions (cf. Viaplana i Perea: 2003), i se'ls ha tractat informàticament amb vista a aconseguir la sincronització entre la seqüència sonora i les corresponents transcripcions

gràfiques. El procediment de cartografia automatitzada, però, s'ha aplicat solament a les dades del qüestionari, atès que es tracta d'un qüestionari tancat pel que fa a les preguntes, i des d'aquesta perspectiva, les dades són fàcilment comparables. Les respostes del qüestionari han estat transcrits ortogràficament i fonèticament i s'han agrupat per tipologies en diverses bases de dades interrelacionades. Les bases de dades, doncs, que conformen el corpus,³ juntament amb el nombre de registres que les integren, són les següents:

	Núm. de registres
Fonètica	49.812
Articles	5.049
Demonstratius	1.512
Possessius	2.006
Locatius	803
Pro. personals	1.608
Clítics	21.591
Verbs regulars	68.152

Taula 1

3. Les característiques dels informants

Atesa la selecció que cal fer de les respostes obtingudes durant el procés d'enquestació per tal d'aplicar uns criteris de representativitat en el procés de cartografia automatitzada, és necessari fer esment dels criteris que acomplien els tres informants enquestats en cadascuna de les localitats:

³ El corpus rep el nom genèric de COD (corpus oral dialectal).

- a) havien d'haver nascut, com també els seus pares, en la localitat seleccionada i no havien d'haver fet llargues estades en altres indrets;
- b) havien de ser homes i dones l'edat dels quals oscil·lés entre els 30 i els 45 anys. [L'edat mínima assegurava, en aquell moment, que els informants no havien estat escolaritzats en català.];
- c) havien de pertànyer a una classe social mitjana;
- d) no havien de tenir una formació acadèmica superior.

4. Els antecedents cartogràfics

Els registres que contenen les transcripcions de les dades dels materials obtinguts a partir de les respostes del qüestionari es poden consultar en les diverses bases de dades on estan emmagatzemats; un petit tast s'ofereix en la pàgina web, encara en desenvolupament, <http://www.ub.es/lincat>. Amb tot, es va creure convenient que una fase addicional del processament dels materials havia de permetre'n la representació cartografiada. Per dur-la a terme es van adoptar els paràmetres que s'havien aplicat en el tractament de les dades de *La flexió verbal en els dialectes catalans* d'Alcover i de Moll (cf. Perea 2001), els quals donaven com a resultat la generació de mapes informatitzats que presentaven, a grans trets —en un àmbit dialectal general i en l'àrea geogràfica dels sis principals dialectes del català: rossellonès, oriental, occidental, balear, valencià i alguerès—, les característiques següents:

- a) Per bé que es poden presentar separadament, és possible d'incloure en una mateixa pantalla els resultats (limitats a una sola resposta, que

incorpora, quan escau, les variants fòniques corresponents) en forma de llista i en forma cartografiada.

b) Quant a la tipologia de la representació cartogràfica, es poden elaborar mapes fonètics (que contenen els resultats fonètics) —vegeu figura 1— o mapes simbòlics —vegeu figura 2—, que superen les dificultats de plasmació dels resultats quan la densitat de localitats distribuïdes en el mapa és molt elevada o quan els resultats presenten polimorfisme. En la pantalla, i situant el cursor damunt un símbol, es pot llegir, transcrit fonèticament, el resultat que hi està relacionat.

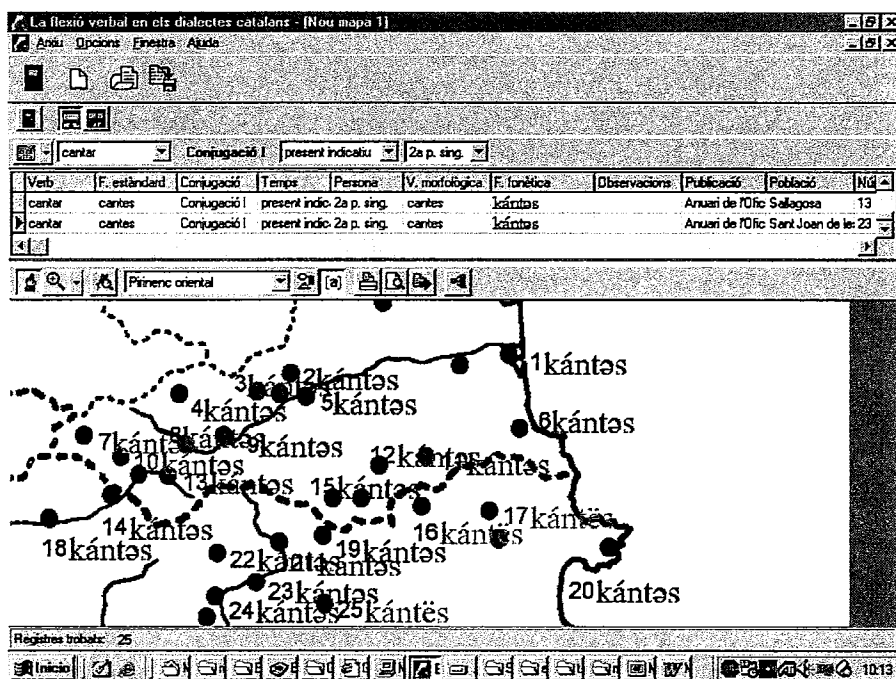


Figura 1. Mapa que inclou les representacions fonètiques de la segona persona del singular del present d'indicatiu del verb *cantar* en l'àrea corresponent al pirinenc oriental.

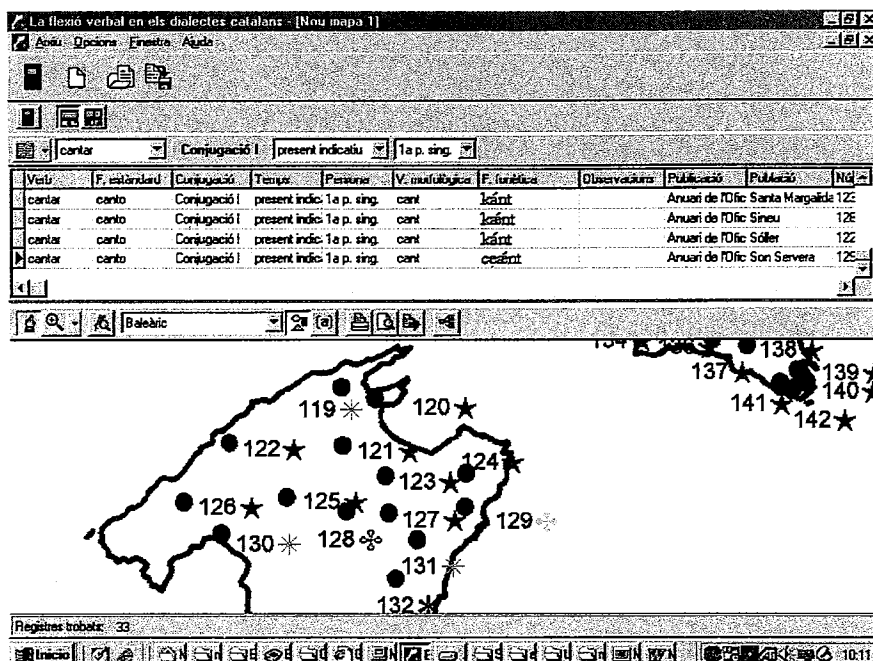


Figura 2. Mapa simbòlic dels resultats relatius a la primera persona del singular del present d'indicatiu del verb *cantar* en l'àrea corresponent al dialecte balear, i en particular al subdialecte mallorquí.

Pel que fa als formats de sortida, cadascun dels mapes confeigits (el mapa general del domini lingüístic català i els sis mapes d'àrees), com també les llistes que es generen de manera paral·lela com a resultat de l'aplicació de determinats criteris de cerca, es poden emmagatzemar en el disc dur de l'ordinador per fer-ne una consulta posterior, o es poden imprimir en format DINA-4. Vegeu en la figura 3, la visualització de la preimpresió dels resultats obtinguts en l'àrea del dialecte valencià corresponents a la primera persona del singular del present d'indicatiu del verb *anar*. En el format d'impressió, hi ha l'opció d'incloure el nom de les localitats.

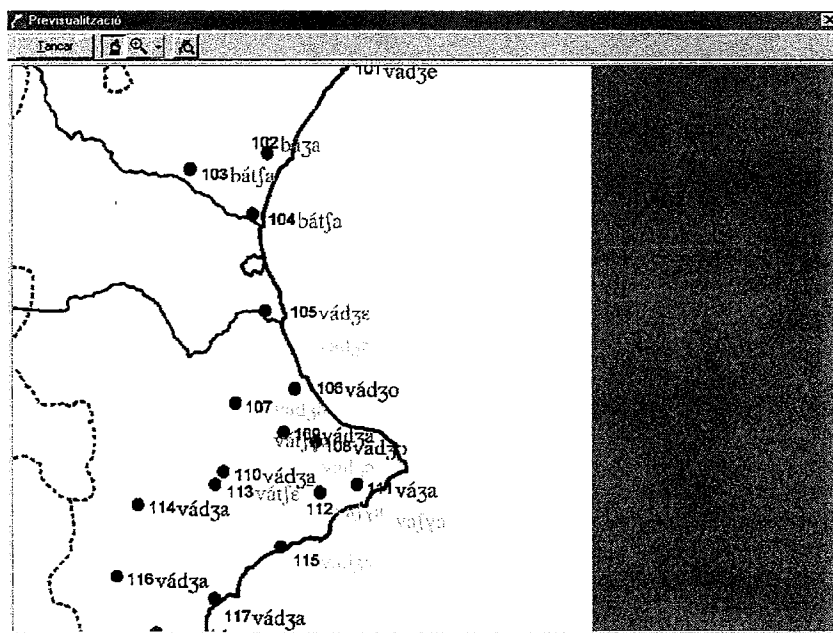


Figura 3.

4. La cartografia automatizada de les dades del COD. Les etapes

Els paràmetres esmentats i els procediments de visualització dels resultats que s'han indicat en l'apartat anterior s'han aplicat, amb petites adaptacions, a les dades del COD (corpus oral dialectal). El resultat ha estat el següent (vegeu figura 4), si es prenen com a exemple les dades corresponents als articles en l'àrea balear, i concretament a Mallorca.

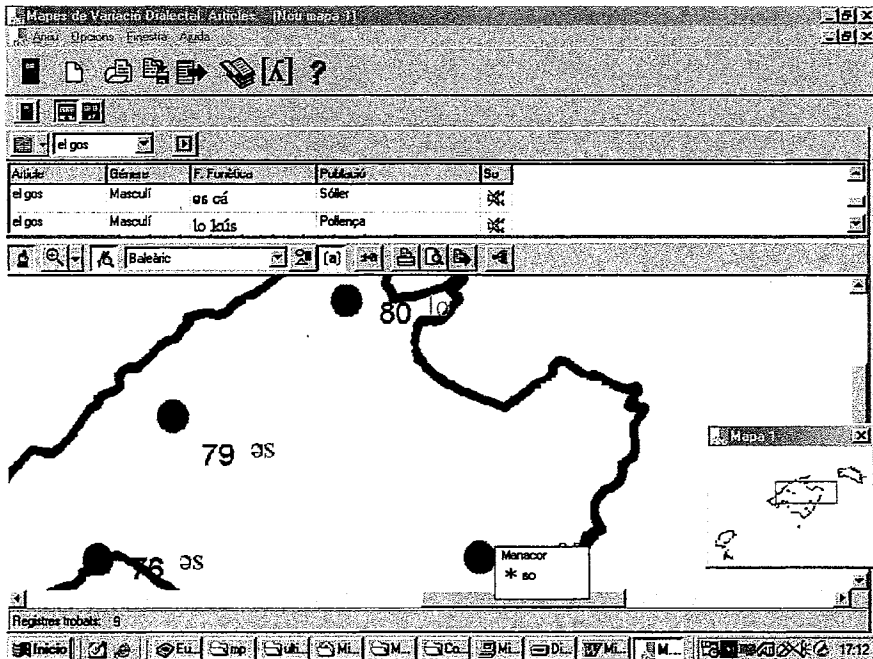


Figura 4.

Tanmateix, el procés de cartografia, amb vista a l'elaboració de mapes sonors, no acaba aquí. Cal accomplir diverses etapes preliminars, que s'indiquen a continuació.

4.1. La selecció dels registres

La base de dades general, on estan emmagatzemades, per classes, les dades obtingudes mitjançant el qüestionari foneticomorfològic —vegeu la taula 1 representada més amunt—, conté un nombre molt elevat de registres, els quals presenten, sovint, solucions alternatives en relació amb les respostes obtingudes en cadascun dels indrets. Ja s'ha indicat que en cada

localitat s'obtenien, per a cada ítem, les respostes de tres informants. Algunes d'aquestes respostes —la majoria— coincidien, però d'altres eren divergents.

Tenint en compte la variació esmentada, no és desitjable que es cartografiïn tots els resultats —coincidents i divergents—, atesa la dificultat de visualització. És evident, com s'ha observat en el procés de cartografia dels articles, que es compta amb un espai limitat. Per aquest motiu, un cop transcrites les dades, i amb vista a l'elaboració de mapes sonors, s'ha aplicat, als materials, un procediment semiautomàtic de selecció. Així, en el cas de coincidència absoluta de respostes, el programa en seleccionava automàticament i aleatòriament una (ja que no existia cap divergència). Quan hi havia dues respostes coincidents i una de divergent, el programa seleccionava el registre majoritari (una de les coincidents). En darrer terme, en els casos on hi havia divergència absoluta (les respostes de cadascun dels informats enquestats eren divergents), els resultats es marcaven en una llista, de manera que, aquest cop manualment —per això s'ha indicat més amunt que la selecció es feia de manera semiautomàtica—, l'investigador seleccionava la forma que es considerava més representativa de l'indret en qüestió.

A continuació es mostren algunes pantalles que inclouen els resultats obtinguts d'activar el programa de selecció, que ha rebut el nom de "Corrector".⁴ Aquest terme no ha de suggerir que el programa fa una correcció en sentit estricte; el que corregeix són les alternances que apareixen en les respostes amb l'objectiu de facilitar-ne la cartografia automatitzada, o de poder-hi aplicar altres mètodes d'anàlisi, com ara el dialectomètric.⁵

⁴ Es tracta d'un programa *ad hoc*, elaborat pels tècnics informàtics de l'empresa Grubit, els quals col·laboren, ja fa uns quants anys, en la informatització de les dades del projecte.

⁵ Vegeu la presentació de Clua que apareix en aquest volum.

Les figures 5, 6 i 7 donen compte dels registres resultants, corresponents, respectivament, a les bases de dades que inclouen els possessius, els locatius i els demostratius de tres localitats (Castelló, Cocentaina i Eivissa), un cop s'ha aplicat el programa "Corrector".

Ortografia	Localitat	Fonètica	Informant
el meu	Castelló	el méw yés	FJBM
la meua	Castelló	la méwa yésa	FJBM
els meus	Castelló	el méwz yécos	FJBM
les meves	Castelló	lez méwez yécos	FJBM
el teu	Castelló	el téw yáí	OMS
la teua	Castelló	la téwa yáía	OMS
els teus	Castelló	es téwz yáíe	OMS
les teves	Castelló	lez téwez yáíe	OMS
el seu	Castelló	el séw amík	OMS
la seva	Castelló	la séwa amíya	OMS
els seus	Castelló	el séwz amíke	OMS
les seves	Castelló	le séwez amíye	OMS
el nostre	Castelló	el nóstre βéí	OMS
la nostra	Castelló	la nóstra βeíma	OMS
els nostres	Castelló	ez nóstres βeíne	OMS
les nostres	Castelló	lez nóstres βeíne	OMS
el vostre	Castelló	el βóstre síí	OMS
la vostra	Castelló	la βóstra síía	OMS
els vostres	Castelló	ez βóstres sííe	OMS
les vostres	Castelló	lez βóstres sííe	OMS

Figura 5. Possessius: registres seleccionats.

Malgrat que alguns cops és coincident (vegeu les inicials corresponents a l'informant de la figura 6), no és necessari que els resultats obtinguts pertanyin a un mateix informant (vegeu les figures 5 i 7); el que és rellevant és la representativitat de les dades d'un determinat indret.

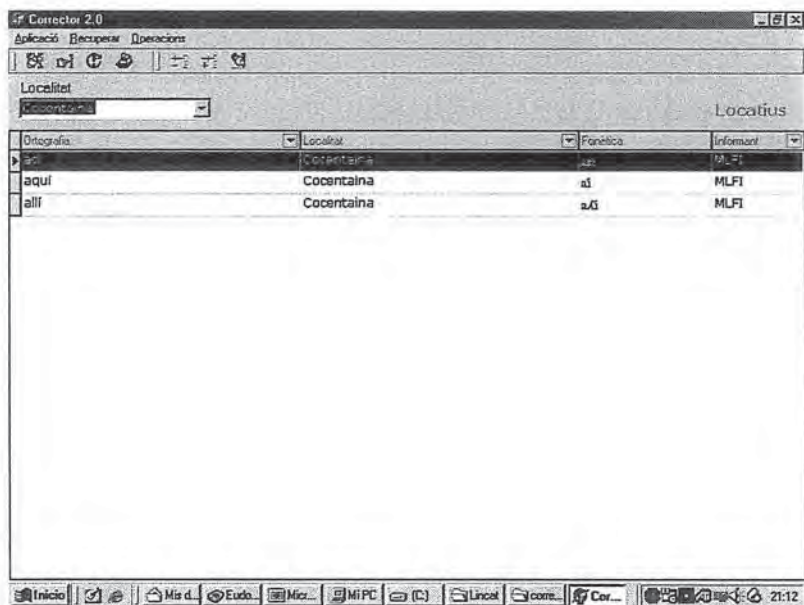


Figura 6. Locatiu: registres seleccionats.

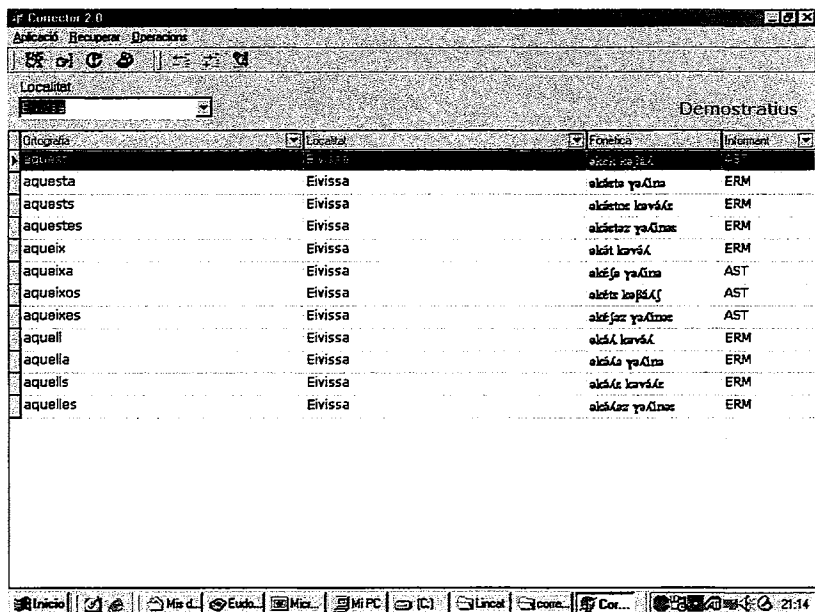


Figura 7. Demostratiu: registres seleccionats.

D'altra banda, el fet que s'hagi seleccionat un registre únic per a cada resposta en cadascuna de les localitats no implica que s'hagi perdut la informació que s'ha eliminat. El conjunt total de registres es troba disponible en una base de dades general, feta amb l'aplicació ORACLE, que es pot consultar sempre que es vulgui obtenir una informació més completa, amb vista a l'elaboració de treballs de caràcter més exhaustiu.

4.2. La segmentació dels registres

Les dades del COD estan basades en registres sonors —a diferència de les dades de *La flexió verbal*, de les quals només es posseeixen les transcripcions ortogràfica i fonètica. Per aquest motiu, es desitjable que els mapes electrònics generats amb els materials del COD incorporin els registres sonors corresponents.

Les operacions de segmentació de la seqüència sonora i d'associació del text (fonètic i ortogràfic) amb el so corresponent són molt laborioses. Per facilitar-les, s'ha desenvolupat una aplicació informàtica *ad hoc*, que s'ha anomenat “Retalls dialectals”. Aquesta aplicació aconsegueix les fases següents:

1. Selecció d'una determinada localitat (vegeu la figura 8).

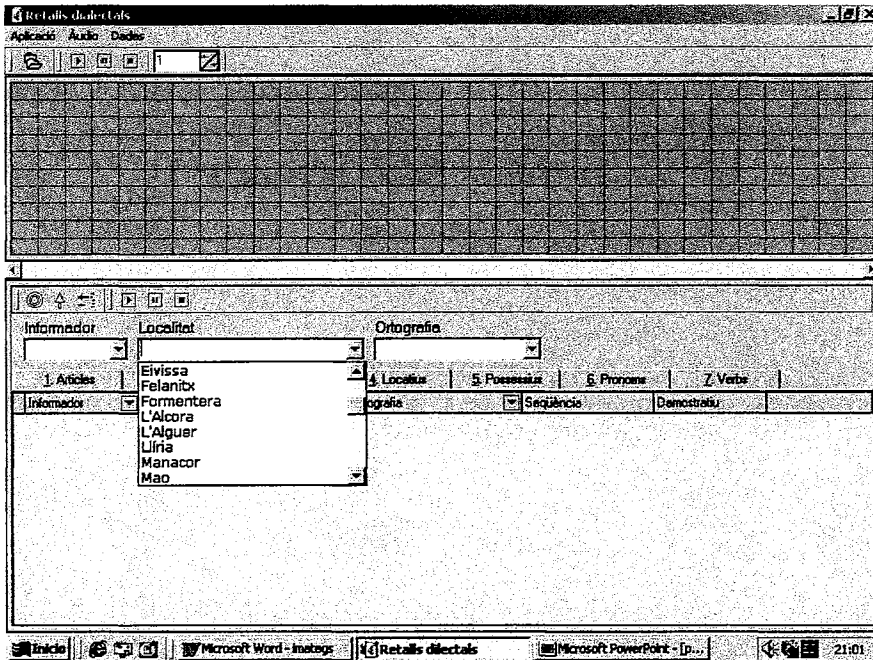


Figura 8.

2. Selecció d'una base de dades determinada, de les set que hi ha disponibles a ORACLE. La figura 9 mostra la base de dades corresponent als possessius i la llista d'alguns dels registres seleccionats.

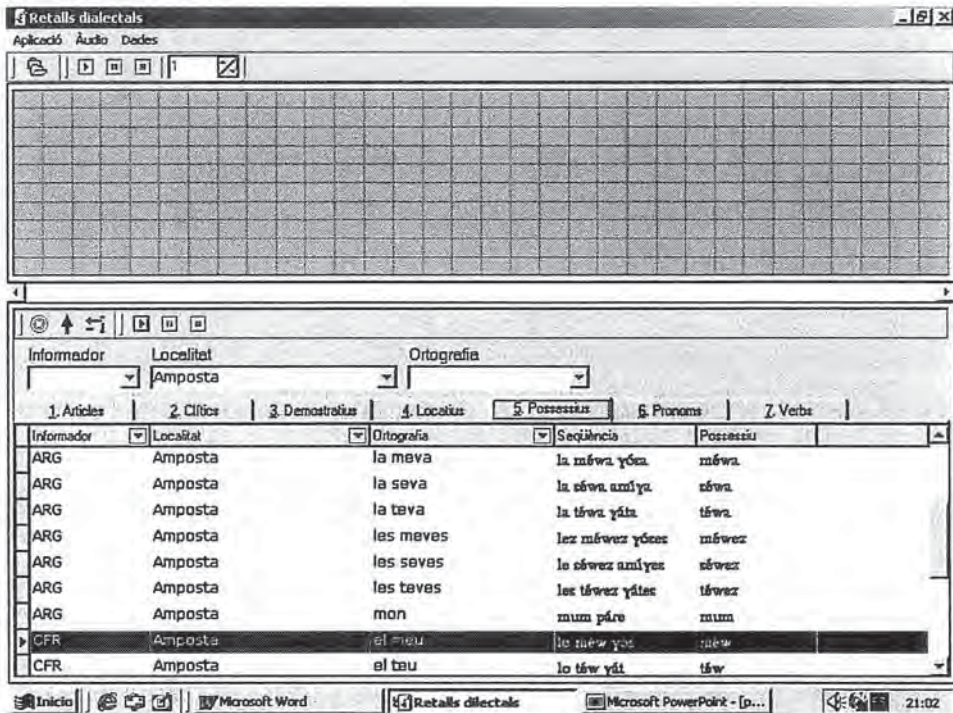


Figura 9.

5. Els mapes sonors

La base de dades corresponent als locatius és la que a hores d'ara s'ha cartografiat automàticament, i els registres, un cop seleccionats, han configurat sengles mapes sonors. Les figures 10 i 11 mostren la intervenció, en aquesta base de dades, de l'aplicació "Retalls dialectals".

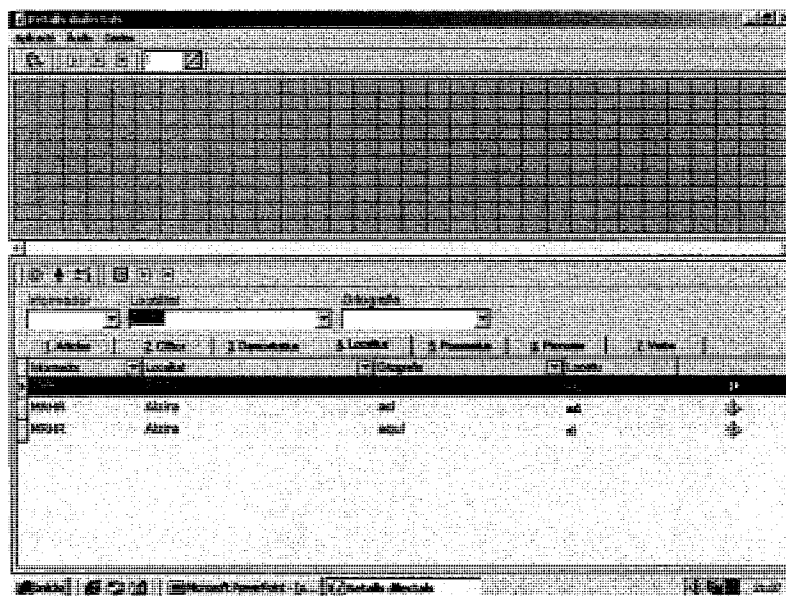


Figura 10.

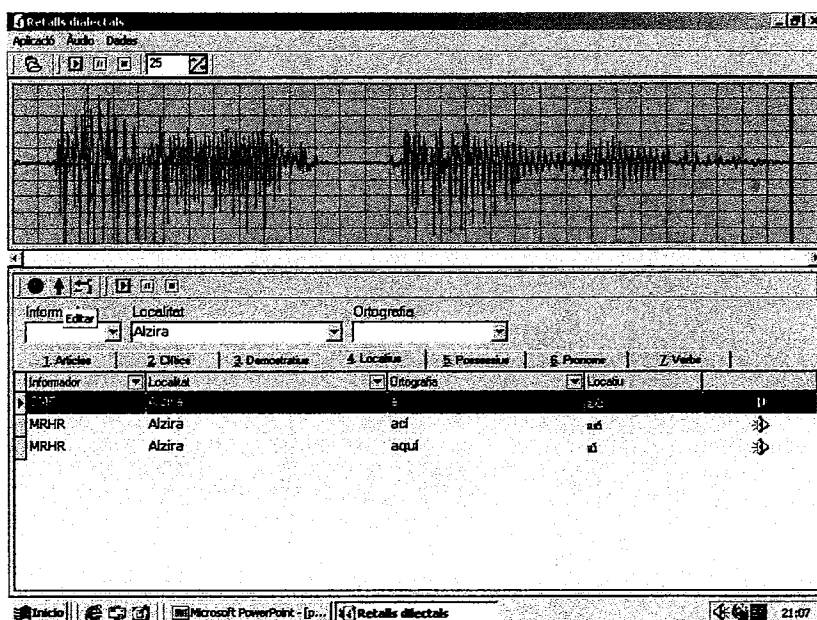


Figura 11.

La figura 11 dóna compte de l'associació de la seqüència sonora amb el registre ortogràfic corresponent, com també amb tota la informació que aquest duu associada (les sigles de l'informant, la localitat i la transcripció fonètica).

Un cop feta aquesta operació amb cadascun dels registres de la base de dades, es pot editar un CD-ROM que permet no sols la cartografia automatitzada dels resultats corresponents —en aquest cas, els locatius— sinó també la possibilitat d'associar, en el mapa, per a cada una de les respostes, les representacions fonètica i ortogràfica amb la seqüència fònica corresponent (vegeu figura 12).

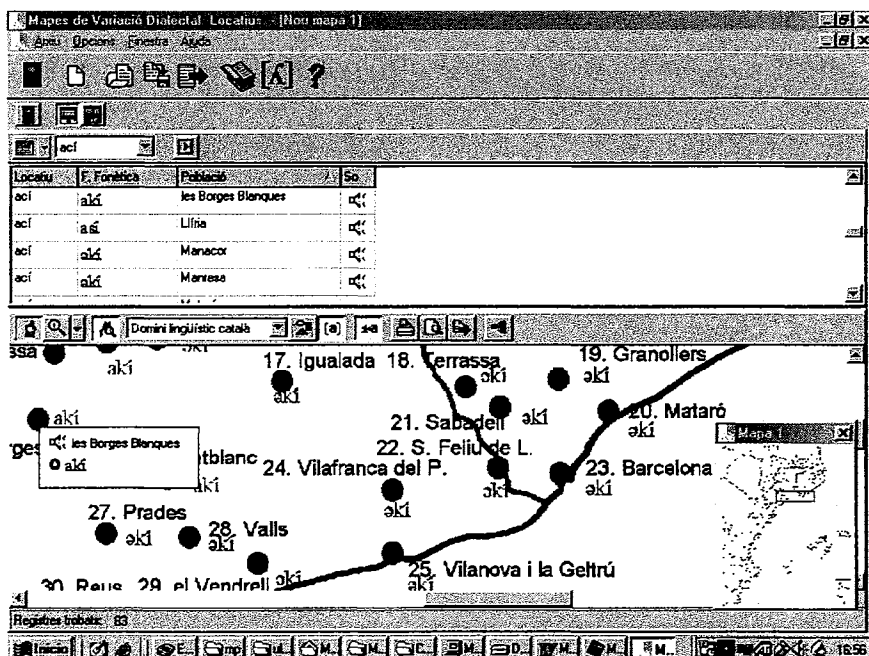


Figura 12.

Referències bibliogràfiques

- ALTURO, Núria, Emili BOIX i Maria Pilar PEREA (2002): «Corpus de Català Contemporani de la Universitat de Barcelona (CUB). A general presentation», dins C. D. Pusch i W. Raible, *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache*, Tübingen: Gunter Narr Verlag, 155-170.
- GOEBL, H. (1992) "L'atlas parlant dans le cadre de l'Atlas linguistique du ladin central et des dialectes limitrophes (ALD)", dins *Actes du Congrès Internacional de Dialectologie/Nazioarteko Dialektologia Biltzarra*, Agiriak (Bilbo, 1991), Academie de la Langue Basque, Bilbo, 397-412.
- _____ (1994) "L'Atlas Linguistique du Ladin Central et des dialectes limitrophes (première partie, ALD I)", dins Pilar García Mouton (1994), *Geolingüística. Trabajos europeos*, Madrid: Consejo Superior de Investigaciones Científicas, 155-168.
- LLORET, Maria-Rosa i Maria-Pilar PEREA (2002): «A report on "The Corpus Oral Dialectal del Català Actual (COD)"», *Dialectologia et Geolingüística*, 10, 49-76.
- PEREA, Maria Pilar (2001): *La flexió verbal en els dialectes catalans d'A. M. Alcover i F. de B. Moll. Les dades i els mapes*, Palma de Mallorca: Conselleria d'Educació i Cultura. Govern de les Illes Balears, edició en CD-ROM.
- VIAPLANA, Joaquim i Maria Pilar PEREA (2003): *Textos orals dialectals. Una selecció*, Barcelona: PPU.
- VIERECK, W. (1988): "The Computerisation and Quantification of Linguistic Data: Dialectometrical Methods", dins A. R. Thomas, *Methods in dialectology. Proceedings of the sixth international conference held at the University College of North Wales*, 3rd-7th August 1987, Clevedon Philadelphia: Multilingual Matters Ltd, 524-552.

