



Idiolectometría Forense e Índice de Similitud Idiolectal

Forensiclab - Unitat de Variació Lingüística
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra

Actualizado: enero 2013

1. Presentación

Esta investigación se ha llevado a cabo con la ayuda de las siguientes financiaciones:

Fase I

Proyecto: **Idiolectometría aplicada a la lingüística forense**

(EXPLORA-INGENIO HUM2007-29140-E)

PI: Dr. M. Teresa Turell

Entidad: Ministerio de Educación y Cultura

Periodo: 2007-2008

Fase II

Proyecto: **Idiolectometría forense e Índice de Similitud Idiolectal** (FII2008-03583/FILO)

PI: Dr. M. Teresa Turell

Entidad: Ministerio de Ciencia e Innovación

Periodo: 2008-2012

Fase III

Proyecto: **Hacia la consolidación de un Índice de Similitud/Distancia Idiolectal (IS/DI) en Idiolectometría Forense** (FFI2012-34601)

PI: Dr. M. Teresa Turell

Entidad: Ministerio de Economía y Competitividad

Periodo: 2012-2015

1.1 Finalidad

En este proyecto se propone estudiar el estilo idiolectal desde una perspectiva aplicada a la **lingüística forense** (en especial a la **fonética forense**, para la identificación de hablantes y el establecimiento de perfiles lingüísticos, por un lado, y la **determinación/atribución de autoría** de textos escritos, por otro). El estilo idiolectal se puede definir como la elección individual por parte de cada hablante de algunas formas lingüísticas (fonéticas, fonológicas, morfológicas, sintácticas, pragmáticas, discursivas) determinadas que le ofrece la lengua (Nolan 1994:331). Por tanto, el estilo idiolectal es individual y único.

La **idiolectometría** tiene como objeto de estudio el **estilo idiolectal**. Hasta ahora esta disciplina se ha ocupado de analizar la distancia lingüística entre hablantes y de establecer fronteras entre idiolectos diferentes, fronteras que por definición separan a cada persona del resto de hablantes. En cambio, este proyecto se propone explorar y desarrollar la posibilidad de medir la diferencia lingüística entre diversos idiolectos y la distancia idiolectal de cada individuo, de manera que al final se pueda conseguir un **Índice de Similitud Idiolectal (ISI)**, una medida numérica que compare diversas muestras lingüísticas y calcule la distancia lingüística entre

ellas. Más específicamente, se trata de poder establecer el nivel de similitud idiolectal a partir del cual se puede considerar que dos muestras lingüísticas (orales o escritas) han sido producidas por una misma persona. La aplicación del estudio del estilo idiolectal al campo de la lingüística forense (disciplina que considera la interfaz entre lengua y derecho) es fundamental, ya que dicha aplicación debería permitir identificar de manera inequívoca a los hablantes o escritores, comparando una grabación o texto escrito **dubitado** (cuya autoría presenta dudas) y un conjunto de grabaciones o textos escritos **indubitados** (cuya autoría es conocida), a partir del análisis de las formas lingüísticas elegidas por cada hablante. Así pues, se propone la creación de un protocolo que permita establecer ese **Índice de Similitud Idiolectal (ISI)** y concluir, una vez analizados diversos parámetros lingüísticos (fonológicos, morfosintácticos, discursivo-pragmáticos), si dos grabaciones o dos textos escritos han sido producidos por una misma persona o no.

En este proyecto a) se aplicará la técnica a cuatro lenguas (español, catalán, inglés y árabe), b) se evaluará el diseño global del protocolo, y c) se diseñarán las pruebas estadísticas para todas las lenguas.

1.2 Antecedentes y estado actual de los conocimientos científico-técnicos

El marco conceptual del tema que se quiere explorar en este proyecto está basado en tres disciplinas lingüísticas: la **lingüística forense** (en particular, la **fonética forense**, para la identificación de hablantes y el establecimiento de perfiles lingüísticos, por un lado, y la **determinación/atribución de autoría** de textos escritos, por otro); la **idiolectometría**, y finalmente, la **teoría de la variación y el cambio lingüístico**, y más específicamente, la **variación sociolingüística**.

Para poder situar conceptual y metodológicamente este proyecto definimos la **lingüística forense** de forma muy general como la interfaz entre lengua y derecho (para una definición más detallada, consúltese: <http://www.iafl.org>). Es preciso comentar desde el punto de vista metodológico que en la mayoría de casos la lingüística forense utiliza técnicas derivadas de las nuevas tecnologías, es decir, que el desarrollo de la investigación y de servicio público en lingüística forense se realiza habitualmente a partir de toda una serie de herramientas informáticas y técnicas estadísticas que posibilitan opiniones, actuaciones y asesoramiento científico y profesional de alta competencia y fiabilidad por parte de los lingüistas forenses, para su posterior uso desde diferentes instituciones y organismos públicos (escuela judicial, ministerio fiscal, policía) y privados (empresas), y por parte de profesionales diversos (jueces, abogados defensores, psicólogos, médicos). En cualquier caso, las pruebas periciales que se presentan no son en ningún caso pruebas definitivas ni concluyentes, sino más bien complementarias de otros tipos de evidencia o bien para introducir una duda razonable.