II

L'EXPLOTACIÓ DE CORPUS LINGÜÍSTICS

La phonologie du français contemporain au Pays basque et son contexte sociolinguistique

Jacques Durand*, Michel Aurnague** et Julien Eychenne*

* CNRS-UMR 5610, Université de Toulouse-Le Mirail

** CNRS-UMR 5478, Université de Bordeaux III

& Université de Pau et des Pays de l'Adour

0. Introduction

Enquêter sur la phonologie du français au Pays basque n'est pas chose facile. Se pose en premier lieu le rapport à la langue basque ou euskara.¹ A-t-on affaire à des témoins qui sont en contact permanent avec la langue basque (bilinguisme équilibré ou pas, actif ou passif) ou à des sujets monolingues? Et même si les témoins ne parlent pas basque, quelle est leur "positionnement" face aux questions d'identité basque et au statut du français dans le dispositif

¹ Ce travail a bénéficié de l'aide généreuse de l'ILF et de la DGLFLF. Nous remercions ceux qui, comme E. Gomez-Imbert, G. Jendraschek, C. Lyche et G. Turcsan, ont eu la gentillesse de lire cet article et de nous faire part de leurs réflexions. Nous remercions également les témoins qui ont accepté sans rechigner de participer à notre enquête. Sans eux, les données qui forment le soubassement de cet article et des études en cours n'existeraient tout simplement pas.

sociopolitique de la France contemporaine? Il faudrait assurément de grandes enquêtes sociolinguistiques sur le français au Pays basque, opposant catégories sociales, classes d'âge, hommes et femmes, villes et campagnes, bascophones et non bascophones, pour mieux comprendre le fonctionnement du français et cerner les domaines les plus pertinents de variation. Les spécialistes de linguistique au Pays basque préfèrent souvent, et on les comprend parfaitement, se concentrer sur la langue basque, à la fois pour l'intérêt intrinsèque qu'elle présente (par exemple, du point de vue typologique) et dans un but patrimonial, puisqu'elle fait partie des langues en péril.²

Nous nous proposons dans cet article d'examiner le français au Pays basque à la lumière d'une enquête réalisée au sein du projet «Phonologie du français contemporain: usages, variétés et structure» (dorénavant PFC). Après avoir présenté le projet PFC (1), nous proposons une synthèse sur la situation sociolinguistique au Pays basque (2), après quoi nous exposerons l'inventaire phonologique du français au pays basque (3). Enfin, nous finirons en abordons deux phénomènes phonologiques plus en détail: le schwa (4) et le /r/ (5).

1. Le projet PFC et sa méthodologie

Le projet PFC vise à constituer un vaste corpus oral couvrant plus de 400 heures d'enregistrements et un minimum de quarante points d'enquête. L'axe central de PFC est socio-phonologique (Durand, Laks & Lyche 2002, Durand

² Les références sur le basque sont nombreuses et le lecteur se reportera à Allières (1979, 1999) pour un point de vue synthétique et à Trask (1997) pour une histoire du basque embrassant tous les domaines de la description linguistique.

& Lyche 2003a) sans suivre les contraintes d'une enquête sociologique *stricto sensu*. S'il se situe dans la continuité des grandes enquêtes phonologiques, PFC n'embrasse pas à la lettre une démarche sociolinguistique au sens technique du terme, dans la mesure où l'échantillonnage dans chaque point d'enquête n'est pas déterminé par un nombre élevé de témoins sélectionnés aléatoirement et en équilibrant de façon rigoureuse tous les paramètres jugés essentiels (sexe, âge, classe sociale, provenance ethnique, etc.). Notre enquête, à la différence de nombreuses enquêtes sociolinguistiques, ne se focalise pas non plus sur quelques variables prises dans l'ensemble du système afin de les corrélérer à des facteurs stylistiques ou sociaux. Nos objectifs sont plus descriptifs et plus globaux. Il s'agit dans un premier temps d'envisager tous les volets phonologiques de la variation pour pouvoir par la suite proposer des analyses de phénomènes typiquement variables comme le schwa ou la liaison par exemple. La couverture géographique ne prétend évidemment pas à l'exhaustivité mais les points d'enquête quadrillent l'Hexagone et comprennent aussi d'autres pays francophones comme le Canada, la Suisse, la Belgique et des départements d'outre-mer comme La Réunion.

Ce travail se rapproche, dans sa conception, de l'enquête préparatoire de Walter (Walter 1982), travail important parce qu'il constitue une enquête sur les variétés du français qui s'appuie sur des méthodes partagées par PFC. H. Walter plaide par exemple pour des enquêtes linguistiques et sociolinguistiques qui fassent une place fondamentale aux idiolectes et note que: «William Labov, dans *Sociolinguistique*, indique bien en effet qu'il a commencé son étude par une enquête préparatoire de 70 interviews (p. 95), mais à aucun moment il ne donne de précisions sur les différents systèmes de ses informateurs. Le lecteur n'a connaissance de cette enquête préliminaire que dans la mesure où elle a permis à l'auteur d'isoler les quelques variables

phonologiques pouvant être mises en relations avec la différenciation sociale» (1982: 18).

Dans PFC, la description porte sur chaque locuteur enregistré avant de s'engager sur le terrain des généralisations pour une région donnée. Walter offre une discussion intéressante des paramètres qui peuvent influencer la variation: l'âge, le sexe, la catégorie socioprofessionnelle, le niveau culturel (ou niveau d'instruction), la situation géographique et le style. PFC est sensible à l'ensemble de ces paramètres et les note de façon rigoureuse. Cependant, la taille de nos échantillons, à savoir entre dix et vingt locuteurs par point d'enquête, ne permet pas d'étudier exhaustivement chacun de ces paramètres. PFC demande une répartition égale entre les sexes pour chaque point d'enquête, au moins deux classes d'âge, un minimum de scolarisation (vu la présence de lecture à haute voix) et, dans la mesure du possible, un certain échantillonnage des catégories socioprofessionnelles. En revanche, notre approche qui s'inspire fortement des travaux des Milroy (1987) en Grande-Bretagne consiste à travailler dans des réseaux denses (famille, amis, amis d'amis) afin d'obtenir des données aussi naturelles que possible. La dimension principale dans cette approche de la variation est le facteur géographique couplé à la différenciation stylistique.

La force de PFC réside dans le rassemblement de données comparables pour tous les points d'enquêtes grâce aux quatre registres requis dont deux de lecture et deux de dialogue. Tous les témoins lisent une liste de 94 mots se terminant par des paires minimales, un petit texte, participent à un entretien semi-directif (que nous appelons «guidé») et à une conversation non guidée ou «libre» dans la terminologie PFC. Dans l'entretien guidé le témoin est dans une situation d'interview où l'enquêteur n'est pas un proche mais un ami d'un ami (Milroy: 1987; Bourdieu: 1993) et idéalement, les enquêteurs évoluent par groupes de deux, où une personne seulement est un familier des témoins.

La personne étrangère au groupe est responsable de la conversation guidée alors que la conversation libre est du ressort du familier du groupe. Le recueil de données aussi naturelles que possible constitue l'objectif de la conversation libre même si le caractère toujours artificiel d'une telle situation (Gadet: 2000) est susceptible de fausser légèrement les données, ne serait ce que du fait même de la présence incontournable du magnétophone. Afin de minimiser le paradoxe de l'observateur, inévitable lors d'un enregistrement, deux ou trois personnes appartenant à un réseau dense sont mises en présence, et de préférence dans un lieu qui leur est familier. Dans cette situation de conversation libre, la présence de l'enquêteur n'est pas indispensable.

L'enquête PFC vise à livrer des résultats d'une grande généralité sur la présence ou l'absence de schwa. Les données que nous recueillons sont de deux ordres: les données sur schwa dans le texte lu et les données dans la parole spontanée (conversations guidée et libre). Nous procédons à des codages à partir de la graphie, mais sans nous rendre esclaves de cette dernière, car elle fournit des renseignements précieux sur les sites possibles déclenchant la présence ou l'absence d'un schwa. De tels codages ne constituent qu'une étape préliminaire de toute analyse mais cette étape nous semble indispensable. Il s'agit de balayer largement le corpus et de recueillir des données statistiques qui permettront quelques grandes généralisations et qui serviront d'assise à des analyses plus fines. Une fois le corpus constitué, nous serons en mesure de comparer les prononciations d'un très grand nombre de locuteurs sur des bases similaires. La mise au point d'utilitaires informatiques dédiés à des tâches de comptage et de classement de ces phénomènes permet d'automatiser et d'uniformiser les observations. Les variétés de français n'ont pas, à notre connaissance, été abordées avec des principes de codage aussi explicites que les nôtres et nous pouvons donc espérer obtenir des données d'envergure comparables.

Le codage est effectué sur la base de la transcription orthographique dans PRAAT sur une tire spécifique. Une fois le codage terminé, notre utilitaire *classeur-codeur* (cf. Durand, Eychenne & Meqqori: 2004) permet d'effectuer un comptage et un classement des présences/absences de schwa. Le système de codage de schwa a pris comme assise les traitements traditionnels, en particulier Dell (1985). Un exposé complet de la méthodologie PFC dépasserait le cadre du présent article: nous renvoyons donc à Durand, Laks & Lyche (2002) et Durand & Lyche (2003a: 244-251 et passim) pour des descriptions exhaustives des codages et une justification de la méthodologie. Le lecteur trouvera par ailleurs dans Eychenne & Hambye (2004) une synthèse des principaux problèmes liés au codage du schwa, ainsi que les solutions préconisées.

L'enquête PFC au Pays basque

L'enquête réalisée dans le cadre du projet PFC, ici comme ailleurs, a des ambitions nécessairement limitées. Elle pose des jalons pour des études plus approfondies et elle défriche le terrain là où les enquêtes sont absentes ou rarissimes. En ce qui concerne la phonologie du français au Pays basque, les publications font cruellement défaut et notre seul véritable point de comparaison est l'étude d'Henriette Walter (1977: 130-135, 1982: 192). Cette enquête, effectuée en 1975, portait sur un seul locuteur, pharmacien à Hasparren (Pyrénées-Atlantique) et âgé de 72 ans au moment des enregistrements. Néanmoins, à travers le portrait phonologique de ce témoin, se dessinent diverses tendances qui nous ont aidé à examiner notre corpus.

La constitution des corpus PFC vise à obtenir des données aussi naturelles que possibles et pour cela nous travaillons autant que faire se peut avec des «réseaux denses» de personnes parfaitement connues d’au moins un des enquêteurs (parents, amis, amis d’amis). Dans notre cas, le point d’ancrage était l’appartenance de l’un d’entre nous (MA) à la communauté basque et c’est dans sa famille et parmi ses amis que les enquêtés ont été choisis. 13 personnes ont été enregistrées du 2 au 5 mai 2002 (voir Annexe 1). JD ne connaissait pas les témoins avant l’enquête mais a été présenté comme un collègue de travail et ami de Michel Aurnague. Ce dernier a organisé les entretiens libres alors que JD était responsable des entretiens guidés. En ce qui concerne la représentativité sociale, nos témoins appartiennent à des couches moyennes ou moyennes supérieures. Un travail complémentaire fournissant un étalonnage social plus diversifié serait utile. En revanche, nous avons obtenu un bon équilibre hommes-femmes (8 femmes et 6 hommes) et une répartition très intéressante en ce qui concerne l’âge puisque nous avons enregistré, dans la même famille nucléaire, une grand-mère de 92 ans (Loc13), sa fille de 65 ans (Loc11), sa petite fille de 38 ans (Loc6) et son petit-fils de 39 ans (Loc8). (Tous les âges mentionnés dans cet article sont calculés par rapport à la date de l’enquête.) Cet échelonnement permet de faire des comparaisons intéressantes et de saisir des différences en temps apparent fortement suggestives de changements en temps réel. Une autre caractéristique intéressante de notre corpus est que les sujets comportent une bonne proportion de bascophones (6 sur 13 sont bilingues) et se répartissent sur deux endroits géographiques: St-Jean-Pied-de-Port et la communauté d’agglomération Bayonne-Anglet-Biarritz (voir Annexe).

St-Jean-Pied-de-Port (Donibane-Garazi en euskara) est une petite cité de 1650 habitants située sur le territoire de la Basse-Navarre (dont elle est la capitale) qui, avec le Labourd (Bayonne) et la Soule (Mauléon), constitue

l'une des trois provinces du Pays basque de France. Cette ville a été longtemps l'un des principaux passages vers la péninsule ibérique à travers les Pyrénées, et a fait partie intégrante du Royaume de Navarre, institué en 824. La Navarre comprenait cinq provinces juridico-militaires sur le versant sud et une au nord (la *Merindad de Ultra Puertos*) dont St-Jean-Pied-de-Port était la capitale. L'accession au trône de France en 1589 du Béarnais Henri IV entraîna le rattachement de la Basse-Navarre à la Couronne de France, rattachement qui eut lieu officiellement en 1620 sous Louis XIII et que le Roi d'Espagne reconnut solennellement lors du Traité des Pyrénées en 1659. Aujourd'hui, St-Jean-Pied-de-Port dépend, sur un plan géographique et administratif, du département des Pyrénées-Atlantiques (64) et de la région Aquitaine. Du point de vue linguistique, elle appartient à la zone du Pays basque où l'euskara a encore une forte présence. Les enregistrements de St-Jean-Pied-de-Port (plus précisément d'Uhart-Cize, village jouxtant St-Jean-Pied-de-Port et qui ne s'en différencie pas d'un point de vue linguistique) ont été effectués chez les enquêtés.

Les enregistrements faits à Biarritz (ville de 30.000 habitants) portent sur une population plus disparate. Biarritz constitue désormais un des trois maillons de la communauté d'agglomération (ex-district) Bayonne-Anglet-Biarritz (BAB, population: 95.500) qui appartient administrativement aux Pyrénées-Atlantique et à la région Aquitaine. Les habitants du (district) BAB ont le plus souvent des origines géographiques dispersées et c'est, on le devine, la zone où le basque est le plus menacé (voir § 2). Nos témoins, bien que fortement enracinés dans le Pays basque et parfois bilingues, reflètent cette diversité. Là aussi, les enregistrements ont été effectués chez les enquêtés.

2. Le contexte sociolinguistique au Pays basque

2.1. Le déclin des langues ‘régionales’

Les données relatives aux pratiques linguistiques contenues dans le recensement de l'INSEE de 1999 (INSEE 2002) montrent que les langues ‘régionales’ ont subi, depuis la fin du 20^e siècle jusqu’à nos jours, une érosion constante, érosion qui s’accélère après la seconde guerre mondiale (avant 1930: une personne sur quatre parlait une langue ‘régionale’, années 1950: une sur dix, années 1970: une sur vingt). De plus, lorsqu’elles sont transmises, ces langues sont utilisées occasionnellement (dans les échanges parents-enfants) alors qu’elles l’étaient de façon habituelle avant-guerre.

Le basque ou euskara n’échappe pas à cette évolution. L’enquête menée en 2001 par le Conseil de la langue basque et l’Institut culturel basque (dans le sillage des enquêtes de 1991 et 1996; CLB 2001 et Walter, 1994: 230-231 pour un résumé) montre que si 36 % des plus de 65 ans du Pays basque nord (France) sont bilingues, 12 % seulement le sont dans la tranche des 16-24 ans (voir le Tableau 2 plus loin). Si la proportion de bilingues diminue constamment dans cet intervalle, on constate, ici encore, une accélération pour les individus nés depuis le début des années 1950 (35-49 ans: 23 %, 25-34 ans: 12 % contre 30 % pour les 50-64 ans).

2.2. La situation du basque en 2001

Selon les résultats de cette même enquête et en considérant la population dans son ensemble (pas de découpage en tranches d’âge), le panorama sociolinguistique du Pays basque en 2001 était le suivant (Tableau 1): 24,7 %

de bilingues (54.700 personnes), 11,9 % de bilingues passifs (26.300) et 63,4 % de monolingues francophones (140.600) (plus de monolingues basques alors que l'on en comptait encore 0,7 % (1.500) en 1996). Cette perspective globale diffère cependant localement. Ainsi, alors que la communauté d'agglomération Bayonne-Anglet-Biarritz (BAB) présente le plus faible taux de bascophones (bilingues: 8,8 % (8.400), bilingues passifs: 8,3 % (7.900), monolingues francophones: 83 % (79.200)) et que l'intérieur des terres —Basse-Navarre (St-Jean-Pied-de-Port) et Soule (Mauléon)— fait apparaître une forte présence de la langue basque (bilingues: 60,9 % (19.500), bilingues passifs: 15,1 % (4.800), monolingues francophones: 24 % (7.700)), le Labourd intérieur —Cambo, Hasparren— est dans une situation intermédiaire (bilingues: 28,5 % (26.800), bilingues passifs: 14,5 % (13.600), monolingues francophones: 57,1 % (53.700)).

	Pays basque nord		BAB		Labourd intérieur		Basse-Navarre et Soule	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Total	221.600	100,0	95.500	100,0	94.100	100,0	32.000	100,0
Bilingues	54.700	24,7	8.400	8,8	26.800	28,5	19.500	60,9
Bilingues passifs	26.300	11,9	7.900	8,3	13.600	14,5	4.800	15,1
Monolingues français	140.600	63,4	79.200	83	53.700	57,1	7.700	24

Tableau 1. Compétences linguistiques par zones en 2001.

2.3. Une résistance toute relative

Bien que la baisse constante du nombre de bascophones au cours du 20^e siècle s'inscrive dans le cadre plus général du déclin des langues 'régionales' en France (cf. supra), les données du recensement 1999 indiquent que le basque résiste mieux à cette érosion que les autres langues, exception faite de

l'alsacien (qui apparaît en première position juste avant le basque). On retiendra que les adultes (de 1999) auxquels leurs parents parlaient basque à l'âge de 5 ans déclarent majoritairement que cette pratique avait un caractère habituel (pour toutes les autres langues 'régionales', excepté l'alsacien, cette pratique est majoritairement occasionnelle) mais aussi que la transmission de ce savoir (par ces mêmes adultes) à la génération suivante est meilleure dans le cas du basque (et de l'alsacien) que dans celui des autres langues. Alors que la chaîne de la transmission est rompue dans 58 % des cas pour le basque (47 % pour l'alsacien), cette proportion atteint 66 % pour le corse et plus de 80% pour l'occitan, le flamand et le breton (atteignant même 90 % dans le cas du franco-provençal).

Les données de l'enquête 2001 du Conseil de la langue basque concordent avec ces observations puisqu'elles font apparaître que 30 % des 50-64 ans sont bilingues contre 12 % chez les 25-34 ans soit une perte de 60 % (voir Tableau 2). Le taux de non transmission semble cependant diminuer puisque 23 % des 35-49 ans et 12 % des 16-24 ans sont bilingues, soit une perte de 48 %. De fait, et même si le début des années 1990 voit encore une forte régression des bilingues chez les jeunes (enquête 1991, 25-34 ans: 25 %, 16-24 ans: 20 %; enquête 1996, 25-34 ans: 14 %, 16-24 ans: 11 %), ce phénomène paraît être en voie de stabilisation au vu des données les plus récentes (enquête 2001, 25-34 ans: 12 %, 16-24 ans: 12 %). D'un point de vue 'qualitatif', il faut noter la part de plus en plus importante des bilingues équilibrés et de ceux pour lesquels le français est prioritaire par rapport aux bilingues ayant le basque comme langue prédominante (enquête 1996, basque prédominant: 32 %, équilibré: 33%, français prédominant: 35%; enquête 2001, basque prédominant: 22 %, équilibré: 37 %, français prédominant: 41%; voir Tableau 3). Selon les données de l'enquête 2001 (Tableau 4), le nombre de bilingues maîtrisant prioritairement le basque chute de façon considérable avec la

diminution de l'âge (≥ 65 ans: 34 %, 16-24 ans: 2 %), le pourcentage de bilingues équilibrés fléchissant lui aussi quoique (beaucoup) plus modérément (≥ 65 ans: 36 %, 16-24 ans: 32 %). La proportion de bilingues pour lesquels le français est prédominant augmente, dans le même temps, de manière spectaculaire (≥ 65 ans: 30%, 16-24 ans: 66%).³

	Tranches d'âge				
Bilingues	≥ 65	50-64	35-49	25-34	16-24
1991	34	38	32	25	20
1996	35	31	27	14	11
2001	36	30	23	12	12

Tableau 2. Bilingues par tranches d'âge (pourcentages).

	Types de bilingues		
	basque prédominant	équilibrés	français prédominant
1996	32	33	35
2001	22	37	41

Tableau 3. Typologie des bilingues (pourcentages).

	Tranches d'âge				
Types de bilingues	≥ 65	50-64	35-49	25-34	16-24
Prédominance basque	34	18	16	8	2
Equilibrés	36	48	32	31	32
Prédominance français	30	34	52	61	66

Tableau 4. Typologie des bilingues par tranches d'âge en 2001 (pourcentages).

³ Cette émergence d'un bilinguisme orienté vers le français est confirmée (plus localement) par le rapprochement de deux enquêtes réalisées à 20 ans d'intervalle (1982-2002) dans la vallée de St-Jean-Pied-de-Port (Aurnague J. et Duguine, à paraître).

2.4. Une langue en péril

Le ralentissement (provisoire?) du processus de recul des bascophones dans les jeunes générations —incessant depuis plusieurs décennies— ne signifie pas que la langue basque soit sauvée sur le territoire français. Il suffit pour cela de rappeler que le basque, comme la quasi-totalité des langues ‘régionales’ de France, fait partie des 3.000 langues en péril répertoriées par l’UNESCO (sur 6.000 langues). Le second ‘Atlas des langues en péril dans le monde’ publié fin 2001 par cette organisation (faisant suite à celui déjà établi en 1996; Wurm: 2001) considère en danger de disparition toute langue apprise par moins de 30 % des jeunes d’une population donnée. Comme on peut le constater, les 12 % de 16-24 ans maîtrisant le basque en 2001 sont bien en deçà de ce seuil et un effort considérable serait nécessaire pour dépasser cette limite fatidique. Mais quelles mesures adopter en termes de politique linguistique?

Le contexte dans lequel s’est effectuée la stabilisation observée chez les plus jeunes est, à ce titre, très instructif. Les conditions d’apprentissage et de diffusion de la langue basque ont assez nettement changé les vingt dernières années, que ce soit dans la sphère de l’enseignement (enseignement en immersion dans les écoles associatives (Ikastola), enseignement bilingue dans le réseau public (Ikas bi) ou privé (Euskal Haziak), cours du soir pour adultes...), dans la vie culturelle (rock basque, radios associatives en langue basque, télévision de la Communauté autonome basque...) ou même dans le domaine institutionnel (ex.: signalétique bilingue, nombreux documents officiels dans les deux langues). De fait, l’image du basque auprès des jeunes a très sensiblement évolué depuis le début des années 1980, l’apprentissage de cette langue et sa connaissance étant souvent perçus comme valorisants. Mais l’élément capital ici est, sans l’ombre d’un doute, l’enseignement, celui-

ci ayant pris, dans bien des cas, le relais de la famille dans le processus de transmission de la langue.

2.5. La place de l'enseignement

Le rôle que joue ce facteur (enseignement) pour la sauvegarde d'une langue est illustré de manière éloquent par la comparaison du taux de bilingues par tranches d'âge en Pays basque nord (France) d'une part et dans la Communauté autonome d'Euskadi (Espagne) d'autre part. Les données correspondant à l'année 2000 (voir le Tableau 5 plus loin) indiquent que, dans la Communauté autonome, le nombre de bilingues augmente de manière très importante chez les personnes de moins de 35 ans (70 ans: 29 %, 60 ans: 21 %, 45 ans: 21 %, 35 ans: 25 %, 25 ans: 33 %, 15 ans: 62 %) alors même qu'il diminue de façon constante sur le territoire français —et tout particulièrement parmi les individus de 45 ans ou moins— (70 ans: 37 %, 60 ans: 32 %, 45 ans: 27 %, 35 ans: 14 %, 25 ans: 11 %, 15 ans: 9 %). Le point d'inflexion (35 ans) suivi de l'accroissement spectaculaire des bilingues parmi les jeunes de la Communauté autonome coïncident avec l'introduction puis le développement du basque dans l'ensemble du cursus scolaire opérés à partir de la fin des années 1970. Le renversement de tendance est dû, en grande partie, à ce facteur.

Faire en sorte que la langue basque sorte de la phase d'extinction dans laquelle elle se trouve en France suppose donc que soit menée une politique volontariste en termes d'enseignement (en 2001, seulement 25 % des élèves de la maternelle et du primaire apprenaient le basque au sein des diverses filières). Une telle démarche doit avoir une dimension géographique et territoriale (possibilité d'accès à cet apprentissage dans l'ensemble des

établissements) mais doit aussi prendre en compte des aspects plus techniques et pédagogiques. Sur ce point —et à la suite des observations des psycholinguistes— il convient de rappeler que, dans un contexte social (activités extrascolaires, télévision, famille...) très majoritairement francophone, un enseignement strictement bilingue (parité horaire dès la maternelle) est souvent insuffisant pour atteindre un réel équilibre linguistique (individuel), l'enseignement en immersion (maternelle et CP en basque puis introduction progressive du français) parvenant tout juste à garantir une certaine harmonie (la grande majorité des enfants parlent français bien avant que cette langue ne soit intégrée à l'école et celle-ci doit donc agir comme élément de rééquilibrage). L'augmentation rapide des bilingues dont la langue prédominante est le français (ayant souvent une maîtrise orale et/ou écrite du basque lacunaire) —notée plus haut— résulte très probablement de la présence massive du français dans la vie courante, l'enseignement à parité du basque et du français (plus répandu aujourd'hui que l'apprentissage en immersion) n'étant pas susceptible de modifier cet état de fait. Ces remarques montrent, en même temps, que des actions en faveur de la langue basque doivent être réalisées dans d'autres domaines que celui de l'école.

Tranches d'âge	70 ans	60 ans	45 ans	35 ans	25ans	15 ans
Pays basque nord	37	32	27	14	11	09
Communauté autonome d'Euskadi	29	21	21	25	33	62

Tableau 5. Bascophones par classe d'âge et par territoire en 2000 (pourcentages).

2.6. La population et la politique linguistique

Quoi qu'il en soit, la demande d'un enseignement bilingue est forte chez les parents du Pays basque puisque, selon un sondage CSA, effectué en 2001 à

l'initiative de l'inspection d'Académie de Bordeaux (CSA 2001), 56 % des parents de la communauté d'agglomération Bayonne-Anglet-Biarritz (BAB) —zone où la présence du basque est la plus faible (voir plus haut)— souhaitent que leurs enfants puissent apprendre la langue basque à l'école (les parents de moins de trente ans (64 %) et ceux dont un enfant a déjà suivi ce type d'enseignement (95 %) étant les plus favorables à cette demande). Les motifs les plus souvent invoqués par les parents sont d'ordre culturel et pratique plutôt qu'idéologique ou professionnel. Par ailleurs, il mérite d'être noté que 54 % des personnes ayant répondu positivement ne sont pas nées en Pays basque.

Nous terminons ce bref panorama sociolinguistique du Pays basque —nettement centré sur l'euskara dans la mesure où toute la population est aujourd'hui francophone— en examinant l'attitude des habitants vis-à-vis des mesures de promotion du basque, telle que révélée par l'enquête 2001 du Conseil de la langue (Tableau 6). Si 42 % des habitants (dans leur ensemble) sont favorables ou très favorables à ces mesures et que 17 % y sont défavorables ou très défavorables —42 % (!) ne se prononçant pas—, d'importantes disparités locales se font jour. L'agglomération du BAB apparaît ainsi la plus en retrait par rapport à une telle politique (favorables ou très favorables: 24 %, défavorables ou très défavorables: 23 %, ni favorables ni défavorables: 52 %) alors que l'intérieur des terres —Basse-Navarre et Soule— y est massivement acquis (favorables ou très favorables: 65 %, défavorables ou très défavorables: 6 %, ni favorables ni défavorables: 28 %) et que le Labourd intérieur —Cambo, Hasparren— relève d'une situation médiane (favorables ou très favorables: 52 %, défavorables ou très défavorables: 13 %, ni favorables ni défavorables: 36 %). Soulignons que ces variations géographiques sont très exactement corrélées à la présence de la langue basque dans les zones concernées (voir plus haut).

	Pays basque nord	BAB	Labourd intérieur	Basse-Navarre et Soule
Très favorable	10	3	12	22
Favorable	32	21	40	43
Ni favorable, ni défavorable	42	52	36	28
Défavorable	15	21	11	6
Très défavorable	2	2	2	0

Tableau 6. Attitude face aux mesures de promotion de la langue basque par zones (pourcentages).

3. La phonologie du français au Pays basque

Au vu des données présentées ci-dessus, étudier la phonologie du français au Pays basque ne peut se faire sans envisager le statut de la langue basque puisque, du point de vue diachronique comme synchronique, les usages du français ne s'y inscrivent pas dans un vide linguistique mais sont en interaction, parfois conflictuelle, avec l'emploi du basque. Bien sûr, tous les Basques ne sont pas bilingues, comme le démontrent les statistiques en (2), mais beaucoup se sentent concernés par la question linguistique et savent à quel point la langue est un élément de leur identité. On le constate d'ailleurs du fait que les usages de nos enquêtés présentent des spécificités lexicales, morphosyntaxiques ou pragmatiques qui marquent leur enracinement dans la communauté basque. Ainsi en va-t-il des liens de famille régulièrement exprimés par des noms empruntés au basque (*ama* (mère), *aita* (père), *amatxi* (grand-mère), *otto* (oncle)...), du vocable *adio* utilisé pour saluer un individu (en le rencontrant ou en le quittant) ou de l'adjectif *pareil* employé, comme son équivalent basque *berdin*, dans des tournures adverbiales indiquant le consentement ou le remerciement (en réponse à une offre/proposition telle que *voulez-vous du café?*).

En faisant ces remarques, nous ne voudrions pas tomber dans le piège du «substrat» utilisé comme explication systématique des différences entre aires

linguistiques. En effet, le français est employé de longue date au Pays basque et s'est en grande partie autonomisé par rapport au basque et il n'est pas impossible que le français ait eu des effets de retour sur le basque dans certains domaines (voir la discussion sur le 'r' plus bas). Il faut donc, à la fois, tenir compte des relations entre le français et le basque, car dans toute communauté multilingues des interférences se produisent entre les systèmes en présence, mais ne pas abuser des explications attribuant toutes les spécificités observés à la langue originelle de la communauté étudiée.

Ce que nous présentons ci-dessous n'est pas une analyse finale de la phonologie de nos locuteurs mais plutôt ce qu'on pourrait appeler des extraits de notre carnet d'enquête. Ils se rapportent surtout aux enquêtés de St-Jean-Pied-de-Port. Les analyses en cours devraient permettre de préciser, de valider ou d'infirmer un certain nombre des affirmations exprimées ici.

3.1. Les voyelles orales

Le système des voyelles orales de nos témoins se caractérise par un inventaire typique des accents du sud de l'Hexagone, à savoir que l'on n'y connaît ni l'opposition /a/ ~ /ɑ/ (*patte-pâte*), ni les oppositions d'aperture moyenne /e/ ~ /ɛ/ (*épée-épais*), /ø/ ~ /œ/ (*jeûne-jeune*) et /o/ ~ /ɔ/ (*côte-cote*). L'absence de ces oppositions se manifeste même dans la lecture à haute voix de la liste PFC et du texte, même si certains témoins esquissent dans la lecture de la liste des oppositions qu'on ne retrouve pas dans la lecture du texte ou la conversation spontanée. En dehors du schwa (cf. section 4), le système ne comporte donc que sept phonèmes vocaliques que l'on peut représenter comme suit:

(1) Voyales orales phonémiques

i	y	u
E	Ø	O
	a	

Nous avons choisi /E, Ø, O/ pour symboliser les voyelles moyennes dans la mesure où leur valeur en surface oscille le plus souvent entre [e] et [ɛ] (*fait-faite*), [ø] et [œ] (*creux-creuse*), [o] et [ɔ] (*chaud-chaude*). Cette différence entre deux principales valeurs phonétiques peut s'analyser soit en termes d'une opposition graduelle du type mi-haut [e, ø, o] vs. mi-bas [ɛ, œ, ɔ], comme dans les systèmes de classification héritées de l'Association Phonétique Internationale, soit à l'aide d'un trait du type [+tendu] ([e, ø, o]) vs. [-tendu] ([ɛ, œ, ɔ]) comme dans les études s'inspirant de la phonologie générative (voir Durand et Lyche (2003a), Lyche (2003b)). Le facteur principal qui détermine cette alternance phonétique est la loi dite de position: voyelle mi-fermée en syllabe ouverte et voyelle mi-ouverte en syllabe fermée. Cependant, ce mécanisme d'ajustement est un peu plus complexe que ne l'admettent la plupart des descriptions. En effet, un schwa dans la syllabe qui suit une syllabe contenant une voyelle moyenne (par exemple, *bêtement* [betəmã]) n'attire pas la consonne qui précède en attaque syllabique. En effet, si la division syllabique était [bɛ.tə.mã] on devrait avoir une voyelle mi-fermée dans la première syllabe qui serait alors ouverte (*[be.tə.mã]), ce qui n'est pas le cas. Ce mot, pris comme représentant d'un vaste ensemble de cas, semble donc avoir la structure syllabique [bet.ə.mã] ou alors une structure avec la consonne /t/ en position ambisyllabique (voir Durand 1976, 1990, 1995). Mais si on laisse de côté cette question épineuse, on peut dire que cet accent, comme les autres accents du midi et de nombreux autres

accents français, favorise la loi de position (au moins pour les locuteurs les plus jeunes):

(2) Loi de position

Les voyelles tendues (ou mi-hautes) apparaissent en syllabe ouverte et les voyelles relâchées (ou mi-basses) en syllabe ouverte.

Cette tendance, qui serait universelle selon certains spécialistes, existe bien sûr en français standard:

(3) Alternances tendu en syllabe ouverte vs. relâché en syllabe fermée

(i) [e] syllabe ouverte inaccentuée vs. [ɛ] syllabe fermée accentuée

[e] – [ɛ]: *céder* – *cède*, *lépreux* – *lèpre*, *agrégation* – *agreg*

(ii) [e] syllabe ouverte accentuée vs. [ɛ] syllabe fermée accentuée

[e] – [ɛ]: *léger* – *légère*, *fermier* – *fermière*

(iii) [o] syllabe ouverte accentuée vs. [ɔ] syllabe fermée accentuée

[o] – [ɔ]: *galop* – *galope*, *sot* – *sotte*, *idiot* – *idiote*

(iv) [ø] syllabe ouverte accentuée vs. [œ] syllabe fermée accentuée

[ø] – [œ]: *peut* – *peuvent*, *veut* – *veulent*, *œufs* – *œuf*

Cependant, dans le français dit de référence, elle y est concurrencée par de nombreuses asymétries, neutralisations, et la présence d'oppositions lexicales du type *épée-épais*, *jeune-jeûne*, *cote-côte*. Dans notre propre enquête, la loi de position semble avoir une forte applicabilité:

(4) Alternances tendu en syllabe ouverte vs. relâché en syllabe fermée

(i) [e] *inquiet*, *tournée*, *tournaît*, *pêcheur*, *pêcheur* vs. [ɛ] *colère*, *infect*, *préfecture*

(ii) [o] *au*, *beauté*, *botté*, *rhinocéros* vs. [ɔ] *roc*, *rauque*, *rhinocéros*, *cause*

(iii) [ø] *jeux*, *jeunet*, *creux* vs. [œ] *jeune*, *jeûne*, *creuse*

Historiquement, la présence d'un inventaire phonémique réduit (sept voyelles distinctives) au Pays basque a sans doute été favorisée par le fait que la plupart des variétés de l'euskara moderne ne comportent que cinq voyelles phonémiques tout comme le castillan: à savoir /i e a o u/ (Michelena 1977; Trask 1997: ch. 3). Mais l'appartenance géographique et culturelle à l'aire dialectale des pays de langue d'oc est peut-être encore plus déterminante en ce qui concerne le système vocalique.

3.2. Les voyelles nasales

Walter notait qu'elles étaient au nombre de quatre: /ẽ/, /œ̃/, /ã/, /õ/ et faisait remarquer: «Les réalisations de /ẽ/ sont assez fermées, et assez proches de [ẽ], celles de /ã/ très avancées, [ã]. L'opposition /ẽ/ ~ /œ̃/ se maintient très bien. On trouve quelques réalisations de consonnes nasales après la voyelle nasale, mais ce n'est pas régulier: *un peu*, [œ̃^mpø]». Nos locuteurs distinguent effectivement *brin* de *brun* et *blanc* de *blond*, comme dans les autres enquêtes réalisées dans le sud de l'Hexagone. Il semble également vrai qu'on n'a pas en règle générale un appendice consonantique qui caractérise l'accent plus conservateur du Languedoc-Roussillon décrit par Durand et Tarrier (2003). La qualité auditive de /ẽ/ ne semble pas aussi fermée que pour le témoin de Walter; le /õ/ semble plutôt un /õ/, et le /ã/ n'atteint pas la réalisation [ã] qui caractérise de nombreux accents méridionaux. Une étude acoustique semble requise pour mieux déterminer la nature des réalisations. Par ailleurs, seul un examen attentif de diverses alternances (liaison, morphologie flexionnelle et dérivationnelle) peut permettre de comprendre le statut des voyelles dites nasales, comme le démontre Durand (1988).

3.3. Les semi-voyelles

Les trois semi-voyelles [j], [w], [ɥ] sont présentes dans le système comme c'est habituellement le cas en français hexagonal. Leur statut sous-jacent est extrêmement complexe et ne peut se déterminer qu'après examen des patrons distributionnels et de nombreuses alternances (Durand et Lyche 1999). A un niveau superficiel, nous avons noté, comme Walter, mais surtout chez les locuteurs les plus âgés, un grand nombre de diérèses: *relier* [rølije], *niais* [nije], *nier* [nije] et, comme en français standard, *prendrions* [prãdrijõ], *prendriez* [prãdrije], *quatrième* [katrijẽm]. Le système chez les locuteurs plus jeunes semble en évolution vers des synérèses (*relier* [rølje], *niais* [nje], *nier* [nje]) comme dans le français de référence.

3.4. Les consonnes

Henriette Walter ordonne les phonèmes consonantiques de son témoin dans le tableau suivant qui constitue un point de départ très utile pour nos observations. (Nous laisserons cependant de côté [j] mentionné ci-dessus.)

Sourdes:	p	f	t	s	ʃ	k
Sonores:	b	v	d	z	ʒ	g
Nasales:	m		n		ɲ	ŋ
Latérales:			l		ʎ	
Vibrantes:			r			ʁ
«Semi-voyelle»:					j	
						h

En ce qui concerne, les obstruantes (plosives et fricatives) la corrélation de sonorité se maintient bien. Néanmoins, nous avons observé de nombreux cas d'assimilation de voisement affectant les fricatives, y compris dans la lecture à haute voix de la liste de mots et même en position initiale: *socialisme* [sosjalizm], *islamique* [izlamik] et *slip* [zlip]. Nous n'avons pas examiné la nature phonétique des réalisations du /s/ - /z/ mais étant donné la présence en basque d'oppositions entre deux types de sifflantes coronales, la réalité phonétique et sociolinguistique mérite sans aucun doute un examen plus approfondi.

Dans l'ensemble des nasales, /m/ et /n/ sont stables de manière prévisible. Comme en français de référence, /ŋ/ a un statut marginal mais clair (*parking, jogging*). La différence possible entre /ɲ/ et /nj/ reste à examiner. En revanche, pour les latérales, nous avons pu observer l'emploi d'une palatale chez deux de nos témoins (Loc13, 92 ans) et sa fille (Loc11, 65 ans) avec l'opposition *étriller* /etriʎe/ - *étrier* /etrije/ dans notre liste de mots. Hors enregistrements, nous avons pu entendre des palatales chez ces locuteurs dans des mots comme *émailler, éveiller* ou *sourciller*. Loc11, qui était enseignante, a même signalé que les prononciations du type /etriʎe/ étaient intégrées à son enseignement (par exemple en dictée), ce qui démontre que des normes régionales ou personnelles se substituent plus souvent qu'on ne le croit à un français de référence bien mal connu en dehors des manuels orthoépiques. Les petits-enfants de la même famille (Loc6 38 ans et Loc8 39 ans) ne font pas cette opposition et, ici comme dans d'autres variétés du français, la latérale palatale est en voie de disparition.

Comme dans l'enquête Walter, une fricative glottale /h/ est également attestée pour l'informatrice la plus âgée (Loc13) qui prononce la séquence *le hasard* comme [lø hazar] dans la lecture du texte. Nous n'avons aucun autre

exemple dans la conversation enregistrée mais ce phonème fait bien partie du système de Loc13 qui l'utilise dans *héron*, *hêtre* ou *héros*.

L'inventaire phonémique du français du Pays basque ayant été exposé, nous nous tournons vers l'examen de deux segments particulièrement intéressants dans cette variété, à savoir le «e caduc» (section 4) et le <r> (section 5), avant de conclure notre présentation

4. La question du «e caduc»

L'enquête PFC au Pays basque met au jour une variété du Midi en évolution par rapport à des accents du Midi plus traditionnels (cf. Durand, Slater & Wise (1987); Durand & Eychenne, sous presse): bien que présentant un système vocalique «typique» du Midi (cf. 3), cette variété présente un comportement pour le schwa qui tend à se rapprocher d'un accent plus standard. Nous disposons pour ce corpus de 6511 codages pour le schwa.⁴

Comparons tout d'abord nos résultats avec ceux de Walter (1977), qui disait ceci de son témoin: «On constate une forte proportion de *e* prononcés dans les monosyllabes (88 % de réalisations, dans *je*, *de*, etc.) et dans les premières syllabes de polysyllabes (86 % dans *reconnaissons*, etc.) tandis qu'il y a seulement 41 % de réalisations de polysyllabes (*dire*, *province*, *encore*, etc.) et 42 % dans les syllabes de polysyllabes autres que la première et la dernière (*appelons*, *arriverons*)». Nous donnons en (6) un tableau récapitulant les résultats de Walter et les nôtres (en %):

⁴ Signalons que Loc10 n'a pas été codé pour les phénomènes du schwa et de la liaison, en raison d'un problème dans l'enregistrement.

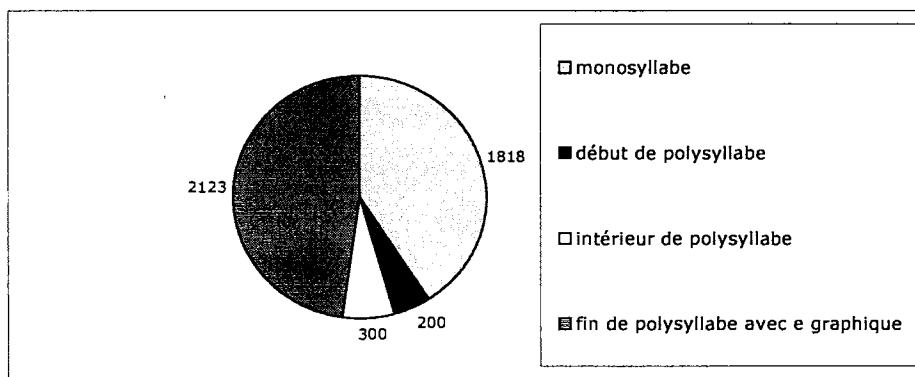
(6) résultats comparés de Walter (1977) et PFC

	Walter (1977)	PFC		
		global	lecture	conversation
Monosyllabes	88	86,4	97,6	80,8
Début de polysyllabe	86	86,5	94,1	83,9
Intérieur de polysyllabe	42	51	76,4	45,3
Fin de polysyllabe ⁵	41	47,1	54,1	43,9

Si l'on compare les résultats de Walter à nos résultats globaux, on remarque qu'ils coïncident assez bien, et l'on peut dégager deux grandes tendances dans les 2 enquêtes: une tendance au maintien de schwa, dans les monosyllabes (Walter: 88 %; PFC: 86,4 %) et en début de polysyllabe (Walter: 86 %; PFC: 86,5 %), et une tendance à l'effacement en syllabe interne (Walter: 42 %; PFC: 51 %) ou finale de polysyllabe (Walter: 41 %; PFC: 47,1 %). Cependant, si l'on considère les colonnes «lecture» et «conversation», qui représentent l'éclatement de la colonne «global» de PFC, on remarque, ce qui n'est pas surprenant, que la lecture favorise grandement l'apparition de schwa, surtout pour les monosyllabes (97,6 %) et le début de polysyllabe (94,1 %), où il se réalise de manière presque obligatoire. La position interne de polysyllabe est celle qui illustre le mieux l'asymétrie lecture/conversation, puisque le schwa, alors qu'il se réalise à 45,3 % en conversation, est prononcé à 76,4 % en lecture. Néanmoins, si l'on compare les résultats PFC pour la conversation aux résultats que rapporte Walter pour son témoin, on constate que les chiffres restent étonnamment proches. Nous tenons cependant à nuancer quelque peu ce point, dans la mesure où les différentes positions sont représentées de manière inégale. La figure (7) illustre la représentation des différents contextes en nombre d'occurrences:

⁵ Contexte <XCe# {C; #}> uniquement, pour rendre les résultats comparables à ceux de Walter.

(7) nombre d'occurrences par contexte codé



Il apparaît que la position finale de polysyllabe avec *e* graphique (2.123 occurrences) et les monosyllabes (1.818 occurrences) rassemblent l'écrasante majorité des codages, alors que le début (200) et l'intérieur (300) de polysyllabe sont comparativement peu représentés, puisque tous les contextes n'ont pas la même représentativité.

Intéressons-nous maintenant au schwa en position finale (tous contextes confondus), et considérons les 4.191 occurrences que délivre notre corpus pour cette position. Si nous regroupons toutes les occurrences de fin de polysyllabe en contexte prépausal ou préconsonantique, nous obtenons la répartition rapportée en (8):

(8) corrélation *e* graphique/schwa en contexte XC_#{C ; #}

e graphique

voyelle	occurrences	pourcentage (%)
absente	1.048	48,2
présente	1.023	47,1
incertaine	103	4,7
total	2.174	100

pas de e graphique

voyelle	occurrences	pourcentage (%)
absente	998	91,5
présente	74	6,8
incertaine	19	1,7
total	1.091	100

Dans ce contexte, il apparaît que, lorsqu'une consonne finale prononcée précède un *e* dans la graphie, elle est suivie d'un schwa dans 47,1 % des cas. En revanche, si la consonne n'est pas suivie d'un *e* dans la graphie, le schwa n'est réalisé que dans 6,8 % des cas. Ainsi, le taux de réalisation sur l'ensemble du corpus est presque sept fois supérieur s'il y a un *e* graphique, ce qui nous semble hautement significatif. Nous interprétons cette asymétrie comme la preuve du fait qu'il existe encore, dans cette variété, une opposition phonologique de type /C#/ (type *net* /nEt/ = [nɛt]) vs /Cə#/ (type *nette* /nEtə/ = [nɛtə]). Ceci étant dit, on remarque qu'environ une fois sur deux, un *e* graphique ne correspond à aucune manifestation phonétique dans la chaîne parlée. Il est ici nécessaire de passer à un niveau de détail plus fin: si ces chiffres globaux nous donnent bien une idée du comportement global du schwa dans notre corpus, ils ne nous renseignent pas sur le système intégré par chacun des locuteurs. Ainsi, plusieurs pistes s'offrent *a priori* à nous pour expliquer ce taux d'effacement important: soit le corpus est relativement

homogène, et l'on s'attend à ce que les locuteurs aient des taux d'effacement très proches (dans ce cas, on peut supposer qu'ils maintiennent tous une opposition phonologique, et il reste à élucider les facteurs favorisant l'effacement de la voyelle); soit le corpus est hétérogène, avec d'un côté des locuteurs conservateurs maintenant le schwa et de l'autre des locuteurs «innovateurs» ayant perdu l'opposition /C#/ vs /Cə#/. Nous donnons ci-dessous en (9), pour les mots de type XVC_# et toutes tâches confondues, le taux de réalisation de schwa en contexte prépausal et préconsonantique.

(9) Taux de réalisation d'un schwa correspondant à un *e* graphique pour chaque locuteur en contexte VC_#{C;#}

locuteur	e graphiques codés	schwas réalisés	réalisation (%)
Loc1	149	40	26,8
Loc2	130	67	51,5
Loc3	165	36	21,8
Loc4	95	46	48,4
Loc5	131	47	35,9
Loc6	126	48	38,1
Loc7	164	52	31,7
Loc8	148	59	39,9
Loc9	138	86	62,3
Loc11	103	41	39,8
Loc12	195	96	49,2
Loc13	115	67	58,2

Les données traduisent une forte variabilité inter-individuelle (coefficient de variation c.v. = 29,3 %), avec une étendue allant de 21% à 62,3 % de réalisation. Cette variété est donc peu homogène quant à la réalisation du schwa: les locuteurs qui présentent le système le plus conservateur (Loc9 et Loc13) ont un taux d'effacement proche de 40 % dans le contexte étudié, ce qui montre l'érosion du schwa final dans cette variété.

En revanche, les locuteurs ayant le système le plus innovateur (Loc3 et Loc1) maintiennent toujours un schwa dans plus de 20 % des cas. Nous interprétons ces chiffres comme la traduction d'une évolution lente d'un système du Midi traditionnel vers un système plus proche de la norme. Il nous est cependant difficile d'affirmer catégoriquement que les locuteurs innovateurs ont complètement perdu l'opposition /C#/ vs /Cə#/. Par exemple, alors qu'il réalise 21,8 % de schwas correspondant à un *e* graphique dans le contexte VC_#{C;#}, le locuteur Loc3 ne réalise que 2,7 % de schwas (soit 4 réalisations pour 146 codages) dans ce même contexte lorsqu'il n'y a pas de *e* graphique. Or, si le locuteur ne disposait plus que d'une seule classe /Cə#/, on devrait s'attendre à ce que les 2 classes s'équilibrent. Les occurrences de schwa pour les mots du type /C#/ sont par ailleurs limitées au contexte prépausal en conversation (*milieu rural* ə//; *et alors* ə//; *cinq jours* ə//; à *Biarritz* ə//), et l'on ne doit pas exclure la possibilité que ce schwa soit comparable au schwa prépausal du français standard (cf. Berrit Hansen 2003).

A l'appui de l'affirmation selon laquelle cette variété serait en évolution d'un système du Midi traditionnel vers un système plus «standard», nous donnons en (9) la corrélation taux de réalisation/âge du locuteur. Nous avons classé les locuteurs par génération, en prenant pour point de référence le locuteur Loc8, co-enquêteur avec JD:

- génération 1: Loc13 (grand-mère de Loc8),
- génération 2: Loc11 (fille de Loc13, mère de Loc8) et Loc12 (oncle de Loc8),
- génération 3: Loc8, Loc6 (soeur de Loc8), et tous les autres locuteurs (connaissances de Loc8).

(10) corrélation taux de réalisation du schwa en contexte VC_#\{C;\#} / âge du locuteur (au moment de l'enquête)

locuteur	âge	génération	réalisation (%)
Loc9	60	2	62,3
Loc13	92	1	58,2
Loc2	30	3	51,5
Loc12	73	2	49,2
Loc4	34	3	48,4
Loc8	39	3	39,9
Loc11	65	2	39,8
Loc6	38	3	38,1
Loc5	34	3	35,9
Loc7	37	3	31,7
Loc1	25	3	26,8
Loc3	31	3	21,8

Idéalement, on devrait s'attendre à ce que les locuteurs les plus âgés (génération 1 et 2) soient ceux qui maintiennent le plus grand nombre de schwas, et qu'inversement les jeunes (génération 3) soient ceux qui l'effacent le plus. Une telle répartition irait dans le sens d'une évolution diachronique de cette variété.

Les données présentées en (10) nous semblent globalement conforter cette hypothèse: ainsi, les locuteurs Loc9 et Loc13, tous deux «sénior», sont ceux qui maintiennent le plus le schwa, alors que les locuteurs Loc7, Loc1 et Loc3, parmi les plus jeunes, sont aussi ceux qui affichent une nette tendance à l'effacement. La situation est néanmoins complexe, et certains locuteurs semblent contredire cette tendance. Considérons par exemple le réseau familial de Loc8 qui, comme nous l'avons dit plus haut, représente les 3 générations. La locutrice la plus âgée, Loc13 (la grand-mère de Loc8), a un

taux de réalisation relativement élevé (58,2 %). Le locuteur Loc12 (l'oncle de Loc8), a un taux de réalisation nettement moindre (49,2 %), mais Loc11 (la mère de Loc8), qui est de la même génération que Loc12, et pour laquelle on pourrait par conséquent espérer un taux de maintien voisin, a un taux de réalisation (39,8 %) presque identique à celui de ses enfants Loc6 (38,1 %) et Loc8 (39,9 %), lesquels représentent la troisième génération. Par ailleurs, au sein de la troisième génération, on observe des variations très importantes, allant de 21,8 % (Loc3) à 51,5 % (Loc2).

Notre interprétation de ces chiffres est que le système est réellement en évolution, mais que cette évolution est disparate: les jeunes locuteurs *tendent* à adopter un système plus standard, mais ils ne le font peut-être pas au même rythme et au même degré. Un élargissement de la base empirique, en particulier pour les générations 1 et 2, pourrait permettre de mieux cerner le rôle de l'âge dans l'évolution du français au Pays basque.

Intéressons-nous maintenant aux groupes consonantiques complexes du type XCC(ə)#C. Bien qu'un schwa se réalise dans la grande majorité des cas (245 réalisations pour 351 occurrences, soit 72,4 %), on relève 18 occurrences de non réalisations, très atypiques pour une variété du Midi, comme les exemples reproduits en (11):

(11) non réalisation d'un schwa dans le contexte CC_#C

D'un autr' côté,
dix kilomètr's de
Premier Ministr' ne cesse (x2)
mes oncl's qui ont travaillé
Premier Ministr' lassé
par contrr' moi
il faut êtr' quand même
notr' liberté

peut-êtr' moins
Je parl' basque,
je parl' moi
parl' bien (x3)
l'agrégation extern' d'économie-gestion
rapport' quand même
l'accent circonflex' par exemple

En ce qui concerne le statut «épenthétique» du schwa (i.e. des réalisations vocaliques ne correspondant à aucun *e* graphique), on en relève 9 occurrences dans le corpus (6 en lecture et 3 en conversation):

(12) réalisation d'un schwa dans le contexte CC_#C

Marc ə Blanc (x3)
Ouest ə France (x3)
film ə très bien (x2)
concept ə de

Ces exemples démontrent que cette variété est clairement en évolution par rapport à des accents plus conservateurs. Il faut aussi se demander, et ce pour chaque locuteur, si un schwa est toujours présent dans des mots du type *quatorze*, *ministre*, ou si une restructuration plus poussée a eu lieu permettant des codas consonantiques complexes comme /str/ à la finale de mot (ce qu'interdisent divers cadres théoriques). L'apparition d'un schwa se réduirait alors à une fonction de «lubrifiant phonétique» dans des jonctures complexes. Une étude de la corrélation statut lexical vs épenthétique du schwa dépasserait le cadre et la portée du présent article. Les outils en cours de développement, qui permettront une exploration plus fine des données, ainsi

que les enquêtes en voie d'achèvement, devraient y apporter des éléments de réponse.

5. La question du 'r': synchronie et diachronie

En ce qui concerne le 'r' nos locuteurs, comme la plupart des francophones actuellement, n'utilisent qu'un seul son distinctif /r/. En le transcrivant /r/, nous employons le symbole 'r' dans une valeur large sans nous préoccuper de sa réalisation phonétique qui est, nous le constaterons, extrêmement variable et soulève des questions épineuses du point de vue synchronique et diachronique.

Une discussion précise de la famille des sons dits 'rhotiques' (du grec 'rho' pour 'r') ne peut se faire sans une bonne connaissance des symboles de l'API (cf. le *Handbook of the International Phonetic Association*, 1999). Rappelons la valeur des symboles utilisés plus bas tels qu'ils sont définis dans la tradition phonétique.

(13) Symboles phonétiques utilisés ci-dessous

[r]: vibrante apicale alvéolaire ou dentale voisée (souvent décrit comme 'r' roulé dans l'usage commun)

[ɾ]: battue apicale alvéolaire ou dentale voisée

[ʀ]: fricative ou approximante uvulaire voisée

[χ]: fricative uvulaire non-voisée (également [ɣ] où le diacritique sous le [ɣ] indique un dévoisement)

[ʁ]: vibrante uvulaire voisée

[ʀ̥]: vibrante uvulaire non-voisée

[x]: fricative vélaire non-voisée

En étudiant la question du 'r', nous n'avons pas observé une opposition entre deux types de 'r' comme celle signalée par Walter qui déclare: «La situation est beaucoup plus tranchée pour l'opposition /r/ - /ʁ/, qui est très nette et parfaitement constante à l'intervocalique: *Paris, parisien, marin, périr*, sont constamment réalisés avec une apicale roulée [r], *arriver, parrain, terrible* avec une fricative uvulaire [ʁ], souvent réalisée avec une très forte friction et sans vibration des cordes vocales à l'initiale: *il n'a pas ri*, que l'on pourrait presque transcrire avec un [x], [inalpaxi]. En finale absolue, on a relevé quelques rares cas de [ʁ] dévoisés dans des mots comme *tard* ou *alors*, mais ce n'est pas systématique».

Le locuteur d'Hasparren que décrit Walter reflète sans doute un usage plus proche d'un substrat basque. En effet, il est probable, du point de vue historique, que toutes les variétés actuelles du basque procèdent d'un système caractérisé par une opposition entre une vibrante apicale /r/ et une battue apicale /ʀ/ comme celle qu'on trouve en castillan moderne (Trask 1997: 144). Cette opposition se maintient de nos jours dans les variétés espagnoles du basque qui opposent une vibrante [r] et une battue [ʀ]: *erre* 'brûler' et *ere* 'aussi' ou *zorri* 'poux' et *zori* 'chance'. Dans la partie bascophone française, il semblerait que la vibrante apicale soit habituellement passée à une dorsale (fricative uvulaire) et il est envisageable que ce changement soit le résultat d'une influence du français sur le basque, hypothèse que nous discutons plus bas. Selon Trask, cette réalisation uvulaire serait voisée [ʁ] en labourdin et bas-navarrais. Dans la variété de basque parlée par nos locuteurs, la fricative uvulaire est en fait normalement dévoisée [χ] ou [ɣ] et peut-être vélaire [x] dans certains cas. Ce n'est pas sans incidence sur le système français qui s'est mis en place.

En ce qui concerne la qualité du phonème /r/ utilisé par nos locuteurs en français, *tous* nos sujets ont une réalisation postérieure (uvulaire) à l'exception de Loc13 (92 ans) qui présente des variations particulièrement intéressantes entre [r] et [χ]. Dans la lecture des chiffres qui précèdent les mots de notre liste, elle utilise plus ou moins aléatoirement des allophones uvulaires ou apicaux. Soit les exemples suivants: *trois* [tχwa], *treize* [tχezə], *quatorze* [katəχzə] face à *trente* [trāt], *trente-deux* [trātdø], *quarante-cinq* [karātsēk], *quatre-vingt trois* [katrəvētχwa], et, pour couronner le tout, *quatre-vingt quatre* qui est prononcé: [katrəvētkatχ]. En dehors de la lecture des chiffres, la forme dominante chez Loc13 dans la liste de mots est le [r] apical vibrant: en effet, sur 34 occurrences du phonème /r/, il y a 33 [r] apicaux avec des réalisations fortes en début de mot: *roc* [r:ək], *ras* [r:a], etc. Le seul «r» uvulaire est utilisé dans l'item 55: *étrier* [etχije] que l'on comparera sur ce plan à 75: *étriller* [etriʎe] avec un [r] alvéolaire vibrant et une latérale palatale [ʎ]. Dans la lecture du texte, tous les «r» sont des vibrantes apicales à l'exception du mot *ministre* qui est lui-même sujet à variation puisqu'il est successivement prononcé [ministr] et [ministχ]! Dans la conversation guidée pour Loc13, nous n'avons qu'un seul exemple de «r» uvulaire et aucun dans la conversation libre. Cela suggère que la réalisation dominante du /r/ dans le vernaculaire de Loc13 est une apicale vibrante. La fréquente utilisation du 'r' uvulaire dans la lecture des chiffres qui sont censés provoquer une prononciation moins surveillée et donc plus proche de la conversation spontanée reste pour nous un mystère. Nous avons à diverses reprises spéculé sur une explication possible liée à l'activité d'institutrice et de secrétaire de mairie exercée par Loc13 et au rôle que les chiffres pouvaient jouer dans ses activités mais nous ne sommes pas en mesure de fournir une explication convaincante. Ce qui semble certain c'est que l'adoption d'une variante uvulaire est probablement très ancienne dans cette zone dialectale:

Loc12, qui a 73 ans, n'emploie que cette dernière en parlant français et fournit la ligne de démarcation pour nos locuteurs puisque pas un seul d'entre eux n'emploie le 'r' dit roulé. Cela suggère, au moins pour ce qui est du français, que le 'r' uvulaire était déjà en place bien avant la deuxième guerre mondiale.

En revanche, ce qui est caractéristique de plusieurs de nos locuteurs est l'emploi d'allophones non-voisés, non seulement en début et fin de mot (ce qui pourrait s'expliquer à partir de tendances générales observées en phonétique), mais également en position intervocalique. Si l'on compare le mot *vérifications* dans le texte PFC on entend: [verifikasjõ] pour Loc13 (92 ans), [veɣifikasjõ] pour sa fille (Loc11) et [veɣifikasjõ] pour sa petite-fille (Loc6). La transition entre dévoisé et voisé n'est pas abrupte entre Loc11 et Loc13 mais on constate une montée des variantes voisées au fil des générations, ce qui rapproche progressivement les usages de la norme parisienne.

Si l'on nous permet quelques spéculations dans cette dernière partie, la transition historique probable de variantes apicales vers des variantes postérieures dans la prononciation du français au Pays basque est sans doute ancienne et liée à l'évolution de la norme en France ainsi qu'à celle de la variété de basque local. On a pu émettre en sociolinguistique l'hypothèse que la dissémination d'un 'r' uvulaire en Europe était due au rayonnement de la norme parisienne. Chambers et Trudgill (1980: ch. 11) posent que les langues européennes ne possédaient primitivement que des consonnes 'r' de type apical (vibrante [r] ou battue [ɾ]). Après s'être répandu à Paris au cours du XVII^e siècle, le r uvulaire ([ʀ] ou [ʁ]) serait graduellement passé dans les zones limitrophes (Belgique, Allemagne, Suisse) pour arriver à Copenhague vers 1780, d'où il serait passé en Suède du sud où il est stationnaire depuis environ 1930 (voir les essais dans Van de Velde et van Hout 2001 pour un

examen détaillé de cette hypothèse). Cette explication exogène, si elle est correcte, pourrait alors s'appliquer à la langue basque, les variétés du nord des Pyrénées se révélant plus perméables à l'influence parisienne que les variétés du sud. Mais, à notre sens, une explication phonologique endogène semble au moins indispensable pour comprendre pourquoi ce sont les vibrantes qui sont passées à une réalisation uvulaire et non les battues.

Dans diverses langues qui présentent l'opposition apicale vibrante/battue, on peut constater qu'elle est neutralisée dans divers contextes. Ainsi, le castillan ne maintient cette opposition qu'à l'intervocalique et un 'r' à l'initiale du mot est systématiquement prononcé comme une vibrante forte. On a pu prétendre que c'était vrai du basque. Trask, par exemple, affirme que le basque n'oppose pas la vibrante et la battue à la finale de mot. (En ce qui concerne l'initiale de mot aucun des deux 'r' n'y est autorisé sauf dans des emprunts du type *radar*). Oyharçabal (1998) conteste néanmoins cette affirmation à la fois diachroniquement et synchroniquement en ce qui concerne le labourdin et le bas-navarrais. Quelles que soient les neutralisations effectives dans le dialecte parlé par nos locuteurs, ce qui est certain est que la présence d'une opposition entre /r/ et /ɾ/ force le système à utiliser un trait de 'longueur' pour différencier ces deux phonèmes. En effet, un [r] vibrant n'est rien d'autre qu'une battue [ɾ] qui est répétée à plusieurs reprises (cf. Ladefoged et Maddieson 1996: 215-245). Mais la présence de cette opposition bref/long est isolée dans le système puisque ni l'espagnol ni le basque (pour s'en tenir à ces deux langues) n'ont des géminées au niveau lexical. On voit donc a priori que cette opposition est fragile. La fragilité d'une opposition ne garantit pas sa disparition mais peut expliquer son élimination si d'autres facteurs interviennent (par exemple, des effets de mode sociolinguistique). Il est donc possible que le basque en France ait subi cette altération de son système en réponse à la structure interne des

oppositions et à l'influence du français. En tout état de cause, une explication sociolinguistique exogène qui attribuerait le passage d'une apicale à une uvulaire seulement à une radiation d'origine parisienne ne semble pas tout à fait convaincante. En effet, on constate qu'en portugais européen standard, malgré les barrières géographiques et culturelles qui séparent le Portugal de la France, l'opposition historique apicale battue vs. apicale vibrante ([karu] *caro* (cher) vs. [karu] *carro* (voiture)) a également laissé la place à une opposition apicale vs. uvulaire (/r/-ʁ/): à savoir, [karu] *caro* vs. [kaʁu] *carro*. Mateus et d'Andrade (2000) déclarent: «The uvular trill [ɐ] co-occurs with other back variants in modern E[uropean] P[ortuguese], namely the voiced uvular fricative [ʁ] or the voiceless one, [χ]; the alveolar trill is common in dialects other than the one under consideration. There is some reason to believe that the voiced uvular fricative [ʁ] is in the process of becoming dominant, today, in the Lisbon region» (p. 11).

Le français lui-même semble avoir suivi la même évolution de manière endogène. Martinet (1969) émet l'hypothèse suivante: «On peut se représenter comme suit le point de départ du phénomène: des réalisations uvulaires accidentelles de la forte, lapsus ou prononciations individuelles défectueuses, ont été favorisées comme offrant un moyen d'éviter des confusions ou des hésitations à comprendre dues à des prononciations insuffisamment énergiques de /-rr-/. L'opposition de /mari/ à /marri/ a ainsi cédé la place à /mari/ - /maʁi/». (Les mots auxquels Martinet fait référence sont *mari* "époux" et *marri* "attristé"). Par la suite, l'uvulaire se serait généralisée à toutes les positions et serait passée d'une vibrante à une fricative (ou approximante). Il n'est donc pas impossible que les locuteurs du basque dans leurs contacts avec le français aient rencontré assez tôt des systèmes présentant des oppositions du type /mari/ - /maʁi/ qui aurait pu

influencer et leurs usages du français et leur propre système basque. Il est également possible que le basque dans la partie française ait évolué comme le portugais vers des variantes uvulaires de la vibrante de manière endogène et que la rencontre avec le français ait renforcé cette tendance. En faveur de la première hypothèse, on notera que d'autres cas d'influence phonologique du français et/ou des langues romanes sur le basque sont attestés, comme, par exemple, l'apparition d'un son /y/ (ü) (variante phonétique puis phonème) —à côté du /u/ (u)— dans certains dialectes. Ainsi, selon R. Lafon (1973/1999: 18): «...le passage de *u* à *ü* devant consonne (sauf *r* douce et *s*) et en fin de mot dans le dialecte souletin et dans quelques parlers voisins est l'effet le plus frappant d'une tendance à la palatalisation de la sonante *u* qui s'est manifestée dans la moitié orientale du domaine basque, et plus particulièrement dans la région du nord-est... Elle est due au contact du basque avec des langues possédant la voyelle *ü* (gascon; provençal en Navarre (de la deuxième moitié du XI^e siècle jusqu'au XIV^e siècle))». Nous n'avons pas tous les éléments pour construire un scénario historique poussé, mais le fait que l'enquête PFC suggère des pistes de réflexion dans le domaine historique est à verser au dossier de notre travail descriptif et théorique.

5. Conclusion

Dans cet article, nous avons essayé de présenter un ensemble de questions relatives à la phonologie du français au Pays basque à la lumière de l'enquête PFC réalisée en 2002 et dont le dépouillement est en cours. Nous avons essayé de résumer une partie centrale du contexte sociolinguistique du Pays basque, à savoir le rapport entre le français et l'euskara. Sur le plan phonologique, nos observations sont parcellaires et provisoires. Nous avons

cependant remarqué qu'un bon nombre de traits de la phonologie du français au Pays basque rattachent cette variété au français d'Oc. Ce n'est pas tout à fait surprenant. Géographiquement les provinces basques françaises sont entourées de la Gascogne au nord et du Béarn au nord-est et à l'est. Encore aujourd'hui des villes comme Bordeaux et Toulouse représentent des choix évidents, voire incontournables, pour certaines formations de niveau universitaire. Historiquement, le Pays basque est fortement lié aux provinces d'Occitanie et administrativement, il fait actuellement partie de l'Aquitaine car il n'a pas le statut de région. Si, à la lumière de notre enquête, le rapport au français d'Oc semble indiscutable sur le plan phonologique, on sait encore trop peu de choses sur la posture sociolinguistique des Basques par rapport aux normes de prononciation. Il faut espérer que l'exploitation de l'enquête PFC et les enquêtes ultérieures qui viendront la compléter permettront de combler ce déficit.

Références

- ALLIÈRES, J. (1979): *Manuel pratique de basque*, Paris: Editions Picard.
- _____ (1999): *Les Basques* (6^{ème} édition revue et corrigée), collection *Que sais-je?*, Paris: PUF.
- AURNAGUE, J. & M. DUGUINE (à paraître): *Euskaldun orok altxa burua: enquêtes sociolinguistiques en Garazi 1982, 2002*, Eusko Ikaskuntza (collection Lankidetzan n. 29).
- BOURDIEU, P. (1993): *La misère du monde*, Paris: Seuil.
- CHAMBERS, J. K. & P. TRUDGILL (1980): *Dialectology*, Cambridge: Cambridge University Press.
- CLB (2001): *Continuité de la langue basque* (3), Enquête du Conseil de la langue basque et de l'Institut culturel basque.

- CSA (2001): *Demande sociale en matière d'enseignement du basque sur la communauté d'agglomération Bayonne-Anglet-Biarritz*, Sondage du CSA pour l'inspection d'Académie de Bordeaux. *Enbata* 1702.
- DELAIS, E. & J. DURAND (eds.) (2003): *Corpus et variation en phonologie du français: méthodes et analyses*. Toulouse: Presses Universitaires du Mirail.
- DELL, F. (1985): *Les règles et les sons. Introduction à la phonologie générative*. (Première édition 1973), Paris: Hermann.
- DURAND, J. (1976): *Generative Phonology, Dependency Phonology and Southern French*. *Lingua e Stile*, XI(1): 3-23. Egalement dans *Essex Occasional Papers*, vol. 16.
- _____ (1990): *Generative and Non-Linear Phonology*, Londres: Longman.
- _____ (1995): «Alternances vocaliques en français du midi et phonologie du gouvernement», *Lingua* 95: 27-50.
- DURAND, J. & J. EYCHENNE (sous presse): «Le schwa en français: pourquoi des corpus?», à paraître dans *Corpus* (numéro dirigé par Tobias Scheer).
- DURAND, J., J. EYCHENNE & A. MEQQORI (2004): «Manuel du 'classeur codeur PFC (schwa et liaison)», dans *Perl'*, 55-62.
- DURAND, J., B. LAKS & C. LYCHE (2002): La phonologie du français contemporain: usages, variétés et structure, dans C. Pusch & W. Raible (eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*, Tübingen: Gunter Narr Verlag, 93-106.
- DURAND, J. & C. LYCHE (1999a): «Regard sur les glissantes en français: français standard, français du Midi», *Cahiers de grammaire* 24, 39-65.
- _____ (2003a): «Le projet 'Phonologie du français contemporain' (PFC) et sa méthodologie», dans E. Delais & J. Durand (2003).
- _____ (2003b): «Structure et variations dans les systèmes vocaliques du français», in C. Sanders, A. Coveney and M.-A. Hintze (eds.), *Variation et francophonie*. Paris: L'Harmattan (à paraître).
- DURAND, J., C. SLATER & H. WISE (1987): «Observations on schwa in Southern French», *Linguistics* 25/2, 983-1004.
- DURAND, J. & J.-M. TARRIER (2003): Enquête phonologique en Languedoc (Douzens, Aude), *Tribune internationale des langues vivantes* 33, 117-127.
- Eychenne, J. & Ph. Hambye (2004): «La FAQ du codeur schwa», in J. Eychenne & G. Mallet (eds.), 29-34.

- EYCHENNE, J. & G. MALLET (eds.) (2004): *Du segmental au prosodique: protocoles, outils, extensions et travaux en cours*, Bulletin PFC n°3, ERSS, Toulouse-Le Mirail.
- GADET, F. (2000): «Derrière les problèmes méthodologiques du recueil des données», dans M. Bilger (ed.), *Linguistique sur corpus: Etudes et réflexions*, Actes du Colloque du 12 mars 1999, Perpignan, *Cahiers de l'Université de Perpignan* 33, 59-75.
- INSEE (2002): *Langues régionales, langues étrangères: de l'héritage à la pratique*. Insee Première n. 830, février 2002 (volet 'Transmission familiale des langues et parlers' du recensement 1999).
- LADEFOGED, P. & I. MADDIESON (1996): *The Sounds of the World's Languages*, Londres: Blackwell.
- LYCHE, C. (2003): «La loi de position et le français de Grenoble», in E. Delais & J. Durand (eds.) (à paraître)
- Martinet, A. (1969): «R, du latin au français d'aujourd'hui», Ch. VIII du *Français sans fard*, Paris: PUF (Version originale in *Phonetica* 8, 1962, 193-202).
- MATEUS, M. H. & E. D'ANDRADE (2000): *The Phonology of Portuguese*, Oxford: Oxford University Press.
- MICHELENA, L. (1977): *Fonética Histórica vasca* (Deuxième édition élargie et révisée. Première édition 1961), San Sebastián: Publicaciones del Seminario de Filología Vasca 'Julio de Urquijo' de la Excm. Diputación de Guipúzcoa.
- MILROY, L. (1987): *Language and Social Networks*, 2^{ème} édition, Oxford: Blackwell.
- OYHARÇABAL, B. (1998): Recension de 'R.L. Trask, *The History of Basque*, Lapurdum 3, 317-333.
- TRASK, R. L. (1997): *The History of Basque*, Londres: Routledge.
- VAN DE VELDE, H. & R. VAN HOUT (eds.) (2001): *R-atics. Sociolinguistic, phonetic and phonological characteristics of /r/, vol. 4 de Etudes & Travaux de l'Institut des Langues Vivantes et de Phonétique de l'Université Libre de Bruxelles*.
- WALTER, H. (1977) *La phonologie du français*, Paris: PUF.
- _____ (1982): *Enquête phonologique et variétés régionales du français*, Paris: PUF.
- _____ (1994): *L'aventure des langues en occident. Leur origine, leur histoire, leur géographie*, Paris: Robert Laffont.
- WURM, S. A. (ed.) (2001) *Atlas of the World's Languages in Danger of Disappearing/Atlas des langues en péril dans le monde*, UNESCO publishing.

ANNEXE 1 Tableau synoptique des témoins de l'enquête au Pays basque

Les informations fournies plus bas ne constituent qu'une sous-partie de nos fiches d'enquête et sont organisées comme suit:

Nom codé (sexe, âge au moment de l'enquête). Lieu de naissance (ville et département) et année de naissance. Lieu de résidence au moment de l'enquête. Profession. Niveau d'études. Compétence en basque.

Loc1 (f, 25 ans), née à St-Jean-Pied-de-Port (64) en 1976. Lieu de résidence: Bayonne. Analyste programmeur. Diplôme informatique (IUT). Bilingue basque/français.

Loc2 (f, 30 ans), née à Bayonne (64) en 1972. Lieu de résidence: Anglet. PRAG (Professeur agrégé) en IUT. Agrégation et DEA de gestion. Apprend le basque.

Loc3 (h, 31 ans), né à St-Jean-Pied-de-Port (64) en 1971. Lieu de résidence: Bayonne. Enseignant. Formation d'ingénieur. Bonne connaissance du basque.

Loc4 (f, 34 ans), née à Compiègne en 1967 mais venue à Biarritz (64) à l'âge de 6 mois. Lieu de résidence: Biarritz. Institutrice. DEUG de lettres et 2 ans de formation pédagogique. Apprend le basque.

Loc5 (f, 34 ans), née à Biarritz en 1967. Lieu de résidence: Anglet. Institutrice. Licence de sciences de l'éducation, 2 ans de formation pédagogique et 2 ans de formation à l'enseignement spécialisé. Bonne connaissance du basque.

Loc6 (f, 38 ans), née à St-Etienne-de-Baïgorry (64) en 1964. Lieu de résidence: Uhart-Cize (64). Conseillère juridique/avocate. DESS de droit des affaires et de la fiscalité. Connaissance élémentaire du basque.

Loc7 (h, 37 ans), né à St-Jean-Pied-de-Port (64) en 1964. Lieu de résidence: Uhart-Cize. Restaurateur. CAP, BEP, BTHD (Ecole Hôtelière). Bonne connaissance du basque.

Loc8 (h, 39 ans), né en 1963 à St-Etienne-de-Baïgorry mais a vécu à St-Jean-Pied-de-Port de 2 à 16 ans. Lieu de résidence: Biarritz. Chercheur (CR) au CNRS. Doctorat en informatique et linguistique et une Habilitation à diriger des recherches (HDR). Bilingue basque/français.

Loc9 (h, 60 ans) né à Saint Palais (64) en 1941. Lieu de résidence: Biarritz. Retraité (ex-médecin). Etudes de médecine + spécialité. Bilingue basque/français.

Loc10 (h, 61 ans), né 1941 à Paris mais venu à Bayonne (64) à l'âge d'un mois. Lieu de résidence: Bayonne. Médecin. Etudes de médecine. Bonne connaissance du basque.

Loc11 (f, 65 ans), née à St-Jean-Pied-de-Port (64) en 1936. Lieu de résidence: Uhart-Cize (64). Retraitée (ex-enseignante). BAC + formation d'enseignement. Bilingue basque/français.

Loc12 (h, 73 ans), né à Bayonne (64) en 1928. Lieu de résidence: Çaro. Retraité (ex-directeur de banque). BAC + conservatoire de musique. Bilingue basque/français.

Loc13 (f, 92 ans), née à Uhart-Cize (64) en 1910 (92 ans au moment de l'enquête). Lieu de résidence: Uhart-Cize. Retraitée (ex-institutrice). Brevet d'études + formation enseignement. Bilingue basque/français.

La sociolingüística de la variació i la fonètica experimental. Per a una complementarietat metodològica

Miquel Àngel Pradilla (URV), Josefina Carrera Sabaté (UB),

Ana Ma. Fernández Planas (UB)

1. La “socialització” de la lingüística

Com és sabut, la consideració de l'ús com a objecte d'estudi de la lingüística, tant la tradicional com la moderna, s'ha vist habitualment arraconat, quan no bandejat. Alhora, com a conseqüència de la desatenció esmentada, la variabilitat amb què es mostra la llengua, ja sigui per raons d'operativitat metodològica o per una aposta teòrica decidida per l'estudi *in abstracto*, ha estat presentada com una mena de nosa d'aprehensió impossible i d'interès, en el millor dels casos, secundari.

En el moment en què els lingüistes van intentar depassar els límits estrictes fixats per a la pròpia disciplina, quan es van interessar per l'estudi de les relacions entre sistemes, usos lingüístics i fets socials, van sorgir tota una sèrie de noves disciplines com l'antropologia lingüística, l'etnolingüística, la psicologia social, la sociologia del llenguatge i la

sociolingüística. Segons alguns autors, la peremptorieta de considerar els aspectes socials de les llengües responia a tres motivacions: a) el desig de trobar una base empírica sòlida per a la teoria lingüística; b) la convicció que els factors socials que influeixen en l'ús constitueixen un tema d'estudi molt legítim en la investigació lingüística; i c) la resposta a la creixent inquietud segons la qual l'esmentat coneixement sociolingüístic hauria de ser aplicat a problemes educatius urgents.

És en aquest marc, inicialment dominat per sociòlegs, antropòlegs i psicòlegs amb coneixements rudimentaris sobre el llenguatge, on, a poc a poc, s'aniran incorporant les primeres investigacions fetes ja per lingüistes. En definitiva, la sociolingüística afegirà a la tradició lingüística un nou component, una nova dimensió: el llenguatge com a manifestació de la conducta humana i com a element central de l'organització social i cultural de les societats.

La importància atorgada al context social —ja patent en alguns enfocaments de la lingüística general—, des de la dècada dels seixanta ençà ha anat donant forma a un àmbit de recerca que acull interessos compartits amb la sociologia, l'antropologia, la psicologia social, l'etnometodologia, la pragmàtica, l'anàlisi del discurs, l'anàlisi de la conversa, la lingüística del text, etc. Malgrat l'oposició de Labov,¹ aquest terme ha fet fortuna a l'hora d'etiquetar un espai de coneixements profundament interdisciplinari, però amb voluntat d'assolir autonomia paradigmàtica. Des d'aquesta perspectiva, hom considera que el coneixement d'una llengua viva és més complet si

¹ La prevenció de Labov cap al terme *sociolingüística* deriva del fet que aquesta etiqueta suposa, paral·lelament, la possibilitat de fer *lingüística* sense tenir en consideració el component social. No cal dir que, malgrat el desacord d'aquest investigador, aquesta "possibilitat" s'ha entès (i continua entenent-se en termes quantitatius que no es poden menystenir) en un sentit de "molta probabilitat".

permet mostrar, no només les relacions estructurals del sistema, sinó també el seu funcionament com a mitjà social de comunicació.

La compartimentació d'aquest gran "calaix de sastre" que acull els treballs que participen, d'una manera genèrica, d'interessos socials i lingüístics ha estat abordada per diferents estudiosos. Tanmateix, no sense aspectes sotmesos a controvèrsia, sembla que s'ha anat conformant un cert consens entre els investigadors. Així, a partir de la proposta terminològica de Fishman, la qual presentaria dos grans blocs, la *macrosociolingüística* i la *microsociolingüística*, se la sol presentar com una disciplina amb tres grans direccions: a) sociologia del llenguatge; b) etnografia de la comunicació; i c) sociolingüística estricta. Mentre que la primera línia monopolitzaria la perspectiva fishmaniana *macro*, les dues restants es repartirien l'enfocament *micro*.²

A propòsit del que s'ha dit, des de l'àmbit català ens agradaria deixar constància d'un parell d'especificitats:

a) El fet que la nostra tradició hagi estat (i continuï essent) essencialment sociològica —no és sobrer recordar que en el nostre context de contacte lingüístic allò que es debat és la supervivència de la llengua pròpia— ha contribuït a la (quasi) invisibilitació de les recerques etnogràfiques i variacionistes.

b) El fet que la sociolingüística presenti unes fronteres absolutament permeables ha permès que de la reflexió interdisciplinària —en fòrums com el que propicia aquesta xarxa temàtica— s'hagi arribat a una mena d'entesa teorimetodològica. Hem assistit, doncs, al naixement d'una *lingüística del corpus*, que ha superat la introspecció i els judicis d'acceptabilitat com a

² Una de les principals discrepàncies a l'hora d'acceptar la taxonomia proposada és la consideració microsociolingüística de la sociolingüística estricta. El fet que aquest model es fonamenti en l'establiment de correlacions entre variables lingüístiques i variables considerades macrosocials, com la classe social, l'ètnia, etc., és considerat per alguns autors com un factor decisiu que reubicaria aquesta disciplina en l'àmbit de la macrosociolingüística.

eina metodològica, i s'ha enfrontat a les dades lingüístiques amb tota la seua cruesa. D'aquesta manera, l'estudi de la variació ha guanyat nous practicants, més enllà de les tradicions dialectològica i sociolingüística.

2. La sociolingüística de la variació

En un article recent, Pradilla (2002) s'ha ocupat de caracteritzar les manifestacions que han pres a casa nostra els estudis sobre la variació lingüística, les “desviacions”, en definitiva, en relació amb el model de referència ortodox, les principals assumpcions del qual sintetitzarem més avall. Aquesta revisió ha pres suport en el plantejament multidimensional de la variació de Moreno (1990). Segons aquest autor, l'esmentada multidimensionalitat tindrà quatre vèrtexs que seran ocupats per la *geolingüística*, la *lingüística històrica*, la *pragmàtica* i la *sociolingüística* —amb la possibilitat de posicions intermèdies. Per a l'avinentesa que ens ocupa, ens centrarem en el vèrtex sociolingüístic, amb el benentès que la resta d'orientacions ens mereixen la mateixa consideració valorativa.

En l'àmbit de la sociolingüística estricta s'han elaborat diversos models destinats a explicar la variació sistemàtica. Els més destacats són el *dinàmic* (Bailey, Bickerton i De Camp) i el *quantitatiu* (Labov). Tanmateix, el decantament majoritari dels nostres investigadors pel segon³ motivarà que li dediquem la nostra atenció en exclusiva.

La proposta sociolingüística de Labov s'ha de considerar com l'elaboració més reeixida sorgida a partir de l'article de Weinreich, Labov i

³ També anomenat dialectologia urbana, dialectologia social, sociolingüística (“estRICTa”, “en sentit restringit”, “pròpiament dita”), estudi quantitatiu de la parla, model de regles variables, (socio) lingüística correlacional, variacionisme i sociolingüística de la variació.

Herzog (1968) «Empirical Foundations for a Theory of Language Change». El treball en qüestió va constituir un veritable salt qualitatiu en la consideració social del llenguatge. Aquest lingüista nord-americà beu, en els seus inicis, del model estructural a través del mestratge d'U. Weinreich. Tanmateix, a les darreries dels anys seixanta s'incorpora a la gramàtica generativa, de la qual proposa una extensió que incorpori un component social. Des d'aquesta perspectiva realista, l'objecte d'estudi serà observat en tota la seva diversitat.

Per a l'enumeració de les principals propostes teoricometodològiques del model labovià, prendrem com a base la síntesi de J. Colomina (1984: 81-82):

1. La variació lingüística lliure no existeix. L'heterogeneïtat de la llengua no és arbitrària, caòtica, sinó que és ordenada. La variació lingüística és inherent al sistema, del qual constitueix una propietat regular i intrínseca. Hi ha una correlació entre la variació lingüística i variables independents de tipus sociopragmàtic, com la classe socioeconòmica, el gènere, l'edat, la raça i l'estil.

2. És possible d'estudiar l'evolució dels canvis lingüístics en la sincronia, en *temps aparent*, mitjançant l'estudi dels canvis en moviment a través dels diversos grups d'edat d'una comunitat de parla.

3. En la descripció i l'explicació dels canvis en moviment es pot postular l'existència de variables sociolingüístiques correlacionades amb variables independents sociopragmàtiques. Les variables sociolingüístiques poden ser: a) *indicadors*, amb distribució social; b) *marcadors*, indicadors amb diferenciació estilística; i c) *estereotips*, marcadors conscients socialment. Les diverses variants tenen un significat social.

4. L'aparició de les diverses variants no pot descriure's adequadament en termes de diferències qualitatives, sinó *quantitatives*, ja que les variants apareixen amb més *frequència* en la parla —i, de vegades, en els diversos

estils— d'uns grups socials que en els altres. La quantificació estadística, amb les aportacions de l'escola canadenca (Sankoff, Cedergren i col·laboradors), dóna més rellevància als resultats probabilístics. Aquest darrer desenvolupament no és gens trivial ja que es fonamenta en la idea, epistemològicament essencial, segons la qual les dades d'actuació no són res més que un reflex estadístic de la competència, que ara serà quantificada mitjançant l'aplicació d'un model logístic multiplicatiu que permetrà convertir les freqüències observades en l'actuació en probabilitats teòriques. Per a l'estudi quantitatiu de la parla, s'ha elaborat tota una metodologia al voltant de l'*entrevista sociolingüística* que incorpora una tipologia dels estils contextuals i destaca la importància de les dades acústiques, especialment l'ús de l'espectrògraf.

5. Les gramàtiques posseeixen tres tipus de regles: *a) invariables* o *categòriques*, sense excepcions ni violacions; *b) semicategòriques*, amb violacions que poden ser interpretades; i *c) variables*. Les regles variables estableixen que per a cada parlant de la llengua la regla s'aplicarà amb més freqüència-probabilitat en un context que en un altre, i es formalitzen amb l'ús de parèntesis angulars:

$$X \text{---} \langle Y \rangle / A \text{---} B$$

Les regles variables donen compte del fet que la variació lingüística és funció de les restriccions sociopragmàtiques, que determinen la freqüència-probabilitat d'aplicació de la regla, i de les restriccions lingüístiques o contextos morfofonològics que afavoreixen o frenen l'aplicació de la regla.

3. La variació fònica en la llengua catalana

L'avaluació de la recerca sociolingüística entre els anys 1985-1995 va motivar l'observació de Boix i Payrató (1996) segons la qual la *sociolingüística del conflicte*, en l'àmbit català, havia donat pas a una *diversitat de sociolingüístiques*. Mentre que abans de l'any 1985 l'enfocament macrosociolingüístic (quasi) monopolitzava la producció, la dècada següent mostrarà una temàtica més variada. Des de la perspectiva variacionista s'albirava, doncs, un desenvolupament que no ha tingut la continuïtat esperada. Aquesta ínfima producció, a partir del 1995, tindrà en la tesi doctoral de Carrera (1999) i el treball de Montoya (2000) les úniques excepcions de gruix.

Aquesta reculada és constatada per Pradilla (2002) en una avaluació de conjunt de la recerca sobre variables fòniques afavorida per la sociolingüística de la variació. Una conclusió que no es veu alterada si hi afegim la consideració de l'estudi de variables morfològiques, sintàctiques i lèxiques. Encara més, tot i la millor salut investigadora de la branca etnogràfica, el tractament conjunt de tota la recerca microsociolingüística continuarà sent marginal en el marc d'una tradició sociolingüística essencialment sociològica.

Tornant a la recerca variacionista, és un fet conegut que la sociolingüística laboviana s'ha emprat de manera aclaparadorament majoritària en els treballs sobre variables fòniques i des d'una orientació essencialment sincronicista. I una de les causes que ha afavorit aquesta supremacia és, sens dubte, la seva millor adaptació a la premissa segons la qual la variació consisteix en la utilització alternativa de formes equivalents semànticament i pragmàticament, cosa que en lèxic i sobretot en sintaxi és

difícil garantir. Els treballs nuclears —amb independència de les publicacions generades— que hom pot inventariar són: Colomina (1985), Mier (1986), Montoya (1989), Alturo i Turell (1990), Escrivà (1993), Montoya (1993), Pradilla (1993), Plaza (1995), Pla (1995), Carrera (1999), Montoya (2000) i Carrera i Freixenet (2003).

4. El mètode experimental en fonètica i la metodologia variacionista

4.1. La fonètica i el mètode experimental

A partir de la segona meitat del segle XX sorgeixen una sèrie de corrents, escoles i disciplines que determinen un abans i un després tant en el món de la lingüística com en d'altres ciències del saber humà. Els anys cinquanta apareix la Gramàtica Generativa i Transformacional i es comencen a desenvolupar àrees de treball com la Semàntica, la Semiòtica i la Fonètica Experimental. Però, més a més, a les ciències tradicionalment conegudes se n'afegeixen d'altres com la lògica i la informàtica, que revolucionarà els estudis de tot tipus encetant unes possibilitats fins aleshores insospitades. Comença un període de connexió, intercanvis i interrelació entre els diferents camps del saber que beneficiarà a tots, entre ells també els de la lingüística amb totes les seves disciplines, inclosa la fonètica, i les anomenades ciències socials, àmbit on tindrien cabuda la psicolingüística, la sociolingüística o la filosofia.

La curiositat per l'estudi dels sons de la parla humana ve de lluny en el temps, ens podríem remuntar a fa més de dos mil anys, als estudis de Panini i la seva preocupació per establir la pronúncia inalterable dels llibres sagrats

en sànscrit per a les cerimònies i els ritus. Però fins a mitjan segle XX els estudis se solien limitar a les impressions articulatòries pròpies de l'investigador —no hi ha havia tècniques invasives i fiables— o a les impressions auditives de l'experimentador (fonètica auditiva). A propòsit d'això, podem recordar la famosa frase de Paul Passy «rien ne vaut une oreille attentive», recollida per Jones (1941). Així i tot, abans de l'esmentada data de referència, la dècada dels anys cinquanta del segle passat, ja s'havien produït intents de trobar la manera de sistematitzar els estudis fonètics i de dotar-los de la rigorositat d'un mètode experimental, inicialment amb eines gràfiques. Ens referim al segle XIX. En aquest sentit, l'obra de Rousselot, a cavall entre el segle XIX i el segle XX, va ser fonamental. Fou l'any 1892 que el considerat pare de la fonètica experimental contradeia Passy amb la coneguda asseveració: «l'oreille n'est pas un instrument suffisant d'analyse»; i el 1897 va fundar a França el que podem considerar el primer laboratori de fonètica experimental de la història. En aquest centre Rousselot va utilitzar àmpliament el quimògraf, instrument inventat feia mig segle per K. Ludwig, a partir del qual va elaborar el 1924 els *Principes de phonétique expérimentale*.

D'aleshores ençà és innegable l'avenç que ha experimentat la fonètica amb l'ús del mètode experimental. Tant la fonètica general com les fonètiques particulars de cadascuna de les llengües que l'han aplicat. Fins al punt que hi ha qui utilitza l'expressió *fonètica experimental* com a sinònim del terme *fonètica*. D'altra banda, també hi ha autors com Malmberg que la utilitzen com equivalent de *fonètica instrumental*. En rigor cap de les dues equivalències són encertades. El mot genèric és fonètica, evidentment, però no tota la fonètica és experimental. Podem parlar de fonètica teòrica, contrastiva o històrica, per exemple, sense fer ús de l'experimentació. D'altra banda, el fet d'emprar utillatge instrumental de laboratori no

assegura que estiguem fent fonètica experimental. Com diu Martínez Celdrán (1991: 13): «los instrumentos y los métodos de registro y medición son complementarios y entran dentro de lo que entendemos por método experimental ya que ayudan a una observación rigurosa y ordenada de la realidad; pero no son el objeto último de la llamada fonética experimental».

La fonètica experimental es distingeix per utilitzar el *mètode experimental*. Aquest constitueix un procediment científic pautat que seguirà la seqüència següent:

- a) planteja el problema a partir de l'observació de fets inexplorats o de fets contradictoris amb investigacions anteriors,
- b) formula hipòtesis que l'expliquin,
- c) observa les dades, les recull, les recopila i les ordena,
- d) les analitza: hi experimenta,
- e) les interpreta,
- f) en comunica els resultats i les conclusions que se'n deriven.

L'ús d'instrumental de laboratori s'ha de situar en els passos *c* i *d*. El fet d'experimentar inclou tot un procés en el qual l'investigador varia de manera controlada i premeditada els fenòmens que observa per tal d'aïllar el(s) paràmetre(s) que considera causant(s) dels fenòmens esmentats. Cal desenvolupar aquest estadi del procés de manera instrumental aprofitant les eines que ens ofereix la tecnologia tant per a l'anàlisi pròpiament fonètica com per a l'anàlisi estadística posterior, ja que les dades que se n'obtidran seran normalment quantificables.

En fonètica l'ús d'aparells específics és necessari per comprovar impressions que qualsevol pot tenir per simple reflexió personal o observació dels altres; per exemple, la constatació mitjançant oscil·loscòpia o sonografia que una oclusiva bilabial retén la sortida de l'aire cap a l'exterior. Esdevé especialment útil per comprovar i quantificar fets

articulatoris i/o acústics que no són evidents a simple vista ni, de vegades fins i tot, accessibles a la pròpia introspecció; per exemple, la constatació de fets assimilatoris que tenen lloc a l'interior de la cavitat bucal o, en general, resultats del fenomen de la coarticulació. I és absolutament imprescindible a l'hora de "manipular" conscientment un sol paràmetre acústic i deixar constants els altres per avaluar perceptivament l'efecte que produeix, com passa habitualment en experiments que treballen amb síntesi de veu.

4.2. L'encaix del mètode experimental en la recerca variacionista

Aquest apartat es proposa destacar la influència mútua creixent que experimenten les metodologies de la sociolingüística de la variació i de la fonètica experimental. En el 15th International Congress of Phonetic Sciences, celebrat a Barcelona l'agost de 2003, hom va poder constatar la presència destacada d'un bon nombre de treballs que s'aixoplugaven sota l'etiqueta *Sociofonètica*. Des que Labov (1966, 1972) va establir correlacions entre la qualitat vocàlica i el perfil social dels informants mitjançant l'ús de tècniques instrumentals, els seus seguidors han proporcionat un degoteig de treballs que amb el temps ha anat fent gruix. I paral·lelament, els resultats obtinguts han contribuït al fet que els fonetistes atorguin més rellevància a les qüestions teoricometodològiques d'índole socioestilística més desateses.

Una de les aportacions fonamentals de la sociolingüística ha estat l'assoliment d'un veritable salt qualitatiu en l'elaboració d'una metodologia empírica i realista. De la mà de la sociologia, s'ha posat molt d'èmfasi en aspectes que posteriorment esdevindran interpretativament determinants. Ens referim, essencialment, a tres qüestions:

- a) Una selecció d'informants que pren suport en l'establiment d'una

mostra estratificada de la comunitat de parla estudiada.

b) L'elaboració d'una eina metodològica de recollida de dades, l'entrevista sociolingüística, que, en funció de la recerca, ha de permetre la segmentació d'un continuum estilístic. A més a més, la recerca global sol completar aquestes dades amb unes altres procedents de l'observació anònima (o participant) i de proves d'avaluació subjectiva.

c) I, finalment, quant al tractament quantitatiu, s'han desenvolupat programes que prenen suport en models matemàtics logísticomultiplicatius. Mitjançant la correlació de variables, la quantificació haurà de visualitzar en termes freqüencials i probabilístics el caràcter ordenat de la variabilitat estudiada.

Si reprenem l'esquema de treball del mètode experimental, sembla, doncs, que són justament els apartats vinculats a l'experimentació (*c* i *d*) els que poden aportar una millora substancial en els treballs de base variacionista. Una altra cosa són les recerques de base experimental que recullen consideracions metodològiques de la sociolingüística de la variació. En aquestes darreres, la quantificació i la discussió dels resultats també segueix les pautes de la fonètica.

Segons la nostra opinió, dues qüestions centrals en què la metodologia sociolingüística es veu enormement afavorida amb la incorporació d'algunes estratègies del mètode experimental en fonètica són: *a*) l'anàlisi qualitativa de determinades variables s'enriquirà amb la delimitació d'al·lofonies fixades amb procediments instrumentals; i *b*) es garantirà la validació empírica de les dades obtingudes, requeriment fonamental quan es treballa amb mètodes d'anàlisi rigorosos.

En relació amb aquestes incorporacions metodològiques, la tradició variacionista catalana no presenta un panorama molt lluït. En la nòmina de recerques variacionistes presentada només trobem anàlisi acústica en els treballs de Montoya (1989 i 1993), Pradilla (1993) i Carrera (1999). Tanmateix,

només a Pradilla tot el conjunt d'ocurrències va ser validat instrumentalment, mentre que a Montoya i Carrera s'analitza una mostra de dades i la resta són interpretades de manera impressionística, mitjançant el criteri auditiu.⁴

És clar, doncs, que el mètode experimental en fonètica pot millorar determinats aspectes de la metodologia sociolingüística. Al nostre parer, un punt central és la incorporació d'un enregistrament professional a l'entrevista sociolingüística, que garanteixi una acurada anàlisi instrumental correcta de les dades. Si l'investigador vol obtenir substància fònica que pugui ser analitzada acústicament, ha de centrar els seus esforços en les qüestions següents:

a) Proveir-se d'un utilatge de qualitat, ja sigui digital o analògic. Convé no oblidar que la tria del micròfon també serà molt important, ja que el resultat final d'un enregistrament sempre coincideix amb la qualitat de l'element més deficient amb què s'ha treballat.

b) Aconseguir unes bones condicions d'enregistrament. El protocol metodològic variacionista recomana que l'entrevista es faci al domicili de l'informant, lloc on se sent més segur i còmode, però que no garanteix, d'entrada, unes condicions d'enregistrament òptimes. Tanmateix, una estratègia d'accés adequada —sovint amb la participació de “mitjancers”— ens ha de permetre crear unes condicions de pseudolaboratori, que evitin els principals factors de distorsió del so, això és, els sorolls interiors i exteriors, la reverberació i l'eco.⁵

⁴ En el cas de Carrera totes les realitzacions dubtoses van ser validades sonogràficament. Finalment, el treball de Pla (1995) constitueix un bon exemple d'exploració acústica prèvia a l'establiment de les variants en el disseny definitiu de la investigació.

⁵ Convé no transmetre a l'informant la nostra preocupació per aconseguir un enregistrament òptim. L'ideal és que la conversa es desenvolupi de la manera més distesa possible, minimitzant la presència de l'utilatge. És per això que la intervenció del mitjancer pot esdevenir fonamental, ja que el fet de conèixer l'informant li permet sol·licitar d'una manera “estratègicament delicada” els requeriments mínims necessaris. D'altra banda, en la planificació de la recerca cal no perdre de vista que les estacions on la bonança del temps convida a tenir les finestres obertes (primavera i estiu) són les menys recomanables per fer el treball de camp, ja que els sorolls exteriors poden dificultar enormement la nostra tasca.

D'altra banda, és fonamental disposar d'analitzadors acústics potents connectats a ordinadors, fet que permetrà emmagatzemar dades sense necessitat d'imprimir-les. L'estalvi de temps i econòmic és obvi, alhora que el nou sistema permetrà tractar les dades en futurs treballs sense haver de recórrer a la tasca, sempre feixuga, de buscar novament els mots dins de les cintes enregistrades.

La comunió metodològica proposada troba el seu principal obstacle en la superació de la *paradoxa de l'observador*. Aquest és el punt essencial, que s'ha d'intentar combatre mitjançant l'adopció de tot un seguit d'estratègies que tendeixen a reduir la formalitat de la conversa. Al capdavant, la intervenció professional de l'investigador acabarà sent clau en l'èxit de l'entrevista.

5. Estudis de cas en l'àmbit de la sociofonètica

La *sociofonètica* està connectada amb l'estudi de la parla en el context social en què es produeix i combina tècniques de la sociolingüística i de la fonètica (tant acústica com articulatòria i perceptiva). Tot i que, en general, el terme ha estat usat per referir-lo a la recerca sociolingüística que ha utilitzat dades provinents de l'anàlisi instrumental acústica, segons (Wright: 2001), la sociofonètica també es pot referir al treball fonètic quantitatiu basat només en anàlisis auditives. A banda d'això, el terme sociofonètica no ha estat l'únic per referir-se a aquesta realitat i s'han proposat alternatives terminològiques com "accent studies", per part de lingüistes com Foulkes i Docherty (1999).

A continuació presentem alguns estudis realitzats a partir de diferents anàlisis instrumentals. Veurem en primer lloc un cas referit al vocalisme; en segon lloc, un fenomen que afecta el consonantisme; i, finalment, en tercer

lloc, un cas suprasegmental. El segon i, en menor mesura el tercer, s'emmarquen en la *sociofonètica*. Tots tres blocs constitueixen exemplificacions a partir de treballs parcials ja publicats que s'integren en projectes d'investigació més amplis.

5.1. El vocalisme tònic mitjà del català

Els estudis sobre transmissió en el vocalisme contrastiu interdialectal de Carrera *et al.* (1999) i (2003) formen part d'un projecte més ampli i són el resultat de proves exploratòries sobre treball vocàlic contrastiu interdialectal.

El corpus d'aquests treballs està constituït per les vocals mitjanes de la sèrie anterior, [e, ɛ] i de la sèrie posterior, [o, ɔ]. Totes han estat emeses en seqüències CVC simètriques. Les consonants precedents i següents al context adjacent contenen uns trets característics articuladoris diferents: oclusiva bilabial ([p]), oclusiva dentoalveolar ([t]), nasal alveolar ([n]), fricativa alveolar ([s]), vibrant ([r]) i lateral ([ɭ]). El resultat d'aquesta combinació de consonants i vocals ha estat un conjunt de logòtoms emesos dins la següent frase portadora: *diu CVC quan vol*. S'ha treballat amb logòtoms perquè hi ha hagut la intenció d'utilitzar algunes de les emissions en els tests de percepció i no s'ha volgut donar cap altra informació dialectal que no fos el timbre de cada vocal per tal de no distorsionar els resultats. D'acord amb això, la frase portadora ha estat constituïda per elements l'emissió dels quals és relativament invariable interdialectalment. Cada frase s'ha repetit tres vegades. Cada vocal s'ha emès 18 vegades per parlant. En definitiva, cada so s'ha pronunciat 54 vegades en cada dialecte.

Els parlants han estat 9 homes (tres per a cada variant dialectal) de 22 a 30 anys d'edat, la pronúncia dels quals no presenta afectació ni influència del castellà. Els parlants procedeixen de diferents àrees geogràfiques: Barcelonès, Segrià i la Safor.⁶ D'aquesta manera l'estudi compta amb representants del català oriental (els informants de l'àrea de Barcelona) i del català occidental (els parlants de l'àrea de Lleida i de Gandia).

Les dades recollides constitueixen mostres de parla de laboratori enregistrades en òptimes condicions per a ser posteriorment analitzades i l'anàlisi acústica ha consistit en la mesura freqüencial (F1, F2)⁷ al període quasi-estable de cadascuna de les vocals investigades mitjançant la tècnica de l'LPC (Martínez Celdrán: 1998). S'ha utilitzat el CSL4300B de KAY Elemetrics Corp. al Laboratori de Fonètica de la UB. Posteriorment, les dades obtingudes han estat sotmeses a tractament estadístic amb el paquet SPSS mitjançant *One Way ANOVA*, *post hoc Scheffé* i l'escalament multidimensional (EM). Els factors tinguts en compte han estat el timbre de les vocals, el dialecte i el context consonàntic. La significació dels contrastos s'ha establert en $p < 0,05$.

Els resultats obtinguts de l'anàlisi de les vocals mitjanes anteriors del lleidatà i barceloní mostren que mitjançant la tècnica de l'Escalament Multidimensional permeten observar una clara separació dels estímuls de les vocals mitjanes segons la sèrie tímbrica i el dialecte, especialment en el cas de les vocals anteriors. (Vegeu la figura 1.)

⁶ Les dades de Gandia s'han obtingut gràcies a la col·laboració de l'Escola Superior de Gandia, i, especialment a Rubén Picó Vila i a Jaime Ramis Soriano.

⁷ De fet, també hem mesurat l'F3, així com la diferència entre F1 i F2, entre F2-F3 i entre F1-F3, però després d'unes anàlisis prèvies hem vist que aquests paràmetres no són prou rellevants per al nostre objectiu i, per tant, s'han descartat.

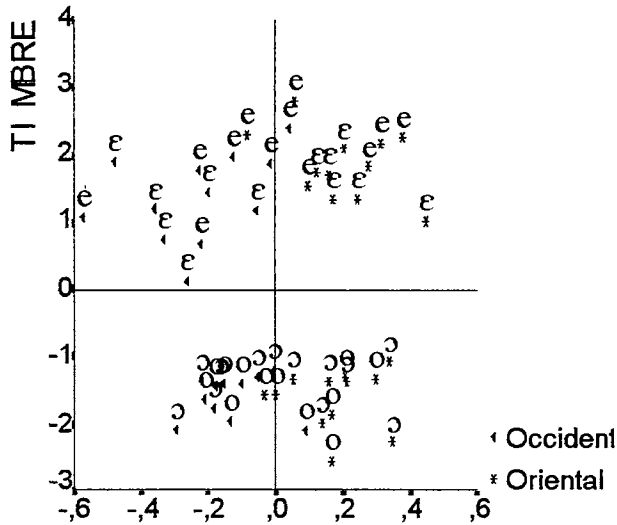


Figura 1. Distribució de les vocals mitjanes a partir de l'MDS.

Després d'analitzar la freqüència del segon formant, s'observa que les vocals del lleidatà presenten un F2 més baix que les del barceloní i per això són més obertes. En el cas del vocalisme posterior, en general F2 és més alt en barceloní que en lleidatà però la diferència no és tan sistemàtica com en el cas de les vocals anteriors i només podem observar una major obertura de la vocal posterior alta del lleidatà en contacte amb l'oclusiva bilabial [p] i, en el cas de la vocal posterior baixa, en contacte amb [t], [p] i [s]. (Vegeu la figura 2.)

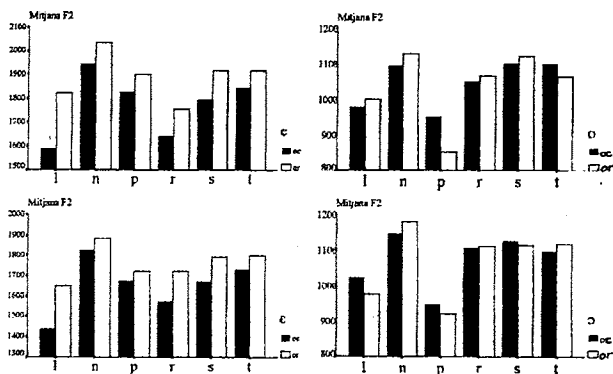


Figura 2. Valors d'F1 i F2.

Les dades provinents de l'anàlisi formàntica de les vocals mitjanes posteriors del lleidatà, del barceloní i del valencià meridional mostren que les vocals posteriors del valencià meridional tenen un F2 sistemàticament molt més alt que el d'altres dialectes, la qual cosa és indicativa d'un timbre molt més obert en valencià que en lleidatà i barceloní. (Vegeu la figura 3.)

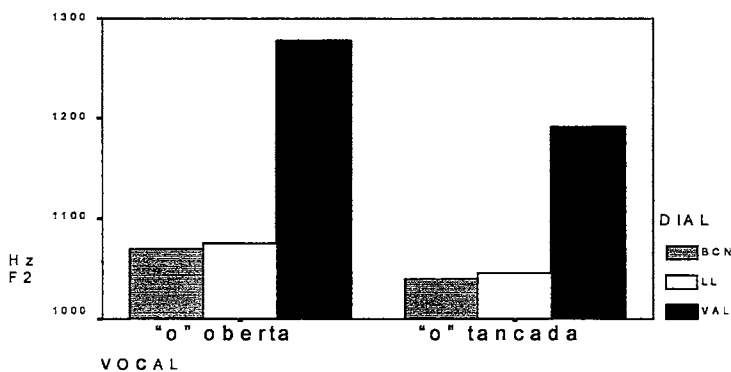


Figura 3. Mitjana de F2 en [o] i [ɔ].

A part d'això, en aquests tres dialectes la influència que exerceixen les consonants sobre les vocals mitjanes posteriors es pot ordenar de menor a major incidència en l'obertura vocàlica de la manera que es detalla a continuació: bilabial <alveolar velaritzada <alveolar i dentoalveolar. (Vegeu els gràfics 4 i 5.)

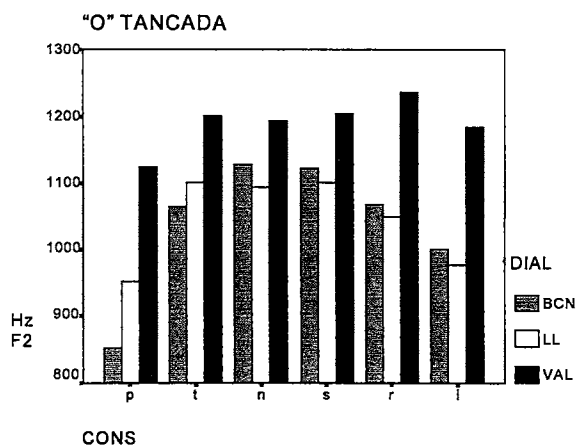


Figura 4. Mitjana dels valors de F2 a [o].

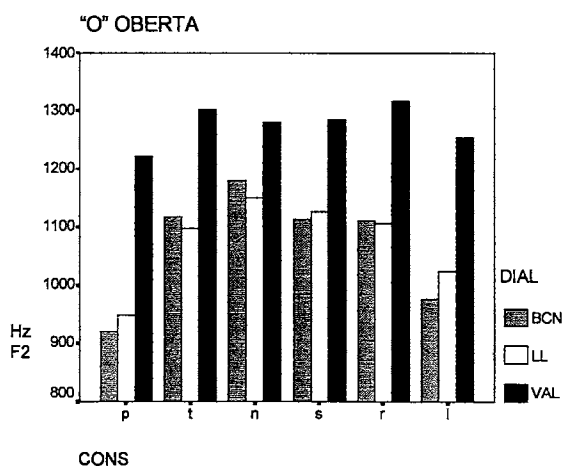


Figura 5. Mitjana dels valors de F2 a [ɔ].

Els estudis de percepció de contrast vocàlic interdialectal de Fernández *et al.* (2000) i (2002) i de Carrera *et al.* (2001) han pretès observar si les diferències d'obertura constatades entre lleidatà i barceloní també tenen lloc en el component de la percepció.

El corpus està constituït per logòtoms d'estructura CVC (on C1=C2). Les consonants tractades han estat utilitzades en treballs anteriors de Carrera, Fernández i Matas (1999) i són les següents: [p, t, s, n, ɫ, r].

Cadascun dels tests elaborats conté realitzacions de tres informants nadius del barceloní i d'altres tres nadius del lleidatà. Aquests sis informants van ser els que havien servit per analitzar les característiques acústiques de les vocals mitjanes a Carrera, Fernández i Matas (1999).

Els informants que han fet de jutges han estat, d'una banda, 50 catalanoparlants: 25 lleidatans i 25 barcelonins, tots amb el català com a L1 (Fernández Carrera i Matas: 2000). Al treball de Carrera, Fernández i Ortega (2001), els informants han estat 30 castellanoparlants fills de castellanoparlants que resideixen a Lleida. Finalment, en l'estudi de Carrera i Fernández (2004), els jutges han estat els 25 catalanoparlants i els 30 castellanoparlants de Lleida. Tots els enquestats⁸ tenen entre 18 i 20 anys i no s'ha tingut en compte la distinció de gènere entre ells.

El desenvolupament de l'enquesta ha estat el següent en tots els casos: en primer lloc, els informants han sentit els estímuls objecte de contrast inserits en la frase portadora següent: *Diu ____ quan vol*; en segon lloc, han sentit els tests que contenien logòtoms aïllats. En total, els tests han estat quatre: a) logòtoms amb frase portadora de les vocals mitjanes anteriors dels dialectes barceloní i lleidatà; b) logòtoms amb frase portadora de les vocals mitjanes posteriors dels dialectes barceloní i lleidatà; c) logòtoms sense

⁸ Agraïm la col·laboració desinteressada dels estudiants de l'IES Fort Pius de Barcelona i de l'IES Ronda i IES Maria Rúbies de Lleida.

frase portadora de les vocals mitjanes anteriors dels dialectes barceloní i lleidatà; i d) logòtoms sense frase portadora de les vocals mitjanes posteriors dels dialectes barceloní i lleidatà. Cada logòtom s'ha sentit dues vegades amb un segon de silenci al mig; cada dos logòtoms, el silenci ha estat de tres segons. Els logòtoms s'han presentat barrejats aleatòriament als oients.

Els resultats assenyalen el següent en relació amb la percepció barceloní-lleidatà:

1) Els estímuls més ben reconeguts són els que es presenten en frase portadora, paral·lelament al que s'observa en altres estudis com els de Cole i Jakimik (1980).

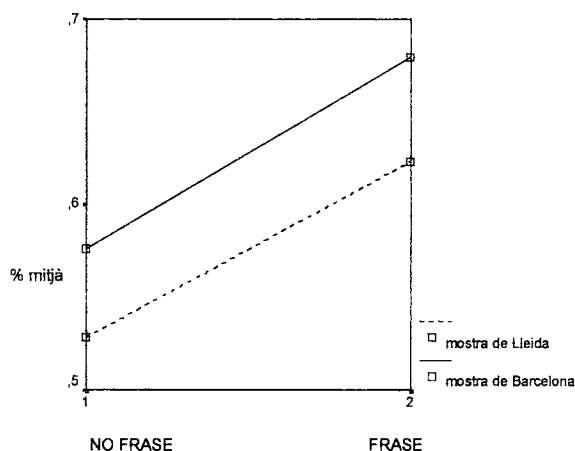


Figura 6. Distincions en la percepció de vocals segons la seva inclusió o no en frase portadora.

2) Els oients de Barcelona reconeixen sistemàticament més bé tots els timbres que els oients de Lleida i presenten millors resultats quan els estímuls provenen de parlants barcelonins. Al mateix temps, els jutges de Lleida perceben millor els estímuls del seu propi dialecte. (Vegeu les figures 7 i 8).

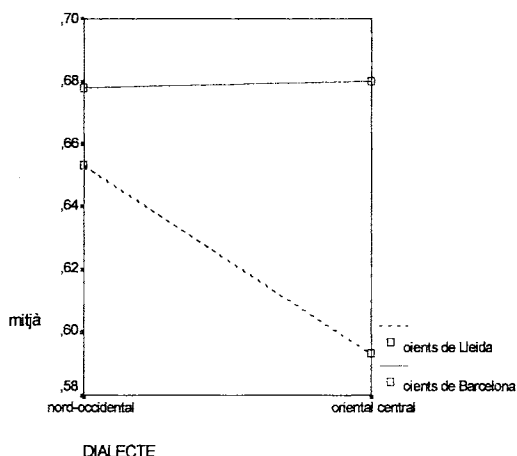


Figura 7. Percepció de logòtoms en frase segons el timbre de la vocal.

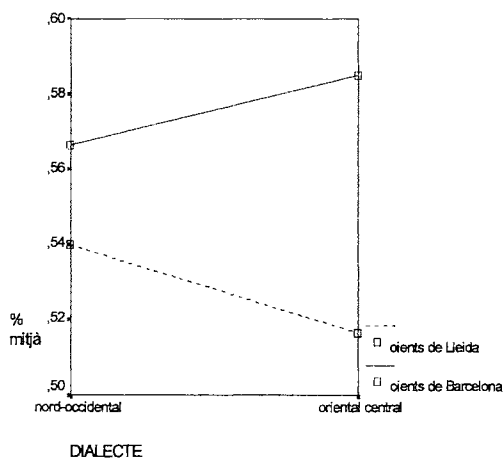
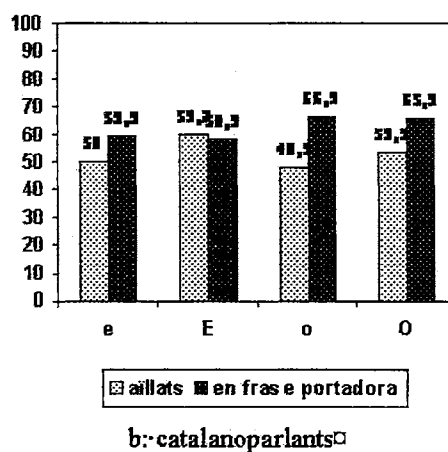
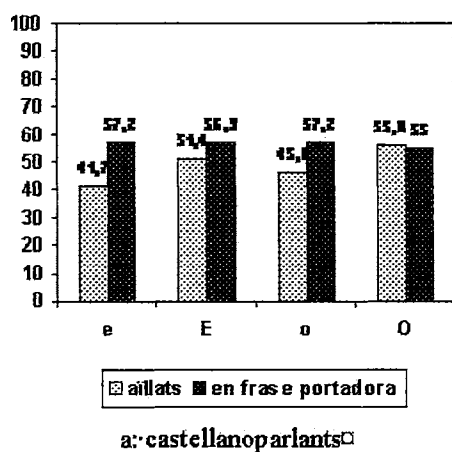


Figura 8. Percepció de logòtoms aïllats segons el timbre de la vocal.

Pel que fa a la percepció català-castellà a Lleida, veiem que tot i que la tendència general indica que els catalanoparlants identifiquen millor els logòtoms que els castellanoparlants (lleidatans), presos els estímuls en bloc les diferències no són significatives estadísticament. Aquesta observació

general, no obstant això, es veu matisada en relació amb la variable context semàntic. Efectivament, en contrastos parcials es constata que els oients catalanoparlants perceben significativament millor les vocals anteriors sentides aïlladament i les posteriors sentides en frase portadora que els oients castellanoparlants. (Vegeu les figures 9 i 10.)



Figures 9 i 10. Encerts en percentatges.

A banda d'això, totes dues mostres d'oients identifiquen millor les vocals mitjanes baixes [ɛ, ɔ] que les mitjanes altes [e, o], un cop sentides sense context. Segurament, aquesta observació pot relacionar-se amb la naturalesa acústica de les vocals obertes respecte de les tancades, i amb una constatació derivada de l'observació del contrast vocàlic en totes les llengües del món, on la diferència més rellevant entre les vocals és l'obertura vocàlica per davant de l'anterioritat lingual (Ladefoged i Maddieson: 1996).

5.2. Les africades alveolopalatals del català

Aquest treball en curs pretén estudiar els diversos fenòmens que afecten els africats en tot el domini català com poden ser l'homorganicitat del elements de l'africat o la seva sonoritat. Per tant, inclou una variant geolingüística (diversos punts d'estudi en tot el territori), també una variant estilística (perquè busca caracteritzar aquests tipus de sons en diferents tipus de parla —bàsicament lectura i espontània—), una variant generacional (l'estudi es vol ampliar a diferents generacions vitals) i, finalment, una variant social (perquè hi haurà diferents estrats socials entre els informants, tant pel que fa a la procedència familiar com pel que fa al grau d'instrucció). La nostra ferma voluntat és estudiar exhaustivament les dades amb instrumental fonètic adient i deixar en un segon pla el paper de l'oïda.

De moment, hem publicat alguns treballs que se centren en aspectes puntuals en aquest marc general. Vegem-ne alguns amb una mica més de detall.

A Fernández Planas y Pradilla (2004) es presenta un estudi articulatori i acústic sobre les característiques de les africades alveolopalatals del català oriental central, concretament del barceloní.

Es pretenia estudiar l'homorganicitat i el comportament de la sonoritat d'aquestes articulacions en tres velocitats de parla (lenta, còmoda i ràpida) i també en funció de la tonicitat o atonicitat de la síl·laba en què s'inseria l'articulació. Les articulacions estudiades van ser la sorda i la sonora sempre en el mateix context vocàlic ([*aCə*] i [*ə'Ca*], segons la posició de l'accent). En cada bloc accentual es van estudiar cinc ítems d'ús habitual que es van repetir cinc vegades en cadascuna de les velocitats d'elocució. Es va treballar a partir de les dades de dues informants entre les quals hi havia una diferència d'edat de tres generacions, si considerem que es produeix una nova generació cada set

anys. El nivell d'instrucció i l'estrat social es pot considerar el mateix en totes dues.

Les variables tingudes en compte van ser, en l'apartat articulatori, els índex d'anterioritat (CA), de posterioritat (CP) i de centralitat (CC) dels contactes (Fontdevila, Pallarès i Recasens, 1994) en cadascun dels segments de l'articulació en el seu centre del seu desenvolupament temporal; i en l'apartat acústic, les durades de l'oclusió, de la sonoritat en l'oclusió (en les sonores), de la fricació i de la sonoritat en la fricació (en el cas de les sonores). Les emissions es van obtenir a partir d'enquestes orals perquè l'escriptura no influís en el comportament lingüístic de les informants, tot i el seu grau d'instrucció, i entre les preguntes se'n barrejaven algunes la resposta de les quals no tenia res a veure amb l'objectiu del treball, però buscaven dissimular-ne el veritable propòsit d'estudi. Així, segons Moreno (1990:105), la tècnica utilitzada va ser la de l'enquesta directa. Segons la tècnica laboviana de segmentació del contínuum estilístic (Labov, 1966; Pradilla 2001a; Pradilla 2001b), entraria dins l'estil acurat propi d'una conversa (en aquest cas de fet entrevista) dirigida.

Els aparells emprats van ser l'EPG3 de Reading i l'Speech Analyzer 24.17 del SIL, respectivament, per a les anàlisis articulatòria i acústica. Les dades numèriques obtingudes van ser tractades estadísticament amb el paquet SPSS. El treball es va desenvolupar al Laboratori de Fonètica de la UB.

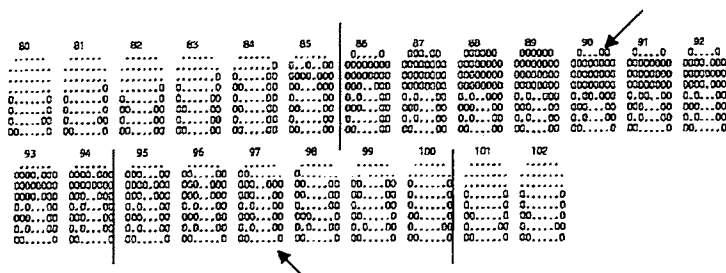


Figura 11. Successió d'electropalatogrames d'una articulació alveolopalatal al mot «encoratjar».

Els resultats acústics, d'altra banda, apunten cap a un procés d'ensordiment de l'articulació afectat molt poc per la posició de l'accent en la paraula i molt per la velocitat de parla de l'emissor, i aquest procés afecta més la informant més jove de totes dues.

Així, es produeixen emissions tan curioses com les dels sonogrames següents (vegeu la figura 12): el primer mostra una africada amb la part oclusiva sonora i la part fricativa sorda; el segon només ensordeix la primera part de la fricativa.

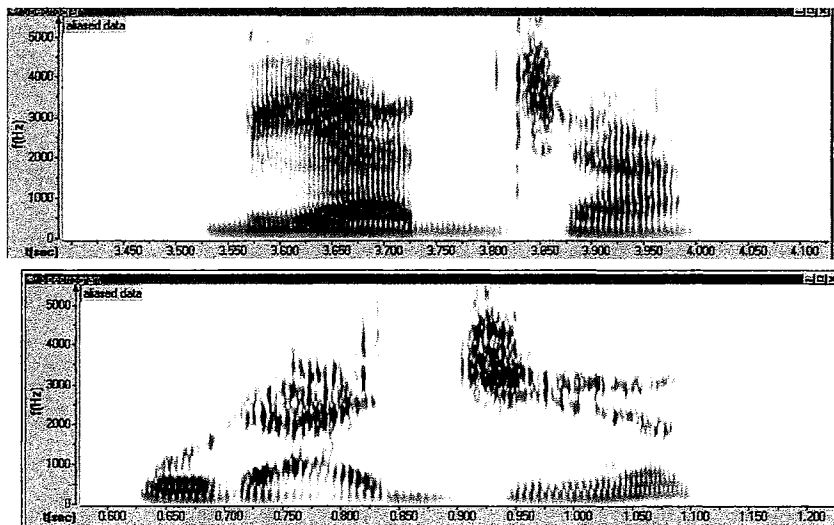


Figura 12. Sonogrames de les seqüències «viatge» (a dalt) i «coratge» (a baix).

En tot cas, en els nostres resultats, es constata que l'ensordiment comença per l'element fricatiu, d'acord amb Salcioli (1981) o Pradilla (1996, 1997); però aquest resultat difereix del que troben altres autors com Recasens (1991). Acústicament es percep que, lògicament, a major velocitat, major reducció de la durada, tant de la global del segment com de la sonoritat, quan n'hi ha. Articulatoriament, una de les informants presenta, a major velocitat de parla, menys superfície de contacte de tots dos segments i una major homorganicitat de punt d'articulació, segons l'esperat; però en la informant més gran aquesta tendència no es pot confirmar. D'aquesta manera s'observen diferents tendències en l'estudi d'aquest fenomen complex i controvertit que demostren l'enorme variabilitat que es produeix i la situació de canvi i d'instabilitat que experimenta. Si noves dades confirmen, en les noves generacions, la tendència apuntada en relació amb l'ensordiment de la fricativa, aquest fet podria tenir repercussions en la transcripció estreta del fenomen.

Els treballs de Fernández Planas, Carrera Sabaté i Pradilla (2003) i també de Pradilla, Carrera Sabaté, Fernández Planas (en premsa) aporten altres exemples que troben aixopluc en el mateix projecte d'estudi global de la variabilitat que dibuixa un perfil de canvi lingüístic en l'ordre alveolopalatal. Novament es tracta d'una recerca que es basa en el mètode instrumental en fonètica. Ara bé, el seu disseny deixa les portes obertes a un tractament *variacionista*, si més no en la línia dels treballs recents de la disciplina emergent anomenada *sociofonètica*.

Pel que fa al corpus, les dades que presentem s'han obtingut a partir d'un corpus de 80 mots (40 amb la forma subjacent africada alveolopalatal sonora i 40 amb la sorda), l'obtenció dels quals s'ha efectuat a partir de la lectura de les frases en què se'ls havia inclòs. Aquestes frases, enunciatives amb sentit complet, anaven intercalades entre d'altres que no contenien cap

mot a analitzar. D'aquesta manera es garantia que l'informant no centrés l'atenció en la variable estudiada. Les paraules del corpus s'han situat al bell mig de la frase amb l'objectiu d'evitar les inflexions de la veu que produeixen l'atac i el nucli prosòdic. Cada frase ha estat emesa tres vegades per cada informant. Per a l'avinentesa que ens ocupa, la lectura ha estat imprescindible per poder garantir l'anàlisi dels factors lingüístics elegits (els veurem més endavant); cal tenir present que les variables objecte d'estudi (les africades alveolopalatals sorda i sonora en posició intervocàlica) tenen una freqüència d'aparició molt baixa.

S'ha volgut que els 20 informants utilitzats constituïssin una població prou homogènia quant a l'edat i a la formació: es tracta d'universitaris d'entre els 18 i els 35 anys. Des de la perspectiva de l'avaluació del canvi generacional (el *temps aparent* labovità), aquesta tria ens permetrà contrastar els resultats amb els de les generacions de més edat. I, en l'àmbit de la variació estilística, també es podrà contrastar l'alt grau de formalitat que imposa la lectura amb l'avaluació de dades més col·loquialitzants.

Els paràmetres acústics mesurats instrumentalment han estat els següents: la durada i l'ensordiment del segment oclusiu; la durada i l'ensordiment del segment fricatiu; i la relació perceptual entre la durada de la sonoritat en el segment oclusiu i la durada total de cada oclusiva.

D'altra banda, les variables lingüístiques controlades han estat: *a*) el nombre de síl·labes del mot; *b*) l'accent; *c*) el context vocàlic (per al barceloní: ['iCə, 'aCə, 'uCə, i'Ca, ə'Ca, u'Ca]; i per al lleidatà: ['iCε, 'aCε, 'uCε, i'Ca, ə'Ca, u'Ca]); i *d*) la freqüència d'ús del mot (el control d'aquest factor ha vingut donat per la necessitat de crear un corpus, on no tots els mots eren habituals per als informants, que permetés avaluar els factors lingüístics precedents). En el seu tractament estadístic hem utilitzat diferents ANOVA d'un factor (One-way ANOVA).

La “socialització” de l’experiment l’obté amb les variables *gènere* i contrast de les *varietats geogràfiques barcelonina i lleidatana*. La distribució dels informants n’ha previst 10 per a cada varietat, 5 masculins i 5 femenins. La inclusió d’aquestes dades en un treball variacionista tindria garantida la representativitat si convenim amb Silva-Corvalán (1991) que el nombre d’informants per casella no ha de ser inferior a cinc.

Les nostres dades constaten que les africades alveolopalatals sonores pateixen un procés acusat d’ensordiment que comença clarament per la part fricativa de l’africat, i que es veu afavorit per diversos factors: I) el dialecte (és més evident a Barcelona que a Lleida —figura 13—); II) el gènere (sobretot a les parlants femenines de Barcelona —figures 14, 15 i 16—); III) la freqüència d’ús de l’estímul (els mots amb una freqüència elevada d’ús causen més ensordiment a les oclusives en barceloní però menys en lleidatà —figura 14—); IV) el context vocàlic (l’ensordiment és més acusat quan l’africat es troba entre vocals baixes —figura 15—); i V) el nombre de síl·labes del mot (com més curt és el mot, més sord es pronuncia l’africat, a Barcelona especialment —figura 16—). La posició de l’accent no sembla afectar el procés d’ensordiment.

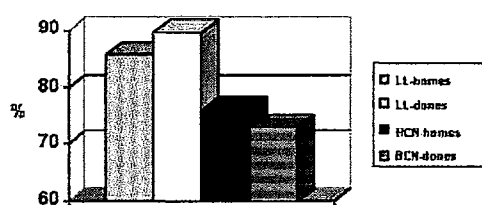


Figura 13. Percentatge de sonoritat de l’element oclusiu de l’africada sonora.

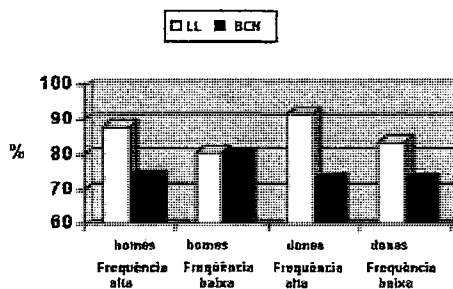


Figura 14. Percentatge de sonoritat del segment oclusiu de les africades sonores relacionat amb la freqüència d’ús de les paraules.

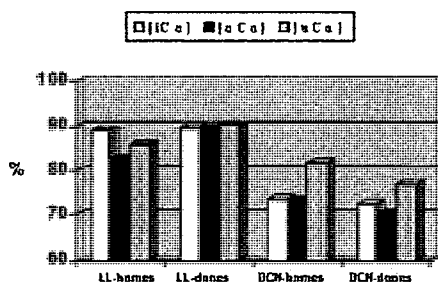


Figura 15. Percentatge de sonoritat del segment oclusiu de les africades sonores.

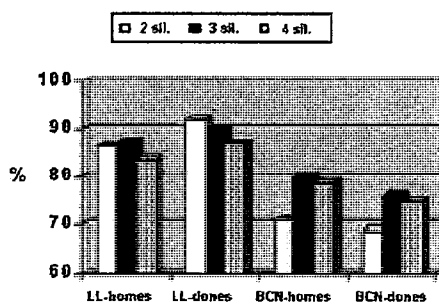


Figura 16. Percentatge de sonoritat de l'element oclusiu de les africades sonores.

Vegeu a la figura 17 uns sonogrames de la seqüència «desitja» que il·lustren diferents graus d'ensordiment a l'oclusiva.

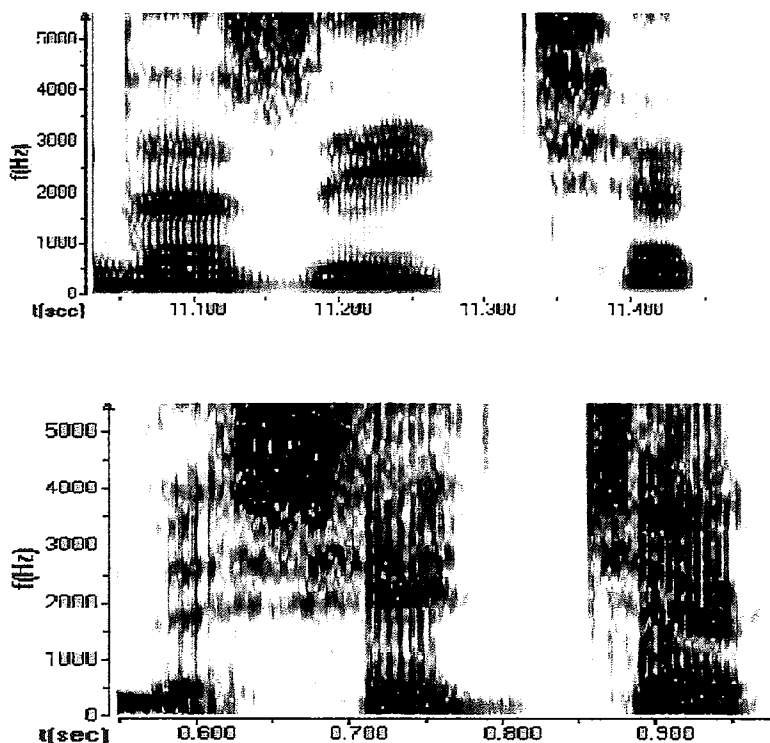


Figura 17. Sonogrames de la seqüència «desitja».

Sabem que la diferència entre els africats sords i sonors no és significativament productiva i, amb aquest benentès, podem entendre l'ensordiment dels africats com una simplificació del sistema; procés en el qual el lleidatà es mostra més conservador que el barceloní.

5.3. La prosòdia del català

A continuació presentem un exemple aplicat als fenòmens suprasegmentals, a la prosòdia. En el marc del projecte internacional AMPER (Atles Multimèdia de la Prosòdia de l'Espai Romànic) l'estudi geoprosòdic aplicat al català el desenvolupa el grup AMPERCAT del Laboratori de Fonètica de la UB. Pretén estudiar les principals variants prosòdiques (això és, d' F_0 , durada i intensitat) de la llengua per tal de contribuir a l'atles general de les llengües romàniques que es coordina internacionalment des del *Centre de Dialectologie de la Université Stendhal Grenoble-3* sota l'impuls del Dr. Contini i del Dr. Romano.

El projecte general és un projecte molt ambiciós pel que fa a la quantitat de material de què disposarà, d'informants tractats i de zones estudiades. El fet de confeccionar un atles sobre prosòdia ja és una novetat, però la seva aportació fonamental resideix a oferir-se als estudiosos i al públic en general en format multimèdia per tal que es puguin veure els gràfics fonètics de les realitzacions i es puguin escoltar a partir dels arxius de so penjats a la xarxa.

El projecte conté quatre tipus de corpus diferents:

a) un corpus formal, molt controlat que podem considerar parla de laboratori obtingut a partir d'alguna mena de lectura (bàsicament alfabètica o pictogràmica);

- b) un corpus induït a partir de supòsits hipotètics plantejats als informants;
- c) un corpus obtingut a partir de la tècnica de Map-Task; i
- d) un corpus constituït per converses no guiades absolutament espontànies.

Així, segons (Labov 1966), els estils obtinguts van de l'espontani (d) al formal (a) passant per l'acurat (b, c).

Els informants són de gènere masculí i femení, dos a cada punt d'enquesta, amb dos graus d'instrucció, fonamentalment. La frontera se situa en la possessió o no d'estudis superiors; en principi, en el primer cas hi haurà gent més jove que en el segon.

L'envergadura de frases obtingudes és absolutament enorme per a cada informant i la feina d'anàlisi, seguint el protocol establert pel Centre de Dialectologie, és estricta. Les frases obtingudes a partir de lectura contenen tres o quatre accents tonals (segons si la frase presenta expansió al subjecte o a l'objecte o no en presenta). Els mots són trisíl·labs aguts, plans i esdrúixols en diferents modalitats de parla. Per això les sessions d'enregistrament de dades solen fer-se en diverses jornades i en sessions força llargues, després d'haver-nos guanyat la confiança dels informants, amb els que prenem contacte després que algú conegut d'ells ens faci d'enllaç.

A AMPERCAT, com en la resta de grups que componen AMPER-ESPANYA, que també es coordinen des del Laboratori de Fonètica (Fernández Planas i Martínez Celdrán, inèdit), de moment s'ha començat la feina amb informants femenines d'una edat compresa entre els 25 i els 50 anys i sense estudis superiors. Pel que fa al català, ja s'han produït diversos fruits que ara mateix estan en premsa de revistes especialitzades o d'actes de congressos.

Presentem com a exemple de mostra de diversitat melòdica les corbes d'F0 característiques del barceloní, del tarragoní, del lleidatà i del tortosí en

les preguntes absolutes introduïdes per la partícula “que” que manifesten, a més a més, dislocació del subjecte al final de la frase (vegeu la figura 18). Aquestes corbes han estat obtingudes experimentalment a partir de la mitjana de tres repeticions de cada frase emeses per la informant de cada subdialecte (Martínez Celdrán i altres, en premsa b). Corresponen a la frase composta per Subjecte, Verb i Objecte plans i sense expansió en cap component:

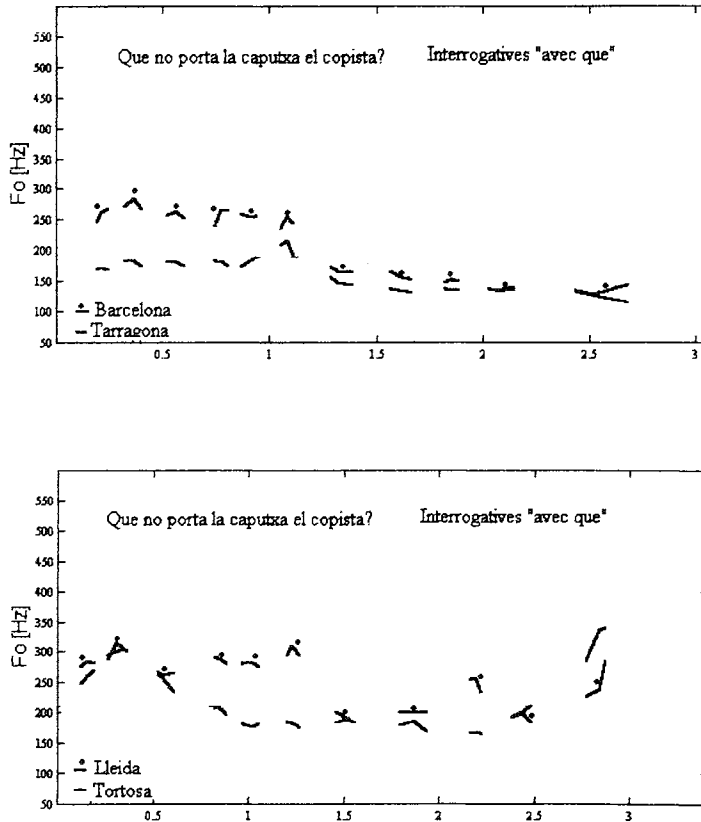


Figura 18. Corbes melòdiques a diferents dialectes.

En aquesta metodologia de treball, les frases es poden sintetitzar d'una manera peculiar segons la qual només en queden els elements prosòdics: F0, durada i intensitat. Aquest tipus de frases sintètiques s'han passat en forma de test perceptiu a grups de 30 estudiants en cadascuna de les localitats on es van enquestar les informants per comprovar si es reconeixia la modalitat (ja que es presentaven frases enunciatives, interrogatives absolutes sense "que" i interrogatives absolutes amb "que" (Martínez Celdrán i altres, en premsa a), o només els dos tipus d'interrogatives (Martínez Celdrán i altres, en premsa b; Fernández Planas i altres, en premsa a; Fernández Planas i altres, en premsa b; van Oosterzee i altres, en premsa; Carrera i altres, en premsa) i si es distingia millor el propi dialecte que un altre. Els resultats indiquen que la prosòdia és suficient per reconèixer la modalitat de parla en qualsevol cas i que cada grup d'informants reconeix millor les dades pertanyents al seu propi subdialecte.

Referències bibliogràfiques

- ALTURO, Núria; TURELL, M. Teresa (1990): «Linguistic Change in El Pont de Suert: The study of variation of /z/», *Language Variation and Change*, 2, 19-30.
- BOIX, Emili; PAYRATÓ, Lluís (1996): «An overview of catalan sociolinguistics and pragmatics (1989-1996)», *Catalan Review*, IX, 317-403.
- CARRERA-SABATÉ, Josefina (1993): «La pretònica inicial e- d'Alguaire», *Anuari de Filologia XVI*, C, 4, 37-72.
- _____ (1999): *L'alternança A/E al Segrià*, tesi doctoral, Universitat de Barcelona.
- CARRERA-SABATÉ, J.; FERNÁNDEZ-PLANAS, A. M.; MATAS-CRESPO, J. (1999): «Estudi contrastiu de les vocals mitjanes del català oriental central i nord-occidental», *Actes del 1r Congrés de Fonètica Experimental*, Tarragona, 143-149.

- CARRERA-SABATÉ, J.; FERNÁNDEZ-PLANAS, A. M.; ORTEGA-ESCANDELL, A. (2001): «Perception of Catalan medium vowels by Spanish speakers», *Proceedings of the 17th international congress on acoustics*, Roma, CD. IV.
- CARRERA-SABATÉ, J.; FERNÁNDEZ-PLANAS, A. (2003): «Posterior medium vowels differences in three catalan dialects», *Proceedings of the XVII International Congress of Linguists*, Praga, CD, S9.
- CARRERA-SABATÉ, J.; FREIXENET-ESTEVE, E. (2003) «[aw]liva o [o]liva? caracterització de l'alternança al segrià», *Noves SL*, Tardor 2003, pàg. 1-11. http://www6.gencat.net/llengcat/noves/hm03tardor/freixenet1_3.htm.
- CARRERA-SABATÉ, J.; VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A. M.; ROMERA BARRIOS, L.; ESPUNY MONSERRAT, J.; MARTÍNEZ CELDRÁN, E. (en premsa): «Les interrogatives al tortosí i al lleidatà. Un element diferenciador de subdialectes», *Estudios de fonética experimental*, XIII.
- COLE, Ronald; JAKIMIK, Jola (1980): «A model of speech perception», dins Cole, R. (ed.), *Perception and production of fluent speech*, Hillsdale, NJ: Erlbaum, 133-163.
- COLOMINA, Jordi (1984 [1992²]): «El parlar de la Canyada de Biar (L'Alt Vinalopó): un estudi de sociolingüística», *Miscel·lània Sanchis Guarner*, I, Barcelona: Publicacions de l'Abadia de Montserrat, 81-87.
- _____ (1985): «La diftongació de /ε/ i d'altres canvis vocàlics a la Canyada de Biar», dins *L'alacantí. Un estudi sobre la variació lingüística*, Alacant: Institut d'Estudis Gil-Albert, 94-109.
- ESCRIVÀ, Vicent (1993): «Aproximació sociolingüística al procés d'assimilació de la variable [ε] àtona final a Oliva», *Actes del Novè Col·loqui Internacional de Llengua i Literatura Catalanes*, III, Barcelona: Publicacions de l'Abadia de Montserrat, 121-136.
- FERNÁNDEZ PLANAS, A. M.; PRADILLA, M. À. (2004): «Características de las africadas alvolopalatales del catalán oriental central», *Actas del V Congreso de Lingüística General*, León: Arco Libros, 2002, 983-995.
- FERNÁNDEZ PLANAS, Ana Maria (2001): *Estudio electropalatográfico de la coarticulación vocálica en estructuras VCV en castellano*, tesi doctoral inèdita, UB.
- FERNÁNDEZ PLANAS, A. M.; MARTÍNEZ CELDRÁN, E. (2003): «El proyecto AMPER en España», *XXXIII Simposio de la Sociedad Española de Lingüística*, Girona (treball inèdit).

- FERNÁNDEZ PLANAS, A. M.; MARTÍNEZ CELDRÁN, E.; VAN OOSTERZEE, C.; SALCIOLI GUIDI, V.; CASTELLVÍ VIVES, J.; SZMIDT SIERYKOW, D. (en premsa): «Proyecto AMPER: estudio contrastivo de frases interrogativas sin expansión del *barceloní* y del *tarragoní*», *Actas del VI Congreso de Lingüística General* (2004), Universitat de Santiago de Compostela.
- FERNÁNDEZ, PLANAS, A. M.; CARRERA-SABATÉ, J.; MARTÍNEZ CELDRÁN, E.; VAN OOSTERZEE, C.; SALCIOLI GUIDI, V.; CASTELLVÍ VIVES, J.; SZMIDT SIERYKOW, D. (en premsa): «Interrogatives absolutes al *barceloní* i al *tarragoní* (estudi contrastiu)», *Estudios de fonética experimental*, XIII, Barcelona, Laboratori de Fonètica UB.
- FERNÁNDEZ-PLANAS, A. M.; CARRERA-SABATÉ, J. (2004): «Percepción de vocales medias del catalán oriental central y del catalán noroccidental por parte de oyentes catalanohablantes y castellanohablantes de Lleida», *Actas del V Congreso de Lingüística General*, 2002, León: Arco Libros, 982-973.
- FERNÁNDEZ-PLANAS, A. M.; CARRERA-SABATÉ, J.; MATAS-CRESPO, J. (2000): «Estudio perceptivo de vocales medias del catalán oriental central y del catalán noroccidental», *Actas del IV Congreso de Lingüística General*, Cadis, 1057-1068.
- FERNÁNDEZ-PLANAS, A. M.; CARRERA-SABATÉ, J.; PRADILLA-CARDONA, M. À. (2003): «A contrastive study of voiced alveolo-palatal affricates in the Catalan of Lleida and Barcelona», *Proceedings of the 15th icphs 2003*, Barcelona, CD, 937.
- FONTDEVILA, PALLARÉS, Maria Dolors i RECASENS, Daniel (1994): «The contact index method of electropalatographic data reduction», *Journal of Phonetics*, 22, 141-154.
- FOULKES, Paul; DOCHERTY, Gerard J. (1999): *Urban Voices*, London: Arnold.
- JONES, Daniel (1941): «Paul Passy», *Le maître phonétique*, 75, 30-39.
- LABOV, William (1966): *The Social stratification of English in New York City*, Washington D. C: Center for Applied Linguistics.
- _____ (1972): *Sociolinguistic patterns*, Philadelphia: University of Pennsylvania Press.
- LADEFOGED, Peter; MADDIESON, Ian (1996): *The Sounds of the World's Languages*, Oxford: Blackwell.
- MARTÍNEZ CELDRÁN, Eugenio (1991): *Fonética experimental: teoría y práctica*, Madrid: Síntesis.
- _____ (1998): *El sonido en la comunicación humana*, Barcelona: Octaedro.
- MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A. M.; SALCIOLI GUIDI, V.; CARRERA-SABATÉ, J.; ESPUNY MONSERRAT, J. (en premsa a): «Approche de la prosodie du

- dialecte de Barcelone», *Geolinguistique. Hors série*, Grenoble: Centre de Dialectologie, Université Stendhal-Grenoble III.
- MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A. M.; CARRERA-SABATÉ, J.; ESPUNY MONSERRAT, J. (en premsa): «Approche de la carte prosodique dialectale de la langue catalane à Catalunya», *Geolinguistique. Hors série*, Grenoble: Centre de Dialectologie, Université Stendhal-Grenoble III.
- MIER, Jeanne Z. (1986): «Estudi sociolingüístic de certs aspectes de la llengua catalana», *Treballs de Sociolingüística Catalana*, 6, 33-112.
- MONTOYA, Brauli (1993): *Variabilitat i prestigi en el català de Valls i de l'Alt Camp*, Valls: Institut d'Estudis Vallencs.
- _____ (1989): «Estratificació de la variació lingüística a Petrer», dins Gimeno, F.; Montoya, B. (ed.): *Sociolingüística*, València: Publicacions de la Universitat de València, 67-95.
- _____ (2000): *Els alacantins catalanoparlants: una generació perduda*, Barcelona: IEC.
- MORENO, Francisco (1990): *Metodología sociolingüística*, Madrid: Gredos.
- PLA, Joan (1995): «L'obertura de [ə] a Barcelona: el xava i altres varietats», dins Turell, M. T. (ed.), *La sociolingüística de la variació*, Barcelona: PPU, 165-219.
- PLAZA, Carme (1995): «Lleialtat lingüística, edat i nivell educatiu. La *e* posttònica a la Conca de Barberà», dins Turell, M. T. (ed.) *La sociolingüística de la variació*. Barcelona: PPU, 117-138.
- PRADILLA, Miquel Àngel (1996): *El Baix Maestrat: una cruïlla fonètica*, Benicarló: Associació Cultural Alambor.
- _____ (1993): *Variació i canvi lingüístic en curs al català de transició nord-occidental/valencià*, tesi doctoral, Tarragona: Universitat Rovira i Virgili.
- _____ (1995): «El desafricament prepalatal en el català de transició nord-occidental/valencià», dins Turell, M. T. (ed.) *La sociolingüística de la variació*. Barcelona: PPU, 53-116.
- _____ (1997): «Sociolingüística quantitativa i anàlisi qualitativa de variables foneticofonològiques: a propòsit del desafricament prepalatal», *Estudios de Fonética Experimental*, VIII, 207-251.
- _____ (2001a): «La sociolingüística de la variació: aproximació metodològica (I)», *Noves SL*, Hivern-Primavera 2001, 1-7, http://cultura.gencat.es/llengcat/noves/hm01hivern-primavera/metodologia/pradilla1_7.htm.

-
- _____ (2001b): «La sociolingüística de la variació: aproximació metodològica (II)», *Noves SL*, Estiu 2001, 1-9, http://www6.gencat.net/llengcat/noves/hm01estiu/metodologia/pradillab1_9.htm.
- _____ (2002): «La variació fònica en la llengua catalana: inventari i avaluació metodològica», *Noves SL*, Tardor 2002, 1-8. http://www6.gencat.net/llengcat/noves/hm02tardor/metodologia/pradilla1_8.htm
- PRADILLA-CARDONA, M. À.; CARRERA-SABATÉ, J.; FERNÁNDEZ-PLANAS, A. M. (en premsa): «De fonètica contrastiva: els africats alveolopalatals del lleidatà i del barceloní» *Actes del 13è Col·loqui Internacional de Llengua i Literatura Catalanes*, Girona.
- RECASENS, Daniel (1991): *Fonètica descriptiva del català*, Barcelona: IEC.
- ROUSSELOT, Jean Pierre (1892): *Les modifications phonétiques du langage étudiées dans le patois d'une famille de Cellefrouin*, Paris: H. Welter.
- ROUSSELOT, Pierre Jean (1924): *Principes de phonétique expérimentale, Nouvelle édition revue et corrigée*, Paris: H. Didier, 2 vols.
- SALCIOLI, Valeria (1981): «Anàlisi espectrogràfica de les prepalatals del català parlat a Barcelona», *Anuario de Filología*, Universitat de Barcelona, 533-550.
- SILVA-CORVALÁN, Carmen (1989): *Sociolingüística. Teoría y análisis*, Madrid: Alhambra.
- VAN OOSTERZEE, C.; FERNÁNDEZ PLANAS, A. M.; ROMERA BARRIOS, L.; CARRERA SABATÉ, J.; ESPUNY MONSERRAT, J.; MARTÍNEZ CELDRÁN, E. (en premsa): «Proyecto AMPER: estudio contrastivo de frases interrogativas sin expansión en *tortosí* y en *lleidatà*», *Actas del VI Congreso de Lingüística General*, Universidad de Santiago de Compostela.
- WRIGHT, Sandra. (2001): «Sociophonetics», R. Mesthrie (ed.): *Concise Encyclopedia of Sociolinguistics*, Amsterdam: Elsevier, 391-396.

Els (sub)corpus lingüístics
del *Diccionari de la pronúncia catalana*

Joan Julià-Muné i Imma Creus
(Universitat de Lleida)

«Verba non volant si phonetica scripta manent»

1. Introducció: objectius i evolució del projecte

El *Diccionari de la pronúncia catalana* —DPC, a partir d'ara— és un dels projectes que desenvolupa el Grup de Recerca en Fonètica del Laboratori de Fonètica Experimental “Pere Barnils”, adscrit al Departament de Filologia Catalana i Comunicació de la Universitat de Lleida, institució que finança el projecte. El grup forma part de la xarxa temàtica «La variació lingüística. Dialectologia, sociolingüística i pragmàtica» (2002/XT/00001), coordinada pel Dr. Joaquim Viaplana de la Universitat de Barcelona.

En termes generals, els objectius del DPC són: *a)* elaborar un diccionari de pronúncia *modern*, pel que fa a l'aspecte metodològic, i al màxim *actualitzat* possible en relació amb el corpus; *b)* contribuir a la descripció

fonètica de les principals varietats del català (nord-occidental, central, valencià i balear); c) elaborar, per primer cop en català, un diccionari de pronúncia tant en suport tradicional (paper) com en suport electrònic (CD-ROM). Més específicament, el DPC pretén aportar dades relatives a l'ús oral actual de la llengua catalana, amb alguna orientació de tipus ortoèpic; proposar l'adaptació fònica al català de mots procedents d'altres llengües, alhora que es familiaritza l'usuari amb la pronúncia originària de mots d'origen forà; actualitzar el cabal lèxic de referència —en relació amb el d'obres similars en l'àmbit català— en funció de les necessitats comunicatives i expressives orals contemporànies, tot incorporant neologismes, marques comercials, topònims, antropònims, sigles, etc. En definitiva, l'objectiu no és elaborar exclusivament una obra útil per als lingüistes interessats en el component oral de la llengua (estudis de fonètica descriptiva, dialectologia, lingüística contrastiva, etc.) sinó sobretot elaborar una eina aplicada útil tant per als professionals de la llengua en general (locutors de ràdio i televisió, actors de doblatge, professors, etc.) com per als usuaris que en un moment determinat, ja sigui per responsabilitat professional o interès personal, es preocupin pel component fonètic de la llengua.

En un recorregut breu pel que ha estat l'evolució del projecte ens hauríem de remuntar a l'any 1990 quan de fet se'n comença a gestar la idea. En aquell any conflueixen un seguit de publicacions que tenen molt a veure amb la fonètica. A casa nostra, per una banda, es publiquen les primeres propostes ortoèpiques oficials amb el títol *Proposta per a un estàndard oral de la llengua catalana. I. Fonètica* (Institut d'Estudis Catalans —IEC d'ara endavant—)¹ i, per l'altra, s'edita el primer diccionari de pronúncia

¹ Recordem que l'any 1992 es publica l'obra homònima relacionada, aquest cop, amb la morfologia. D'aleshores ençà s'han anat fent reedicions d'aquests opuscles que, més enllà de retocs puntuals, no n'han modificat el contingut. Tampoc no s'han editat les propostes referents a la sintaxi i al lèxic.

catalana, el *Diccionari ortogràfic i de pronúncia* de Jordi Bruguera.² Paral·lelament, apareix el *Longman Pronunciation Dictionary* (LPD) de John Wells,³ un diccionari de pronúncia de l'anglès que ha estat un referent essencial per a la nostra tasca.

Pel que fa al desenvolupament pròpiament dit del nostre projecte, podem establir diferents etapes que evidencien tant la seva evolució com l'establiment de nous objectius. Així, en la primera fase (1990-1994) es recullen únicament dades fonètiques —mitjançant enquestació— en l'àmbit nord-occidental (varietat lleidatana) corresponents a la pronúncia de lèxic comú. En una etapa posterior (1995-1998) s'amplia el corpus, a partir del DIEC (1995), alhora que s'afegeix lèxic patrimonial i s'incorpora la pronúncia central. Durant la darrera fase (1999-2003) és sens dubte quan més s'ha incrementat el nombre d'entrades, si tenim en compte que s'hi ha afegit lèxic d'especialització (neologismes, terminologia, etc.), onomàstica pròpia i forana, gentilicis, onomatopeies, sigles, acrònims, manlleus, marques comercials, etc. L'increment considerable d'entrades en aquest cas respon al desig de satisfer tant com sigui possible les necessitats expressives dels parlants actuals, ja que possiblement les respostes a la pregunta *què busca l'usuari en un diccionari de pronúncia* són tan diverses com ho són les necessitats que planteja la comunicació oral. En l'etapa actual del DPC es treballa amb una ampliació de l'equip amb l'objectiu d'incorporar-hi la pronúncia valenciana i balear,⁴ de manera que el diccionari pugui esdevenir

² Disposem d'una edició actualitzada d'aquesta obra, apareguda el maig de 2004, que, tot i que n'amplia el repertori lèxic respecte a la primera, elimina, per exemple, informació fonètica i la informació relacionada amb les variants morfològiques o amb els verbs conjugats que apareixen com a annex.

³ També en aquest cas disposem d'una segona edició del diccionari, concretament de l'any 2000. A banda d'incorporar-hi noves entrades, l'aspecte més nou són els gràfics que indiquen percentualment la varietat de pronúncies d'alguns mots seleccionats.

⁴ Tot i que la col·laboració encara és incipient i no definitiva, agraïm el treball desenvolupat fins ara per Jesús Jiménez, Vicent Garcia Perales i Nicolau Dols.

un referent per a un major nombre de parlants dels diferents dominis dialectals del català. La finalització del projecte, així com l'edició posterior en paper, en principi, està prevista per al 2006 i l'edició en format electrònic per al 2008.

2. El DPC com a corpus lingüístic

És evident que, en essència, el DPC és —o aspira a ser— un diccionari de pronúncia, és a dir, una recopilació de lèxic —necessàriament selectiva— amb les corresponents transcripcions fonètiques de cada entrada. La part que correspon a les transcripcions, diverses en funció de la variació fonètica intradialectal i interdialectal, és la que atorga a l'obra la categoria de diccionari de *pronúncia*; ara bé, si ens fixem en el diccionari de manera integral, i això òbviament inclou el cabal lèxic de referència i sobretot el procés de selecció lèxica a què ens referíem, aleshores pensem que podríem considerar el DPC com una mostra de *corpus lingüístic*. Tot seguit aportem elements d'anàlisi que ens permetin justificar aquesta idea.

D'entrada, cal tenir en compte que, tot i que possiblement quan parlem de corpus lingüístics pensem sobretot en compilacions de textos, la definició de *corpus* admet altres possibilitats. David Crystal defineix el terme com «una col·lecció de dades lingüístiques, ja siguin textos escrits o una transcripció de parla enregistrada, que pot ser utilitzada com a punt de partida per a la descripció lingüística o com un mitjà per verificar hipòtesis sobre una llengua» (2000: 149-150).⁵ Un corpus convé que sigui representatiu d'un estat o nivell de llengua determinat, però el cert és que,

⁵ Traduïm directament de l'edició espanyola (2000) de l'obra *A Dictionary of Linguistics and Phonetics* (1980 [1997]).

per raons òbvies, és sempre limitat. En aquest sentit es parla de *corpus textual* quan es tracta d'un conjunt de documents complets i de *corpus de referència* quan es tracta de fragments textuais (Blecua 1996: 16). Fins aquí, doncs, hauríem determinat els aspectes que tenen a veure amb les *característiques* i la *tipologia* d'un corpus. Si apliquem els comentaris anteriors al DPC, el primer que podem afirmar és que, efectivament, podem considerar-lo un exemple de corpus en la mesura que és una col·lecció de dades que combina la representació ortogràfica (entrades) amb la fònica (transcripcions fonètiques) que, en un nombre elevat, provenen de material enregistrat. Les transcripcions, a més, reflecteixen un estat de llengua actual que obre les portes a una possible anàlisi lingüística des de diferents perspectives. Tipològicament, en una primera fase —la que correspondria a l'edició en paper— el DPC es pot considerar fonamentalment un corpus escrit, ja que el referent oral és representat de manera gràfica mitjançant la simbologia fonètica; tanmateix, la segona fase, corresponent a la digitalització del contingut del diccionari, ens permetrà parlar ja pròpiament d'un corpus oral. A més, el suport informàtic afavorirà l'ús de l'obra com a recurs lingüístic i la seva explotació com a tal. Pel que fa a les característiques concretes del DPC, podríem dir que és un corpus format per unitats lèxiques que a la vegada s'agrupen en diferents subcorpus lèxics.

Prèviament, però, al que acaba sent un corpus hi ha sempre l'etapa de constitució d'aquest, que cal planificar conscientment no només en funció dels objectius del projecte, sinó sobretot perquè el resultat final, és a dir, el corpus en si mateix, assoleixi garanties de representativitat. Llisterra (1996) sintetitza —partint de Carré (1991, 1992)— les fases principals en la constitució d'un corpus:⁶ *a*) definició del corpus (disseny del contingut), *b*)

⁶ Tot i que Llisterra es refereix a corpus lingüístics orals, considerem que els diferents apartats que estableix es poden fer extensius a altres tipus de corpus, amb les modificacions oportunes segons els objectius inicials i les possibles aplicacions posteriors del corpus.

obtenció de les dades, c) preparació de les dades, d) gestió i tractament de les dades, i e) disseminació de les dades (documentació i suport). Traslladat novament aquest esquema al DPC, la definició del corpus podríem dir que s'ha anat redefinint en la mesura que s'ampliaven els objectius del projecte i augmentava el volum de les entrades, de manera que l'obtenció de dades també s'ha diversificat. D'una banda, i en relació amb el corpus de referència gràfic —és a dir, amb les entrades—, les dades s'han obtingut bàsicament a partir de fonts lexicogràfiques descriptives —ja siguin enciclopèdiques o monogràfiques especialitzades—, encara que, més puntualment, també s'han tingut en compte reculls lèxics de tipus dialectal, premsa, mèdia orals, etc. D'una altra banda, per a les transcripcions, sempre que ha estat possible, s'han obtingut les dades mitjançant enquestació directa⁷ de parlants nadius dels àmbits dialectals corresponents. Pel que fa a la preparació o presentació de les dades, les decisions afecten sobretot la transcripció fonètica, ja que des del punt de vista ortogràfic únicament hem hagut de plantejar-nos algunes grafies relacionades amb mots patrimonials o neologismes no recollits en les obres lexicogràfiques consultades, o resoldre casos de divergència gràfica entre diferents fonts. Quant a la transcripció fonètica, hem optat per un tipus de transcripció funcional, de base segmental, utilitzant la simbologia de l'AFI,⁸ i simplificant al màxim la representació de les variants a fi d'evitar repeticions excessives i problemes d'interpretació per part de l'usuari. Finalment, la gestió i el tractament de

⁷ Aquest procediment es va poder seguir de manera sistemàtica, i amb un disseny de treball empíric força definit sociolingüístament, sobretot en la primera fase del projecte, en què es treballava amb lèxic bàsic. Posteriorment ha estat més complicat, ja que el tipus de lèxic —menys freqüent i fins i tot especialitzat— no afavoria les respostes espontànies per part dels informants enquestats. Conseqüentment, algunes de les dades s'han obtingut mitjançant una enquestació menys planificada i més selectiva pel que fa a la tria d'informants. En altres casos, no ha estat possible disposar de dades empíriques més enllà de la competència lingüística activa —el que pronunciem— i passiva —el que sentim pronunciar habitualment— dels membres de l'equip.

⁸ Ens servim del sistema SIL Doulos IPA-Regular per a Apple Macintosh.

les dades, així com la seva difusió, com ja hem dit anteriorment, s'ha de plantejar en dues fases: la que correspon a l'edició de l'obra en suport tradicional i la que correspon a l'edició electrònica. En aquests moments es treballa en la primera.

En els següents apartats ens centrarem en els aspectes que tenen a veure fonamentalment amb el disseny del contingut del DPC o, dit d'una altra manera, aprofundirem en les característiques dels diferents subcorpus del diccionari. Per tant, prescindirem de comentaris referents a altres fases del projecte perquè o bé ja han estat tractades en altres treballs (Julià-Muné & Creus (2003), Julià-Muné (en premsa)) o bé ara per ara no ens semblen prioritàries.

3. Els subcorpus lèxics del DPC

En aquest bloc, a banda de llistar els principals subcorpus lèxics del diccionari, exposarem alguns dels criteris que hem adoptat en la incorporació de determinat lèxic i en la transcripció fonètica que hi consta. De totes maneres, cal advertir que algunes de les consideracions que s'exposen en aquest apartat, així com les característiques d'algunes entrades, poden veure's modificades, si tenim en compte que es tracta d'un projecte en procés d'elaboració.

3.1. Vocabulari comú

La inclusió de lèxic comú en el DPC és sens dubte indispensable per diferents raons. La més òbvia de totes es fonamenta en el fet que aquest tipus de vocabulari és la base de qualsevol diccionari descriptiu i, naturalment, el DPC encaixaria en aquesta categoria. Des del punt de vista

fonètic, en principi, hom podria considerar que el lèxic comú —s’hi inclouria el vocabulari bàsic— és el que menys dubtes de pronúncia planteja. Tanmateix, tot i acceptant aquesta premissa —discutible, si més no—, és evident que la informació fonètica d’aquestes unitats aporta dades interessants per al contrast dialectal, per exemple, alhora que es converteix en una eina imprescindible per als aprenents de català.

Pel que fa a la selecció lèxica, el criteri que hem seguit ha estat buidar de manera sistemàtica el contingut dels diccionaris generals de la llengua catalana més recents: en una primera fase el diccionari normatiu de l’IEC (DIEC, 1995) i, en una segona, el *Gran diccionari de la llengua catalana* (GDLC) d’Enciclopèdia Catalana (1998). Aquest cabal lèxic, que és el més nombrós del DPC, s’ha completat amb vocabulari complementari que hem extret d’obres enciclopèdiques actualitzades (*Gran Enciclopèdia Catalana*, *La Gran Enciclopèdia en català* (2004)) i diccionaris/reculls lèxics diversos de tipus monogràfic (Bover (²1997), Bruguera (²2004), Camps & Botet (2001), Domènech (coord., 2002), Lacreu (ed., 2001), Ortega (2003), Paloma & Rico (2000), Pomares (1997) i Verdaguer (1999), entre altres).

Formalment, hem apostat per algunes modificacions respecte a les fonts emprades. En primer lloc, pel reagrupament semàntic i morfològic; és a dir, que hem unificat en una sola entrada possibles divergències semàntiques —que en el nostre cas ens interessen poc— i morfològiques —s’assenyalen les possibles variants— de mots que no presenten ni variació ortogràfica ni fònica. Així, si en un diccionari general (GDLC, posem per cas) trobem dues entrades per al mot ‘amant’, una amb tres accepcions semàntiques i tres variants morfològiques (*adj.*, *m.* i *f.*), i un altra amb dos significats diferents i una indicació morfològica (*m.*), en el DPC hi constaria una sola entrada (*amant*) amb les indicacions *adj.*, *m.* & *f.*, perquè la pronúncia seria la mateixa. En segon lloc, hem considerat convenient incloure informació

complementària, com ara la relativa a algunes formes de plural o de la conjugació verbal, per la senzilla raó que, més enllà dels canvis ortogràfics, aporten informació fonètica rellevant. Vegem-ne algun exemple:

- (1) **assaig** *m.* [a'satʃ̃ | an'satʃ̃ (!)] || [ə'satʃ̃ | [ən'satʃ̃ (!)]
assaigs/assajos *pl.* [a'satʃ̃ \ a'sa.ʒos | 'saʝ.] || [ə'satʃ̃ \ ə'saʒus]

En relació amb la informació fonètica, en el DPC intentem reflectir tant dades referents a la variació dialectal (N-O, nord-occidental; V, valencià; C, central i B, balear) com a la variació funcional (p. e., pronúncia de diftongs o hiats en funció de la velocitat d'emissió o del registre), amb limitacions evidents, la primera de les quals és la impossibilitat d'abraçar totes les manifestacions de la variació fonètica. El resultat de tot plegat és la incorporació de variants fonètiques, sempre que escaigui, d'acord amb els usos orals contemporanis: forma estàndard, que correspon generalment a la primera transcripció i, de vegades, també a la segona (es tractaria de formes secundàries); formes col·loquials reservades per a ús informal o formes clarament desaconsellables —‘no recomanables’ utilitzant la terminologia de l'IEC—, que en els casos més rellevants assenyallem amb el símbol (!). Vegem-ne alguns exemples:

- (2) **abarrocament** *m.* N-O [a.βa.ro.ka'men] V [aβaroka'ment | 'men] ||
 C [əβərukə'men] B [əβərokə'ment | ru | cə']

pressió *f.* N-O, V [pre.si'o | 'sjo | .zjo (!)] || C, B [prəsi'o | 'sjo | zjo (!)]

3.2. Terminologia especialitzada

Respecte a la terminologia especialitzada, cal dir que en una primera fase del projecte es va excloure del DPC, bàsicament per dues raons. En primer lloc, perquè es tracta d'un tipus de lèxic d'ús restringit i, en segon lloc, per la dificultat que suposava obtenir les dades a partir d'enquestació amb la metodologia que s'havia aplicat en la recollida del lèxic comú. Posteriorment, però, es va considerar que, justament els aspectes que inicialment ens havien fet optar per l'exclusió, es convertien en arguments sòlids per a la seva inclusió. Dit d'una altra manera, el fet que es tracti d'un vocabulari poc utilitzat —en la seva forma catalana— fins i tot pels professionals que integren els àmbits on l'ús de la terminologia és imprescindible (els àmbits literari, científic, jurídic, tecnològic, artístic, etc.), entre altres raons, perquè per formació —o per deformació— solen recórrer a les formes castellanes o angleses equivalents, és el que acaba generant nombroses dificultats de pronúncia. Conseqüentment, vam considerar que, incloent aquest tipus de lèxic en el DPC, convertíem el diccionari en una eina de consulta útil per a professionals, mínimament sensibilitzats, de la medicina, del dret, de la bioquímica, dels mèdia, etc. Cal fer notar, a més, que la incorporació de lèxic terminològic és una constant en les noves edicions de diccionaris descriptius (cf. el DIEC (1995) amb el GDLC (1998), o la primera edició del diccionari de Bruguera (1990) amb la segona (2004) o amb el diccionari de Lacreu (ed., 2001)).

Les obres que acabem d'esmentar constitueixen algunes de les fonts utilitzades per al buidatge lèxic. N'hi podríem afegir alguna altra (*La Gran Enciclopèdia en català* (2004), Doménech (coord., 2002)), tot i que hem d'admetre que no hem fet una selecció lexicogràfica exhaustiva d'obres que en aquest cas necessàriament s'haurien de buidar. Ens referim als diccionaris

terminològics en català, que ja comencen a ser abundants. Aquesta tasca, però, depassaria els objectius del projecte i sobretot l'abast del DPC.

Del que acabem de comentar es desprenen, doncs, les principals característiques de la informació corresponent a aquestes unitats lèxiques. Des del punt de vista tipogràfic, les entrades apareixen sense cap marca diferencial, perquè, com dèiem, generalment es tracta de formes ja incorporades en els diccionaris descriptius actuals, fins i tot encara que siguin d'origen forà —adaptades o no a l'ortografia catalana. Quan es tracta de mots no adaptats, apareixen en lletra cursiva. Pel que fa a la informació fonètica, bàsicament s'assenyala la variació dialectal —la variació funcional és més excepcional per l'ús que es fa d'aquest lèxic— i, de tant en tant, s'especifiquen recomanacions d'ús, marcant el que es consideraria que són pronúncies no recomanades (!), malgrat que siguin molt freqüents (desplaçaments d'accents, ensordiments, pronúncies interferides, etc.). En els casos de mots no adaptats, es transcriuen també en la forma originària. Vegem-ne una mostra:⁹

- (3) *aeròstat m.* [a.e'rɔs.tat |.tət |'tat (!)] || [æ'e'rɔstət |'tat (!)]
attrezzo m. [a'tre.ðzo] *it.* [at'trettso]

3.3. Lèxic propi i exclusiu (?) de cadascuna de les varietats

El conjunt de mots propis d'una varietat lingüística, el que en sentit genèric anomenem *dialectalismes*, acostuma a passar pel sedàs de la

⁹ En la relació d'exemples d'aquest article, si no hi figura la indicació de la varietat fonètica (NO, V, C, B), s'entén que la primera és nord-occidental i la segona (a la dreta de la doble barra vertical) la central. Vegeu, com a complement, els *Annexos I i II*.

selecció restrictiva en els diccionaris generals de la llengua catalana, sense que el lector en conegui exactament els criteris d'admissió/exclusió. D'altra banda, hi ha diccionaris que opten o bé per assenyalar la procedència dialectal de les entrades en qüestió —seria el cas del GDLC, que afegeix *dial* a les entrades corresponents— o bé per eliminar-ne qualsevol marca —és el procediment que generalment utilitza el DIEC—, perquè sovint són mots que ja han perdut la seva categoria de dialectalismes i s'utilitzen arreu.

La decisió d'afegir lèxic patrimonial en el DPC s'explica pel fet que un diccionari que es fonamenta en el reconeixement de la variació fonètica catalana —si més no, de les principals varietats dialectals— cal també que opti pel reconeixement lèxic dialectal,¹⁰ incloent-hi mots que, com dèiem, no apareixen als diccionaris generals. De totes maneres, aquesta no és una tasca senzilla i, a més, implica prendre una sèrie de decisions —tampoc fàcils de resoldre—, algunes de les quals es poden deduir del que hem comentat a l'inici d'aquest apartat.

El principal problema deriva sens dubte de la dificultat d'establir uns límits clars entre el que és lèxic propi i/o exclusiu d'una àrea dialectal. En primer lloc, perquè sovint es tracta de lèxic compartit per més d'un territori. En segon lloc, perquè no sempre es tracta de dialectalismes pròpiament lèxics sinó que s'haurien d'interpretar com a tals des del punt de vista semàntic (un mateix mot adquireix determinades accepcions en un territori) o morfològic (p. e., *l'amanit*, *la fred*, etc.) (vg. Creus, González & Julià i Muné, 1998). Hi hauríem d'afegir encara les dificultats que afecten l'ortografia i la relació d'aquesta amb la variació fonètica. Així, hi ha mots que han pervingut bàsicament per transmissió oral i que, per tant, plantegen dubtes a l'hora de proposar-ne una grafia (p. e., el mot que s'utilitza amb el

¹⁰ En aquest cas, però, la variació dialectal abraçaria zones més concretes. La intenció és incorporar mots ribagorçans, pallaresos, tortosins, tarragonins, gironins, etc.

significat de ‘sabatilla esportiva’ en algunes contrades de l'àrea nord-occidental i comarques de Tarragona: *queds/quets/quecs* (?) < Keds® (cf. *vamba* < Wamba®)). D'altres vegades cal valorar si es tracta d'un mateix mot amb variants fonètiques o si es tracta de mots diferents (p. e., *genoll* vs. *jonoll*, *tardor* vs. *tardaor*, etc.). Aquestes, per tant, són les principals decisions que cal prendre en un tipus d'obra com la nostra. Vegem-ne un exemple, amb algunes de les possibles opcions:

- (4) a) *genoll m.* [dʒe'noʎ | dʒi' | dʒo'] || [ʒə'noʎ | ʒu' | ʒu'noj]
 b) *genoll m.* [dʒe'noʎ | dʒi'] || [ʒə'noʎ]
 b') *jonoll m.* [dʒo'noʎ] || [ʒu'noʎ | 'noj]

Complementàriament, un dels dubtes que se'ns planteja —com d'altres, encara no resolts— és la idoneïtat d'incloure, en el DPC, la transcripció fonètica només en una o dues varietats dialectals o en totes de forma sistemàtica. Com a argument a favor de la primera opció, hi hauria el criteri de la coherència, de manera que semblaria poc adequat incloure-hi la pronúncia relativa a una àrea dialectal on el mot en qüestió no s'utilitza, tot i que, com hem dit, prèviament caldria establir el seu abast dialectal exacte. Com a argument a favor de la segona opció, hi hauria el que podríem anomenar necessitats expressives compartides, ja que quan s'emptra un mot poc corrent en l'àmbit dialectal propi (en llegir un text, posem per cas) el que se sol fer és adaptar-lo a les característiques fonètiques particulars del dialecte en qüestió.

La suma de les dificultats esmentades, doncs, fan imprescindible la consulta de fonts diverses i resulta gairebé obligatòria la recollida de dades ‘sobre el terreny’. Aquesta és la raó per la qual al costat de diccionaris generals hem emprat diccionaris descriptius —el més útil en aquest aspecte continua sent el DCVB—, vocabularis/diccionaris de tipus dialectal (Vený

(1958/59, 1960, 1979, 1984), Recasens (1985), Coll (1991), Sistac (1993, 1998), Turull (1996), Casellas & Casellas (2003), Ortís (2002) i Massana (2004), entre altres), alhora que hem aprofitat la nostra experiència com a lectors (de premsa o obres literàries que recullen lèxic patrimonial) i com a parlants de diferents dialectes.

3.4. Sigles i acrònims

A diferència del lèxic que integra algun dels subcorpus anteriors, en general, les sigles i els acrònims s'utilitzen de manera més aviat restrictiva, tot i que n'hi ha un nombre reduït que formen part de la quotidianitat d'una bona part de la població (Renfe, TV3, IVA, Muface, Sida, Banesto, BBVA, ONCE, PSOE, ERC, CiU, etc.). Per contra, n'hi ha un nombre força més elevat que difícilment es poden descodificar sense consultar algun dels diccionaris especialitzats que s'han editat per a les diferents llengües i que, sortosament, també podem consultar en el cas del català (Mestres & Guillén (²2001)). Per tant, com a criteri per a la incorporació d'aquest subcorpus al DPC podríem dir que hem optat per l'equilibri: sigles i acrònims d'ús corrent al costat d'altres de menys freqüents, i d'origen català juntament amb d'altres d'origen forà.

Més enllà de la semàntica d'aquestes unitats lèxiques que, com diem, tenim més o menys satisfeta amb l'edició de diccionaris diversos, no és gaire la informació que podem trobar relativa a la seva pronúncia, atesa la manca de reculls lexicograficofònics.¹¹ En qualsevol cas sí que podem extreure algunes de les orientacions que contenen obres com ara el *Llibre*

¹¹ Els diccionaris de pronúncia del català de què disposem fins ara (vg. § 1) no incorporen ni sigles ni acrònims a diferència, per exemple, de l'LPD de Wells.

d'estil de TV3 (1995) o Castellanos (2004), entre altres. Com a principals problemes que afecten la pronúncia de sigles i acrònims hi ha: *a)* els relacionats amb la lectura de la pronúncia lletrejada o acronímica, *b)* les neutralitzacions del sistema àton en aquells dialectes en què són pròpies, *c)* la pronúncia de hiat i/o diftong, i *d)* la pronúncia de sons consonàntics en determinats contextos. En el primer cas sovint el resultat són interferències fonètiques a partir d'una mala lectura: BBC [biβi'si], llegit 'a l'anglesa' o [beβe'the] 'a la castellana', en lloc de [beβe'se] com correspondria en català; ICE [i'the] en lloc de ['ise]/['isə], etc. Pel que fa a la resta de casos, tanmateix, les orientacions són menys clares i obren la possibilitat a pronúncies diverses. Així, per exemple, es poden neutralitzar o no les *a*, *e* i *o* àtones d'acrònims —en els dialectes que correspongui— sempre que no s'hagin lexicalitzat, ja que aleshores la neutralització és obligatòria (Renfe ['remfe]/['remfə]/['remfə], però *sida* ['siðə]; Talgo ['taɫɣo]/['taɫɣu]). També és opcional la pronúncia diftongada o amb hiat en sigles com ara IEC ([j'jek]/[i'ek]), encara que en certa manera depèn de la lectura que se'n faci (*el IEC*/*l'IEC*, respectivament). En darrer terme, tampoc no hi ha un criteri únic a l'hora d'admetre pronúncies del tipus: ESO ([e'so\u]/[e'zo\u]), NASA ([nasa\ə]/[naza\ə]).¹²

La informació fonètica que inclou el DPC amalgama informació relativa: *a)* al tipus abreviació —es dedueix per la transcripció si es tracta d'una sigla o d'un acrònim, encara que en aquest darrer cas pròpiament s'hauria de parlar d'una subclasse dins les sigles—; *b)* a la pronúncia més corrent; *c)* a les tendències fòniques que caldria evitar —interferències—; i *d)* a l'origen d'aquest tipus d'unitats —en els casos que corresponen a

¹² Castellanos (2004: 113) és del parer que és preferible pronunciar sorda la 's' entre vocals (i sense reduir les vocals *e* i *o*).

llengües foranes s'afegeix la transcripció en la llengua originària. De totes maneres, cal precisar que no totes les entrades contenen la mateixa informació i, a més, aquesta no sempre s'explicita. Vegem-ne alguns exemples:

(5) **AFAC** *f.* [a'fak] || [ə'fak]

ACEM *f.* [a'sem | a'θem (!)] || [ə'sem | a'θem (!)]

AFI *m. & f.* ['a.fi] (cf. *ang. IPA* ['aɪp^hi:'eɪ])

3.5. Onomatopeies i mots expressius

Les reflexions amb què encetàvem l'apartat anterior es podrien, de fet, fer extensives a les que dediquem a les *onomatopeies* i als anomenats *mots de creació expressiva*. En el primer cas es tracta de mots que intenten imitar sons o sorolls emesos per animals, objectes, fenòmens naturals etc., mentre que, en el segon, ens referim a mots que suggereixen alguna idea o que evoquen alguna qualitat física o moral a través del valor que s'atorga als sons vocàlics i/o consonàntics que els representen (Riera-Eures & Sanjaume (2002: 12-13)). En els dos casos, en principi, corresponen a unitats lèxiques d'un ús restringit —si més no si apel·lem a un ús conscient de la llengua— però és evident que omplen molts racons de la nostra vida diària (*ai*, quan expressem dolor; *bff*, quan bufa el vent; *pst*, quan cridem algú, etc.).

La representació graficofònica d'aquestes unitats crea sovint problemes. Quant a la representació ortogràfica, fins no fa gaire —i durant molts anys— la principal font ha estat sens dubte els còmics, les revistes infantils i els textos literaris; ara, però, ja disposem d'algun diccionari de tipus monogràfic (vg. Riera-Eures & Sanjaume (2002), el principal referent

de què ens hem servit). En relació amb el component oral, les pel·lícules infantils, les sèries de televisió o els dibuixos animats han omplert els seus guions d'onomatopeies i mots expressius; tanmateix, no disposem encara —si més no fins on arriba la nostra informació— de materials que ens puguin ajudar a ‘estandarditzar’ una sèrie de ‘pronúncies’ o de produccions fòniques, encara que no és pas fàcil, com tampoc ho és intentar sistematitzar la grafia de certes onomatopeies o mots d’expressió creativa sense deixar un marge obert a la variació.

Els criteris que hem aplicat al DPC en relació amb aquest subcorpus són els següents. Per a la selecció lèxica hem optat per incorporar les unitats d’ús més corrent, alhora que hem evitat al màxim la multiplicitat formal. Per a la informació fonètica en aquest cas hem primat el que podríem anomenar la variació expressiva en detriment de la variació dialectal, per exemple, que, d’altra banda, és més aviat escadussera. Així, no és estrany trobar per a una única entrada diferents transcripcions del tipus:

(6) *ai!* *interj.*, *m.* ['aj | 'a:ɟ | 'ʔaj | 'ʔajʔ]

Amb independència que es pugui considerar un criteri de transcripció inadequat —potser per a alguns fins i tot un mostrar de transcripcions arbitrari—, el cert és que el DPC obeeix a un objectiu molt específic: aportar propostes per a la representació fònica dels diferents valors psicològics, físics, morals, etc. que contenen aquestes unitats.

Imaginem-nos una situació real que, discursivament, pogués encaixar en la transcripció següent:

l'a(l)tre dia

vai(g) (a)nar

'tatre'ðie

βaj'na

al	aɫ
dentista	den'tiste
i encara	jaŋ'kare
no (ha)vien passat	no'βiempa'sat
tres segons	treseyonz
i dic	i'ðik
ai	'aj
...	...

Segurament tots estaríem d'acord en el fet que no expressem de la mateixa manera la sensació de sorpresa, de queixa o de dolor. Per tant, no ens sembla gaire adequat representar fonèticament les tres sensacions amb la mateixa simbologia, la qual cosa justificaria transcripcions com les de (6). Complementem-ho amb el referent discursiu anterior:

(a) *expressió de sorpresa*

ai	'aj
m'he (e)quivocat	mekiβo'kad
de consulta	dekon'suɫte
perdoni	per'ðoni

(b) *expressió de queixa*

ai	'a::j
quin mal	kim'maɫ
me fa	me'fa
aquest qu(e)ixal	a'ketki'ɟaɫ
doctor	dok'to

(c) *expressió de dolor*

ai	'ʔaj 'ʔajʔ
el dentista	eɫden'tiste
ja m'(ha)via	jam'bie
obert	aw'βer
la boca	la'βoke

3.6. Neologismes, manlleus, estrangerismes i expressions foranes: inadaptats i en vies d'adaptació

La vitalitat lingüística de l'idioma pot proporcionar nous termes i expressions des de dins de la pròpia llengua (*neologismes*), tot i que sovint recorrem al lèxic procedent de les llengües amb més difusió i més influència del nostre entorn, ja sigui mitjançant l'adaptació al català (*manlleus*) o mitjançant la conservació de la forma forana (*estrangerismes*). No hi ha dubte que entre aquests idiomes forans destaquen, en primer lloc, l'anglès, seguit de l'espanyol i del francès. Ara bé, malgrat la importància de la creativitat lèxica i la capacitat recursiva dels catalanoparlants, com constata una publicació recent (Observatori de neologia, 2004), és evident que hi ha una manca de reculls lexicograficofònics —amb transcripció fonètica— relatius als neologismes, alhora que en el cas dels manlleus i estrangerismes és convenient conèixer-ne l'origen i l'adaptació fònica al català.

En el cos del DPC hem optat per incorporar les unitats d'ús més freqüent que integren els subcorpus esmentats. Pel que fa a *neologismes*, *manlleus* i *estrangerismes*, a banda de beneficiar-nos de la recerca que, en aquest camp, s'ha dut a terme en català (TERMCAT: *Neoloteca*, 2001a,b, 2003; IULA (UPF): *Observatori de neologia*, 1998, 2003, 2004; Colell, 2003), s'hi han

introduït mots que, per la seva actualitat, encara no han estat recollits en cap altra obra lexicogràfica, i que procedeixen majoritàriament dels mitjans orals i escrits. També hem recorregut a llibres d'estil, com ara el d'*El Periódico* (2002). Respecte a *expressions foranes*, des de llatinismes fins a mots de domini prou comú d'origen japonès o àrab, ens hem servit, entre altres, d'Escolà (1997 [31999]) per a expressions llatines i d'Ehrlich (1990) per a termes i expressions corresponents a llengües contemporànies del català.

En relació amb la informació fonètica, com s'observarà seguidament, s'hi inclou tant la pronúncia catalana com l'originària (vg. el § 3.9 per a la relació de llengües). A la catalana es té en compte tant l'adaptació parafònica —a partir de la pronúncia originària— com la paragràfica —d'acord amb l'ortografia de la llengua de procedència “pronunciada a la catalana”. Vegem-ne una mostra:

(7) a) *Neologismes, manlleus i estrangerismes:*

aznarisme m. [az.na'riz.me] || [əz.na'riz.mə]

background m. [bag'grawn] || [bag'grawn | bəg'] *ang.* ['bækgʌʊnd]

bacon m. ['bej.kon | 'bej.] *ang.* ['beɪkən] (vg. *béicon*)

culebrot m. [ku.ʔe'βrət] || [ku.ʔə'βrət]

handbol m. [am'bɔt] || [ə'm'bɔt] *ang.* ['hændbɔ:t | 'hæmbɔ:t]

lobby m. ['tɔ.βi] *ang.* ['lɒbi || 'lɑ:bi]

màster m. ['mas.ter] || ['mastər] *ang. master* ['ma:stə || 'mæstə]

mobbing m. ['mɔ.βiŋ] *ang.* ['mɒbiŋ || 'ma:biŋ]

(8) b) *Expressions llatines:*

a priori loc. ll. [a.pri'ɔ.ri]

de facto loc. ll. [de'fak.to]

modus vivendi loc. ll. ['mɔ.ðuz.βi'βen.di]

(9) c) *Expressions diverses, pròpies de llengües específiques:*

Achtung! ['ax.tuŋ] al. ['ʔaxtuŋ]

au revoir! [or'βwar] fr. [oʁvwaʁ]

auf Wiedersehen! [ˌawv.βi.ðər'ze.en | 'zen] || [ˌawvβiðər'zeən | 'zen]
al. ['ʔawfvidəʁze:ən]

money m. ['mɔ.ni | 'mɔ.nej] ang. ['mʌni]

please [plis] ang. [p^hli:z]

3.7. Marques registrades/comercials (®, TDMK)

El camp de les marques comercials és un terreny completament erm pel que fa a treballs lexicogràfics i, consegüentment, tampoc no disposem de recursos lexicograficofònics. Malgrat això, considerem que és un subcorpus important ja que pot contribuir a una tasca normalitzadora rellevant, tant en relació amb els productors (publicistes, dobladors, etc.) com en relació amb els consumidors. Així, doncs, hem seleccionat els termes d'ús més corrent entre els productes de consum, empreses, maquinària (vehicles), etc., que en bona part són d'origen forà. Hem partit bàsicament de treball de camp propi amb la recollida de terminologia *in situ* (llar, centres comercials, etc.) així com de la publicitat dels mitjans orals i escrits.

Com en l'apartat anterior, el DPC proporciona, a més de la pronúncia catalana pròpia o adaptada, la pronúncia originària del referent forà, si escau, com podem observar a la mostra següent:

- (10) **Alitalia**® [a.ʎ̥ta.ʎ̥ja |.ʎ̥jə] || [aʎi'taʎja | əʎi'taʎjə] *it.* [ali'ta:ʎja]
Bayer® ['ba.jer] *al.* ['bajə] *ang.* ['beɪə]
Bailey's® ['bej.tis |'bej. |'baj.] *ang.* ['beɪlɪz]
BMW® [be.e.me'βe] || [beemə'βe] *al.* [be:em've]
Citroen® [si.tro'ɛn |'ɛn |'ɛŋ |'tro. |θi. (!)] || [sitru'ɛn |'ɛŋ |'tro |θi (!)]
fr. Citroën [sitʁo'ɛn]
Mercedes (Benz)® [mer,se.ðez'βɛns |'βɛns |,θe. (!)] ||
 [mer,seðez'βɛns || mər,seðəz' |mer,θeðez' (!)] *al.* [mɛɾ,tse:ðes'bɛnts]
Sony® ['so.ni |'sɔ.] *jap.* ['sopii]

3.8. Onomàstica auctòctona

L'onomàstica pròpia constitueix un subcorpus al nostre parer imprescindible per a tot recull lexicofònic d'una llengua. En el DPC aquest subcorpus comprèn l'onomàstica relativa als països de parla catalana que es troba recollida en aplecs lexicogràfics, com ara enciclopèdies, atles i nomenclàtors: *antroponímia* (noms, cognoms, etc.) i *toponímia* major i menor (des de veïnats i pobles corresponents a municipis petits fins a les ciutats més grans, passant per les comarques i les seves capitals; rius, llacs, serralades, costes, etc.). Les dades que aporten les fonts especialitzades (Moll (1959 [21982]); GEC (1969 [21986]); Albaigès (1980 [12000]), Moran, Batlle & Rabella (2002); *Nomenclàtor* (2003); *La Gran Enciclopèdia en català* (2004)) s'han combinat amb treballs de camp puntuals, de manera que es poden oferir les diferents possibilitats de pronúncia. Com es pot veure als exemples següents, en els topònims prioritzem la pronúncia local:

- (11) **Alcoletge** [aɫ.ko'ɫe.ðʒe |.tʃe (!)] || [əɫku'ɫedʒə |'ɫe]
Andorra, _ la Vella [an'dɔra \ ɫa'βeɫa | an'do.rɐ \ ɫa'βe.ɫɐ] ||
 [ən'dɔrə |'do \ ɫə'βeɫə]
Anna f. ['an.nɐ | 'a.nɐ] || ['annə] *ang.* ['ænə] *al.* ['ʔana] (cf.
esp., port. Ana)
Baix Empordà, el [eɫ βaj.ʒem.por'da |.ʒam. |.pur'] || [əɫβazəmpur'da]
Barberà del Vallès [bar.βe'ra.ðet.βa'ɫes] || [bərβə'raðətβə'ɫes]

3.9. Onomàstica forana: relació de llengües

Si bé no és tan imprescindible com l'onomàstica autòctona, és indiscutible que en l'entorn comunicatiu 'global' en què ens movem, els noms propis pertanyents a idiomes forans adquireixen una rellevància extraordinària, com podem comprovar, posem per cas, en un simple butlletí de notícies. Pel que fa a aquest subcorpus, el DPC incorpora, com en el cas de l'onomàstica autòctona, antropònims i topònims. S'han seleccionat antropònims coneguts i ben arrelats (de polítics; actors, actrius, cineastes; artistes, esportistes, científics; escriptors, pensadors, filòlegs...), topònims majors i menors: propers (Jaca, Nimes...) i de rellevància pel seu estatus o població (Vigo, Dover, Waterloo, Nova York...), capitals d'estat, de província, de departament, de regió, de nació, etc. Aquests de vegades apareixen en la forma tradicional establerta segons la versió 'traduïda' (p. e., Aquisgrà, Londres...) i, quan això no és possible, en la forma originària.

Amb relació a la informació fonètica, tant en els antropònims com en els topònims se sol presentar l'adaptació al català en doble versió: la pronúncia *parafònica*, a partir de la pronúncia en la llengua d'origen, i la

grafista o *paragràfica*, basada en l'ortografia originària, com es pot observar a la taula que clou aquest apartat.

A part de recórrer a la informació que ens forneixen els mèdia orals i escrits, ens hem servit dels reculls lexicogràfics següents: Alberich & Ros (1993); Duden *Das Aussprachewörterbuch* (2000); GEC (1969 [²1986]); *La Gran Enciclopèdia en català* (2004); *La Enciclopedia* (2004); Jones (1917 [¹⁶2003]) (ed. per Roach, Hartman & Setter); Krech & Stötzer (ed., 1982); Warnant (1962 [³1968]) i Wells (1990 [²2000]). També hem aprofitat les dades de treballs previs dels membres de l'equip, com ara de Julià-Muné (2000, 2004) i dels diccionaris i/o manuals de pronúncia de les llengües involucrades, sempre que ha estat possible. D'altra banda, per precisar la transcripció ens resulta de molta utilitat la darrera versió dels principis de l'AFI i el mostrari de pronúncies que conté (The International Phonetic Association, 1999). No cal dir que, sempre que hem pogut, hem enquestat parlants nadius cultes de la llengua en qüestió i, si convenia, hem consultat especialistes en fonètica de la llengua implicada. Com a últim recurs solem contactar per via telefònica amb parlants, de qui ens consta la genuïnitat de pronúncia, com poden ser els responsables culturals dels consulats i les ambaixades. Vegem tot seguit quins són els idiomes implicats des dels entorns més immediats, o que s'estudien entre nosaltres com a segones llengües o llengües foranes, fins als més remots.

Relació de llengües:

Entorns més immediats / estudi com a L2 o LF:

espanyol (generalment castellà) i francès

Domini mundial i estudi com a LF:

anglès (britànic || nord-americà)

[indicat de manera sistemàtica]

Entorns menys immediats:

basc

gallec/portuguès (europeu || brasiler)

[indicat de manera no sistemàtica]

italià, occità

Altres entorns europeus rellevants (cf. membres UE):

alemany, neerlandès; islandès/danès/suec/noruec;

(gaèlic) irlandès, gal·lès, gaèlic (escocès), bretó;

rus, polonès, txec, eslovac, eslovè, serbocroat, macedoni,

búlgar, bielorús, ucraïnès; letó, lituà;

estonià, finlandès, hongarès;

albanès; grec; maltès

Entorns no europeus força immediats, especialment pel cabal immigratori:

àrab (marroquí, algerià...), berber (o amazig)

mandinga, hausa, amhàric (altres nord-africanes)

Entorns remots però rellevants pel pes demogràfic i/o econòmic:

japonès

(xinès) mandarí / taiwanès, (xinès) cantonès

hindi/urdu, bengalí, marati

coreà

Entorns remots però rellevants pel pes històric:

turc

hebreu, àrab (saadí, egipci, sirí, iraquí...)

persa, afganès (o paixto), kurd

Entorns remots menys rellevants:

swahili

zulu/xosa

(altres sud-africanes)

indonesi
vietnamita
(altres extremerorientals)

A continuació incloem un seguit d'exemples que serveixen per il·lustrar alguns dels problemes que afecten l'onomàstica forana.

(12) a) *Supeditació a l'espanyol*

Algunes adaptacions al català es realitzen mediatitzades pel sistema fònic espanyol i, consegüentment, són pronúncies inadequades (s'indiquen amb el símbol (!)). De fet, la pronúncia 'catalanitzada' resulta molt més propera a l'originària. Vegem-ne alguns exemples:

Anderlecht ['an.der.ɫɛxt | .ɫɛkt | .ɫɛtʃ (!)] || ['andərɫɛxt | ɫəxt | ɫɛkt | ɫɛtʃ (!) | ɫətʃ (!)] *neerl.* ['andərɫɛxt]
Cluj [kɫuʃ | kɫuɣ (!)] *rom.* [kluʒ]
Coch [koɣ | kok | kotʃ (!)] *gal·lès* [ko:ɣ] (cf. *ang.* [k^həʊx | k^həʊk || k^houk])

(13) b) *Termes homògrafs però no homòfons, si més no en la versió parafrònica* (p. e., en català i hongarès):

Brassó [bra'so] || [brə'so]
Brassó ['brɔʃo | bra'so] || ['brɔʃo | brə'so] *hong.* ['brɔʃʃɔ]

(14) *c) Adequació de la pronúncia a la multiplicitat ortogràfica:*

Andròpov, Andropov [an'drɔ.paf |.pɔf \ an.dro'pɔf] || [an'drɔpəf
|pɔf \ əndru'pɔf | andro'pɔf] *rus* [an'drɔpɐf]

Adaptació de l'onomàstica i ® foranes

(pronúncies —en N-O, V || C, B— proposades per als mèdia)

IDIOMA	PRONÚNCIES		
	<i>Tradicional</i> (assimilable a la forma 'traduïda')	<i>Catalanitzada</i> 1. <i>Adaptació a partir de la pronúncia originària:</i> pronúncia parafònica. 2. <i>Adaptació segons l'ortografia:</i> pronúncia grafista o paragràfica.	<i>Originària</i> (de llengües contemporànies del català)
Espanyol (castellà)	Osca ['ɔs.ke\ə 'ɔskə]		Huesca ['weska]
		Badajoz (1. = <i>orig.</i>); 2. [ba.ða'xos] [bæðə'xos] Cáceres (1. = <i>orig.</i>); 2. ['ka.se.res 'kasərəs]	Badajoz [baða'xoθ] Cáceres ['kaθeres]
Francès	París [pa'ris pə'ris]	Bacqueville 1. [bag'viʔ] 2. [bag'biʔ] V, B 1./2. [bag'viʔ] Bailleul 1. [ba'jɔʔ 'jeʔ] Chartres 1./2. [ʔʃar.tre\s 'ʃartre\s]	Paris [pa'ʁi] Bacqueville [bak'vil] Bailleul [ba'jœl] Chartres ['ʃartʃ]
Portuguès (europeu) (brasiler)	Lisboa [ʔiz'βo.ɛ\ə ʔiz'βoə]	São Paulo 1. [saw'paw.ʔo 'saw'pawʔu] 2. [sao'paw.ʔo]	Lisboa [ʔiz'βoɐ] São Paulo [səwm'pawʔu]

Anglès (britànic)	Londres [ˈtɒn.dres ˈɒndrəs]	Bacon 1. [ˈbeɪ.kən] ˈbeɪ. ˈbeɪkən]; 2. [ˈbeɪ.kən ˈbeɪ. ˈbeɪkən] Hitchcock 1./2. [ˈhɪtʃ.kɒk] Laughton 1. [ˈtɔ.tən ˈtɒtən]; 2. [ˈtaw.tən]	London [ˈlʌndən] Bacon [ˈbeɪkən] Hitchcock [ˈhɪtʃkɒk] Laughton [ˈlɔːtən]
Alemanys	Colònia [koˈtɔ.ni.ɛ\ə ˌnɛ\ə] C [kuˈtɔniə ˌnɛə] B [kuˈtɔniə ˌkoˈ ˌnɛə]	Bosch® 1. (2). [ˈbɔʃ] Hitler 1. [ˈhɪd.ʔa ˈhɪd.ʔə]; 2. [ˈhɪd.ʔer ˈhɪd.lər]	Köln [kʰœln] Bosch® [ˈbɔʃ] Hitler [ˈhɪtlər]
Neerlandès	L'Haia [ˈʔa.je\ə ˈʔajə]	Cruyff 1. [ˈkrɔɪf]; 2. [ˈkrwɪf] Van Gaal 1. [fɑŋˈxɑt]; 2. [bɑŋˈgɑt]; V, B 2. [vɑŋˈgɑt]	Den Haag [denˈhɑːx] Cruyff [ˈkrœɪf\ˈkœœɪf] Van Gaal [fɑŋˈxal]
Hongarès	Budapest [bu.ðəˈpest buðəˈpest]	Bács 1. [bɑʃ]; 2. [baks]	Budapest [ˈbudɒpeʃt] Bács [bɑʃ]
Rus	Moscou [mosˈkɔw C [musˈkɔw] B [mosˈkɔw ˌmusˈ]	Boris 1. [baˈrɪs]; 2. [ˈbɔ.rɪs] Khruixtov 1./2. [kruˈʔʃɔf kruˈʃɔf]	Moskva [mɐskˈva] Boris [bɐˈrɪs] Khruixtov [xruɐˈʔɔf]
Mandarí	Pequín [peˈkin pəˈkin]	Beijing 1. [peɪˈʔɪŋ pəɪˈʔɪŋ]; 2. [beɪˈʔɪŋ bəɪˈʔɪŋ] Báihuà 1. [pajˈχwa]; 2. [bajˈχwa]	Běijīng [ˈpeɪ ˈʔeɪŋ] Báihuà [ˈpaj ˈχwa]

4. El corpus complementari biobibliogràfic i terminològic: relació de quadres

Quadres biobibliogràfics

ALCOVER, Antoni M.	JONES, Daniel
ARTEAGA, Josep M.	MARTINET, André
ASSOCIACIÓ FONÈTICA INTERNACIONAL (AFI/IPA)	MOLL, Francesc de B.
BARNILS, Pere	NAVARRO TOMÁS, Tomás
BELL, Alexander Graham	PASSY, Paul
[CHOMSKY, N. & HALLE, M.]	PIGMALIÓ
COROMINES, Joan	PIKE, Kenneth L.
FABRA, Pompeu	PRIMER FONÒLEG (Anònim islandès)
JAKOBSON, Roman	SANCHIS GUARNER, Manuel
JESPERSEN, Otto	SCHÄDEL, Bernhard
	SWEET, Henry

Quadres terminològics

Abreviacions, abreviatures, sigles i acrònims	Arrodoniment (arrodonit >< tibant)
Accent, accentuació	Articulació
Afèresi	Aspiració (moment d'engegada laríngia, MEL/VOT)
Africat	Assimilació
Alfabet Fonètic Internacional (AFI/IPA)	Pronunciació de 'B'
Allargament (durada; llarg >< breu)	Barbarisme fònic
Al·literació	Batec, bategant
Al·lòfon	Bilabial
Alveolar, alveolopalatal, postalveolar	Bucal
Alternança vocàlica	Pronunciació de 'C'
Anglès, anglicismes	Cardinal (vocal _)
Apòcope, forma apocopada	Clic
Aproximant	Coalescència
	Coarticulació

Coda	<i>Pronunciació de 'L, LL, L·L'</i>
Consonant, contoid	Labial, labialització
Constrictiu	Labiodental
	Laringe, laringi
<i>Pronunciació de 'D'</i>	Lateral
Dental, dentalització	Líquids
Diftong	Llatí, llatinismes
Distribució complementària	Llengua, lingual; apical, dorsal
Dissimilació	Lletres 'mudes'
Elisió (apostrofació)	Nasal, nasalització
Enllaç (<i>liaison, linking</i>)	Neologismes
Entonació	Neutra (vocal _), <i>schwa</i>
Epèntesi	Neutralització (arxifonema)
Epítesi	
	Obertura
Faringi	Oclusiu
Fonema	Onomàstica forana
Fonètica	Onomatopeia (simbolisme fònic)
Fonologia	Ortoèpia (ortologia)
Fonotaxi	
Francès (gal·licismes)	Palatal, palatoalveolar
Fricatiu	Pausa
	<i>Pinyin</i>
<i>Pronunciació de 'G'</i>	Plecs glotals
Glotal, glotis, glotalització	Prefix
	Pronúncia canònica, connectada (parla contínua)
<i>Pronunciació de 'H'</i>	Pronúncia originària / grafista o paragràfica / parafònica
H aspirada/muda	Pròtesi
Hiat	
Homòfons	
	<i>Pronunciació de 'R, RR'</i>
Ieisme, ieista	Reducció vocàlica
Iod, iodització	Retroflex (apicoalveolar)
<i>Pronunciació de 'J'</i>	Rima

Ritme (cadència, tempo)

To, tonal (tònic >> àton)

Ròtic

Triftong

*Pronunciació de 'S, SS, Ç'**Pronunciació de 'V'*

Semivocàlic

Vel del paladar (paladar tou)

Sensibilització

Velar, velarització

Síl·laba, sil·labificació, estructura

Veu

sil·làbica

Vibrant

Síncope, forma sincopada

Vocal, vocoid

Sonor, sonorització

Uvular

Sons opcionals

Sord (ensordiment)

Sufix

Pronunciació de 'X'

Tancament

Pronunciació de 'Z'

Tensió (lax >> tens)

ANNEX I: MOSTRA PROVISIONAL EL DPC SEGONS LES VARIETATS DEL PRINCIPAT [Nord-occidental || central]

[...]

androide *adj.*, *m.* [an'droj.ðe] ||

[ən'drojðə]

andròlatra *m.* & *f.* [an'drɔ.ʎa.trɛ] ||

ən'drɔ ʎətrə]

androlatria *f.* [an.dro.ʎa'tri.ɛ] ||

[andruʎə'triə]

andròleg, **-òloga** *m.* & *f.* [an'drɔ.ʎək\.

ʎɔ.ɣɛ] || [ən'drɔʎək\ ʎɔɣə]

andrologia *f.* [an.dro.ʎɔ'zi.ɛ |. ʎɔʎ] ||

[andruʎɔ'ziə]

Andròmac, **-a** *m.* & *f.* [an'drɔ.mak\ɛ] ||

[ən'drəmək\ə]

andromania *f.* [an.dro.ma'ni.ɛ] ||

[andrumə'niə]

baconada *f.* [ba.ko'na.ðe] || [bəku'naðə]**baconera** *f.* [ba.ko'ne.ɾɛ] || [bəku'neɾə]**baconià**, **-ana** *adj.* [ba.ko.ni'a\ .nɛ |'nja]

|| [bəkuni'a\nə |'nja]

baconisme *m.* [ba.ko'niz.me] ||

[bəku'nizmə]

bacora *f.* [ba'ko.ɾɛ] || [bə'korə]**bacoreta** *f.* [ba.ko're.ɾɛ] || [bəku'retə]**Bacqueville** [bag'viʎ |'biʎ] *fr.* [bak'vil]**Bács** [batʃ | baks] *hong.* [bartʃ]**Bácska** ['batʃ.kɔ | 'baks.ka] *hong.*

['bartʃko]

bacsonià, **-ana** *adj.*, *m.*; **Bacsonià**

[bak.so.ni'a\ .nɛ |'nja] || [bəksuni'a\nə

|'nja]

Andròmeda *f.* [an'drə.me.ðe] ||
[ən'drəmæðə]

Andromèdids, els *m. pl.*

[eɫ,zan.dro'me.ðĩs] ||

[əɫ,zandru'meðĩs]

andròmina *f.* [an'drə.mi.ne] ||

[ən'drəminə]

andromonoècia *f.*

[,an.dro.mo.no'ɛ.si.ɛ |.sje] ||

[,andrumunu'esiə |sjə]

andromonoic, -a *adj.*

[,an.dro.mo'nɔjk\ɛ] ||

[,andrumu'nɔjk\ə]

andromorf, -a *adj.* [,an.dro'morf\ɛ

'morf] || [,andru'morf\ə]

andróna *f.* [an'dro.ne] || [ən'dronə]

Andrònic *m.* [an'drə.nik] ||

[ən'drɒnik]

Andronovo [an'drə.na.βa |

an.dro'nə.βo] || [ən'drɒnəβə |

andro'nəβo] *rus* [ən'drɒneve]

andronvià, -ana *adj., m.*

[an.dro.no.βi'a\ .ne 'βja] ||

[əndrunuβi'a\ nə 'βja]

andropausa *f.* [,an.dro'paw.zɛ] ||

[,andru'pawzə]

andropètal *m.* [,an.dro'pe.taɫ |.tɛɫ] ||

[,andru'petəɫ]

Andròpov, Andropov [an'drə.paf

||.pof \ an.dro'pɒf] || [an'drɒpɒf

|pɒf \ əndru'pɒf] | andro'pɒf] *rus*

[an'drɒpɒf]

Andros *m.* ['an.dros] *grec* ['andrɒs]

[...]

back-up *m.* [bək'ap] *ang.* ['bækʌp]

backyard *m.* [bæg'jart 'jar] *ang.*

[,bæk'jɔ:ɔ || 'jɔ:ɔɔ]

bacó, -ona *m. & f.* **Bacó** [ba'ko\ .ne]

|| [bə'ko\ nə]

Baço [ba'so]

[...]

baihuà *m. & f.* [paj'χwa | baj'χwa]

mand. báihuà [²paj⁴χwa]

Baikal [baj'kaɫ] *rus* [bɛj'kaɫ]

Baikonur [baj.ka'nur |.ko'] || [bajkə'nur

|ko'] *rus* [bajkə'nur]

baila f., Baila ['baj.tɛ] || ['bajtə]

Bailach [baj'tak] || [baj'tak]

Bailaguet [baj.ta'ɣet] || [bajtə'ɣet]

bailaor, -a *m. & f., -ores, -oras* *pl.*

[baj.ta'or\o.ra \ .res\ .ras] *esp.*

[bajla'or\ora res\ras]

Bailet [baj'tet] || [baj'tet]

Bailey ['bej.ti 'bej. 'baj.] *ang.* ['beili]

Bailey's® ['bej.tis 'bej. 'baj.] *ang.*

['beiliz]

Bailina [baj'ti.ne] || [baj'tinə]

Baillaud [ba'jo] *fr.* [ba'jo]

Bailleul [ba'jɔɫ 'jɛɫ] *fr.* [ba'jœɫ]

Baillon [ba'jo 'joŋ] *fr.* [ba'jɔ]

[...]

Bardina [bar'di.ne] || [bər'dinə]

Bardinet [bar.ði'net] || [bərði'net]

bardiota *f.* [bar.ði'ɔ.te] || [bərði'ɔtə]

bardissa f., Bardissa [bar'di.se] ||

[bər'disə]

bardissaire *m. & f.* [bar.ði'saj.re\ɛ] ||

[bərði'sajrə]

bardissam *m.* [bar.ði'sam] ||

[bərði'sam]

bardissar *m.* [bar.ði'sa] || [bərði'sa]

bardissenc, -a *adj.* [bar.ði'sɛŋ\ .kɛ] ||

[bərði'sɛŋ\ kə]

bardisser, -a *adj., m. & f.* [bar.ði'se\ .rɛ]

|| [bərði'se\ rə]

bardo *m.* ['bar.ðo] || ['barðu]

bardó m., Bardó [bar'ðo] || [bər'ðo]

Bardolet [bar.ðo'tet] || [bərðu'tet]

bardoll *m.* [bar'ðoɫ] || [bər'ðoɫ]

bardollar *v.* [bar.ðo'la] || [bərðu'la]

Bacon ['bej.kon |'bej.] || ['bejkən
|kon] *ang.* ['beikʰn]
bacon *m.* ['bej.kon |'bej.] *ang.*
['beikʰn] (vg. *béicon*)

Bardot, Brigitte [bri'ʒid.bar'ðo |'ðot] ||
[bri'ʒidbar'ðo |bər'ðot] *fr.* [bʁiʒitbaʁ'do]
[...]

**ANNEX II: MOSTRA PROVISIONAL DEL DPC ELABORAT
SEGONS LES QUATRE VARIETATS
[Nord-occidental, valencià || central, balear]**

abarrotar *v.* N-O [a.βa.ro'ta] V
[aβaro'tar] || C [əβəru'ta] B
[əβəro'ta |ru']
abarrotat, -ada *adj.* N-O
[a.βa.ro'tat\ə.ðe]
V [aβaro'tat\aða] ||
C [əβəru'tat\aðə] B [əβəro'tat\aðə
|ru']
abarset *m.* N-O, V [a.βar'set] || C
[əβər'set] B [əβər'set |'set]
abarta *f.* N-O [a'βar.te] V [a'βarta]
|| C, B [ə'βartə]
abartra *f.* N-O [a'βar.tre] V
[a'βartra] || C, B [ə'βartre]
abartrell *m.* N-O, V [a.βar'treɫ] || C,
B [əβər'treɫ]
abartrosi *f.* N-O, V [a.par'tro.ʒi
|a.βar' |.ʒis (!) |.sis (!)] || C, B
[apər'trozi |əβər' |ʒis (!) |sis (!)]
Abascal N-O, V [a.βas'kaɫ] || C
[əβəs'kaɫ], B [əβəs'kaɫ |'kaɫ] *esp.*
[aβas'kal]
abasardar *v.* N-O [a.βa.zar'ða] V
[aβazar'ðar] || C, B [əβəzər'ða]
abàsia *f.* N-O [a'βa.ʒi.ɛ |.ʒjɛ] V
[a'βazia |ʒja] || C, B [ə'βaziə
|ʒjə]

abastable *adj.* N-O [a.βas.ta.βte |.bte
|.pte (!)] V [aβas'taβte] || C
[əβəs'tabte |pte (!)] B [əβəs'tabte]
abastador, -a *adj., m. & f.* N-O
[a.βas.ta'ðo\ .ɾe] V [aβasta'ðor\ə] || C,
B [əβəstə'ðo\ɾə]
abastament *m.* N-O [a.βas.ta'men] V
[aβasta'ment |'men] || C [əβəstə'men]
B [əβəstə'ment]; **en vol** *loc.* N-O
[a.βas.ta.me.nem'boɫ] V
[aβasta,mentem'vɔɫ |tem'boɫ] || C
[əβəstə,menəm'boɫ] B
[əβəstə,mentəm'vɔɫ]
abastar *v.* N-O [a.βas'ta] V [aβas'tar] ||
C, B [əβəs'ta]
abastat, ada *adj.* N-O [a.βas'tat\ə.ðe] V
[aβas'tat\aða] || C, B [əβəs'tat\aðə]
abastidor, -a *m. & f.* N-O [a.βas.ti'ðo\ .ɾe]
V [aβasti'ðor\ə] || C, B [əβəsti'ðo\ɾə]
abastiment *m.* N-O [a.βas.ti'men] V
[aβasti'ment |'men] || C [əβəsti'men] B
[əβəsti'ment]
abastir *v.* N-O [a.βas'ti] V [aβas'tir] ||
C, B [əβəs'ti]
abat *m., Abat* N-O, V [a'βat] || C, B
[ə'βat]
abatallar *v.* N-O [a.βa.ta'ɫa] V

abassegador, -a *adj.*, *m.* & *f.* N-O

[a.βa.se.ɣa'ðo\ .rɛ] V

[aβaseɣa'ðor\ a] || C

[əβəsəɣə'ðo\rə] B [əβəsəɣə'ðo\rə
|jə']

abassegement *m.* N-O

[a.βa.se.ɣa'men] V [aβaseɣa'ment

'men] || C [əβəsəɣə'men] B

[əβəsəɣə'ment |jə']

abassegar *v.* N-O [a.βa.se'ɣa] V

[aβase'ɣar] || C [əβəsə'ɣa] B

[əβəsə'ɣa |'ja]

abast *m.* N-O, V [a'βast] || C, B

[ə'βast]

[aβata'kar] || C, B [əβətə'ka]

abatedor, -a *adj.*, *m.* & *f.* N-O

[a.βa.te'ðo\ .rɛ] V [aβate'ðor\ a] || C, B

[əβətə'ðo\rə]

abatible *adj.* N-O [a.βa'ti.βte |.bte |.pte

(!)] V [aβa'tiβke] || C [əβə'tib:tə |ptə

(!)] B [əβə'tib:lə]

abatiment *m.* N-O [a.βa.ti'men] V

[aβati'ment 'men] || C [əβəti'men] B

[əβəti'ment]

abató *m.* N-O, V [a.βa'to] || C, B

[əβə'to]

Referències bibliogràfiques

- ALBAIGÈS, Josep M. (1980 [¹2000]): *Diccionari de noms de persona*, Barcelona: Edicions 62.
- ALBERICH, Joan i ROS, Montserrat (1993): *La transcripció dels noms propis grecs i llatins*, Barcelona: Enciclopèdia Catalana.
- ALCOVER, Antoni M. i MOLL, Francesc de B. (1926/1930-1962/1968 [²1978]): *Diccionari català-valencià-balear* (10 vol.), Palma de Mallorca: Moll (DCVB)
- BLECUA, José Manuel (1996): «Reflexiones al margen de los corpus escritos», dins Payrató, L.; Boix, E.; Lloret, M. R.; Lorente, M. (ed.), *Corpus, corpora*, Barcelona: PPU, 15-26.
- BOVER, August (1996 [²1997]): *Diccionari dels gentilicis catalans*, Barcelona: Edicions 62.
- BRUGUERA, Jordi (1990 [²2004]): *Diccionari ortogràfic i de pronúncia*, Barcelona: Enciclopèdia Catalana.
- CAMPS, Cristià & BOTET, Renat (2001): *Diccionari català-francès*, Barcelona: Enciclopèdia Catalana.

- CARRÉ, R. (1991): «Los bancos de sonidos», dins Vidal Beneyto, J. (dir.), *Las industrias de la lengua*, Salamanca-Madrid: Fundación Sánchez Ruipérez-Pirámide, 108-118.
- _____ (1992): «Speech Databases», dins Ainsworth, A. (ed.), *Advances in Speech Hearing and Language Processing. A Research Annual*, 2, Londres: Jai Press, 199-216.
- CASELLAS, Fèlix & CASELLAS, David (2003): *Vocabulari de les comarques gironines*, Barcelona: La Busca.
- CASTELLANOS, Josep Anton (2004): *Manual de pronunciació. Criteris i exercicis d'elocució*, Vic: Eumo.
- COLL, Pep (1991): *El parlar del Pallars*, Barcelona: Empúries.
- COLELL, David (2003): «La pronúncia dels neologismes en català», *Sintagma*, 15, 5-21.
- CREUS, Imma; GONZÁLEZ, Rodolf; JULIÀ I MUNÉ, Joan (1998): «Dialectalismes i diccionaris: entre la tradició i la variació», *Sintagma*, 10, 61-80.
- CRYSTAL, David (2000): *Diccionario de lingüística y fonética*, Barcelona: Octaedro. Versió espanyola de *A Dictionary of Linguistics and Phonetics* (1980 [⁴1997]), Oxford: Blackwell.
- DOMÉNECH, Josep Lluís (coord.) (2002): *Diccionari bàsic de la Comunicació*, València: Nau Llibres.
- DUDEN (2000): *Das Aussprachewörterbuch*, Mannheim: Dudenverlag.
- EHRLICH, Eugene (ed.) (1990): *Dictionary of foreign terms and phrases*, London: Penguin.
- ESCOLÀ, Josep M. (1997 [³1999]): *Diccionari de llatinismes i expressions clàssiques*, Barcelona: Edicions 62.
- GINEBRA, Jordi i MONTSERRAT, Anna (1999): *Diccionari d'ús dels verbs catalans*, Barcelona: Edicions 62.
- Gran Diccionari de la llengua catalana* (1998): Barcelona: Enciclopèdia Catalana (GDLC)
- Gran Enciclopèdia Catalana* (1970 [²1986]): 24 vol. i 3 supl., Barcelona: Enciclopèdia Catalana (GEC)
- INSTITUT D'ESTUDIS CATALANS (1990 [³1998]): *Proposta per a un estàndard oral, I: Fonètica*, Barcelona: IEC.

-
- _____ (1992 [⁴1998]): *Proposta per a un estàndard oral, II: Morfologia*, Barcelona: IEC.
- _____ (1995): *Diccionari de la llengua catalana*, Barcelona / Palma de Mallorca / València: Edicions 3 i 4, Ed. 62, Editorial Moll, Enciclopèdia Catalana, Publicacions de l'Abadia de Montserrat (DIEC).
- JONES, Daniel (1917 [¹⁶2003]): *English Pronouncing Dictionary*, Cambridge: CUP (ed. per P. Roach, J. Hartman & J. Setter; amb CD) (EPD)
- JULIÀ I MUNÉ, Joan (2000): «L'onomàstica a la ràdio: descripció i propostes ortoèpiques», dins Julià i Muné, J. (ed.) *Llengua i ràdio*, Barcelona: Publicacions de l'Abadia de Montserrat, 183-194.
- _____ (en premsa): «L'ortoèpia de l'onomàstica forana», dins *Actes del 13è Col·loqui Internacional de Llengua i Literatura Catalanes* (Girona, setembre 2003).
- _____ (en premsa): «Els diccionaris de pronúncia del català: passat, present i futur», *Revista de Filologia Romànica*.
- JULIÀ-MUNÉ, Joan & CREUS, Imma (2003): «A New Dictionary of Contemporary Catalan Pronunciation», dins Solé, M. J.; Recasens, D. & Romero, J. (ed.), *Proceedings of the 15th International Congress of Phonetic Sciences* (Barcelona 3-9 August), Australia/Barcelona: Causal Productions/UAB, vol. 1, 969-972.
- KRECH, Hans i STÖTZER, Ursula [AUTORENKOLLEKTIV] (ed.) (1982): *Grosses Wörterbuch der deutschen Aussprache*, Leipzig: VEB Bibliographisches Institut.
- La Enciclopedia* (2004): 20 vol., Madrid: Salvat.
- La Gran Enciclopèdia en català* (2004): 20 vol., Barcelona: Edicions 62.
- LACREU, Josep (ed.) (2001): *Diccionari valencià de pronunciació*, Alzira: Bromera.
- LLISTERRI, Joaquim (1996): «Els corpus lingüístics orals», dins Payrató, L.; Boix, E.; Lloret, M. R.; Lorente, M. (ed.), *Corpus, corpora*, Barcelona: PPU, 27-70.
- MASSANA, Josep M. (2004): *Diccionari de lleidatanismes*, Lleida: Pagès.
- MESTRES, Josep M. i GUILLÉN, Josefina (1992 [²2001]): *Diccionari d'abreviacions*, Barcelona: Enciclopèdia Catalana.
- MORAN, Josep; BATLLE, Mar; RABELLA, Joan Anton (2002): *Topònims catalans. Etimologia i pronúncia*, Barcelona: Publicacions de l'Abadia de Montserrat.
- Nomenclàtor Oficial de Toponímia Major de Catalunya* (2003), Barcelona: Generalitat de Catalunya/IEC.
- Nouveau Petit Larousse* (1970), Paris: Larousse.

- OBSERVATORI DE NEOLOGIA IULA/UPF (1998): *Diccionari de paraules noves. Neologismes recollits a la premsa*, Barcelona: Enciclopèdia Catalana.
- _____ (2003): *Neologismes documentats en textos escrits i orals en català (2001)*, Barcelona: UPF, IULA.
- _____ (2004): *Llengua i neologia*, Barcelona: Meteora.
- ORTEGA, Rudolf (2003): *Diccionari de futbol*, Barcelona: Ed. 62.
- ORTÍS, Pere (2002): *La parla de l'Urgell*, Bellpuig: Pere Ortís.
- PALOMA, David & RICO, Albert (2000): *Diccionari de pronunciació en català*, Barcelona: Edicions 62.
- PERIÓDICO, EL (2002): *Llibre d'Estil*, Barcelona: Primera Plana.
- POMARES, Joaquim (1997): *Diccionari del català popular i d'argot*, Barcelona: Edicions 62.
- RECASENS, Daniel (1985): *Estudi lingüístic sobre la parla del Camp de Tarragona*, Barcelona: Curial/Publicacions de l'Abadia de Montserrat.
- RIERA-EURES, Manel i SANJAUME, Margarida (2002): *Diccionari d'onomatopeies i mots de creació expressiva*, Barcelona: Edicions 62.
- SISTAC, Ramon (1993): *El ribagorçà a l'Alta Llitera. Els parlars de la vall de la Sosa de Peralta*, Barcelona: Institut d'Estudis Catalans.
- _____ (1998): *El català d'Àneu. Reflexions a l'entorn dels dialectes contemporanis*, Esterri d'Àneu: Consell Cultural de les Valls d'Àneu.
- TELEVISIÓ DE CATALUNYA (1995): *El català a TV3. Llibre d'estil*, Barcelona: Edicions 62.
- TERMCAT (2000 [²2003]): *Societat de la informació. Noves tecnologies i Internet*, Barcelona: Termcat.
- _____ (2001a): *Nou Diccionari de neologismes*, Barcelona: Edicions 62.
- _____ (2001b): *Diccionari d'Internet*, Barcelona: Enciclopèdia Catalana.
- THE INTERNATIONAL PHONETIC ASSOCIATION (1999): *Handbook of the International Phonetic Association*, Cambridge: CUP.
- TURULL, Albert (1996): «Mots d'ordre. Notes sobre el *Diccionari* de l'Institut i el lèxic nord-occidental», *Tirant al Blanc* 9, 70-74.
- VENY, Joan (1958-59, 1960): «Paralelismos léxicos en los dialectos catalanes», *Revista de Filología Española* LXII, 91-149; XLIII, 117-202.

- _____ (1979): «Dialectalismes en els poetes lleidatans», *Homenaje a Samuel Gili Gaya (in memoriam)*, Barcelona: Biblograf, 193-200.
- _____ (1984): «Lexicografia i dialectalismes. A propòsit del Diccionari català-llatí d'Antoni Font (s. XVII)», dins *Miscel·lània Sanchis Guarner I*, València: Universitat de València, 381-390.
- VERDAGUER, Pere (1999): *Diccionari de renecs i paraulotes*, Perpinyà: Trabucaire.
- WARNANT, Léon (1962 [³1968]): *Dictionnaire de la prononciation française*, Gembloux: J. Duculot.
- WELLS, John C. (1990 [²2000]): *Longman Pronunciation Dictionary*, Harlow: Longman. (LPD).



UNIVERSITAT DE BARCELONA



Secció de Lingüística Catalana
Departament de Filologia Catalana

COL·LECCIÓ
LINGÜÍSTICA
CATALANA

9