La **idiolectometría**, disciplina lingüística que se encuentra en fase inicial de desarrollo, tiene como objeto de estudio el **estilo idiolectal** de los hablantes y escritores de una determinada lengua. El estilo idiolectal se puede definir como la elección, entre las diferentes opciones que ofrece la lengua, que los hablantes hacen de algunas formas lingüísticas (fonéticas,

fonológicas, morfológicas, sintácticas) determinadas. Por tanto, podemos entender que una lengua está formada por la suma de dialectos y sociolectos, y estos a su vez se definen como la suma de idiolectos (o estilos idiolectales de la lengua). Así, los factores sociales (grupo socioeconómico, grupo de edad, nivel educativo, género, profesión) se manifiestan en la configuración de los idiolectos. La figura 1 ofrece una representación esquemática de lo que se acaba de comentar:

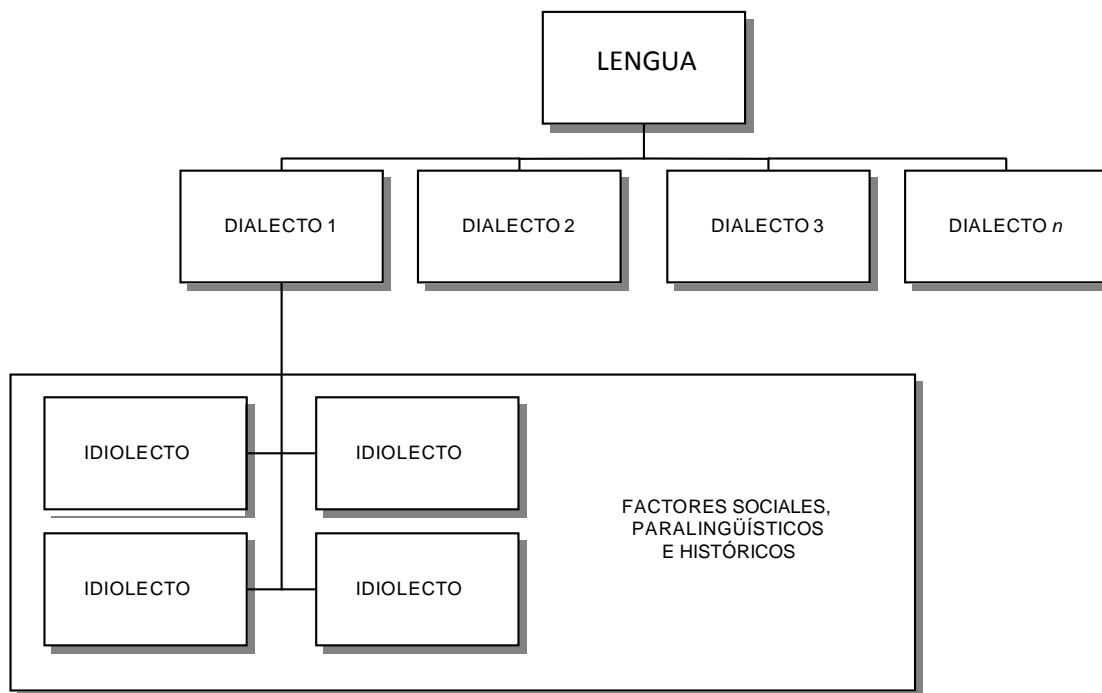


Figura 1. Esquema de la estructura de las lenguas, formadas por dialectos, y estos formados por el conjunto de idiolectos. Los idiolectos están condicionados por factores sociales, paralingüísticos e históricos.

Los estudios que analizan la influencia de los factores sociales en las realizaciones lingüísticas de los hablantes se centran en la **teoría de la variación y el cambio lingüístico** (Labov 1994, 2001; Turell 1995), que postula la existencia de variación lingüística inherente a todas las lenguas y que afecta a todos los niveles lingüísticos: fonética y fonología, morfología, sintaxis, semántica, pragmática y discurso. La variación sociolingüística estudia la forma en que se estructura esta variación, es decir que considera aquellos factores internos (lingüísticos) y externos (sociales) que intervienen en la estructuración de la misma. Diversos estudios han demostrado la influencia del estilo (Schilling-Estes 2002), del género, la edad, la etnicidad y la clase socioeconómica (Labov 1966, 1994, 2001), o la de los grupos o redes sociales (Milroy 1987, Eckert 2000) en las producciones lingüísticas, tanto a nivel fonético y fonológico, como morfológico, sintáctico y semántico. Así, la sociolingüística estudia la estructura de la variación desde diversos niveles y diversas perspectivas, tal como resume el diagrama siguiente:

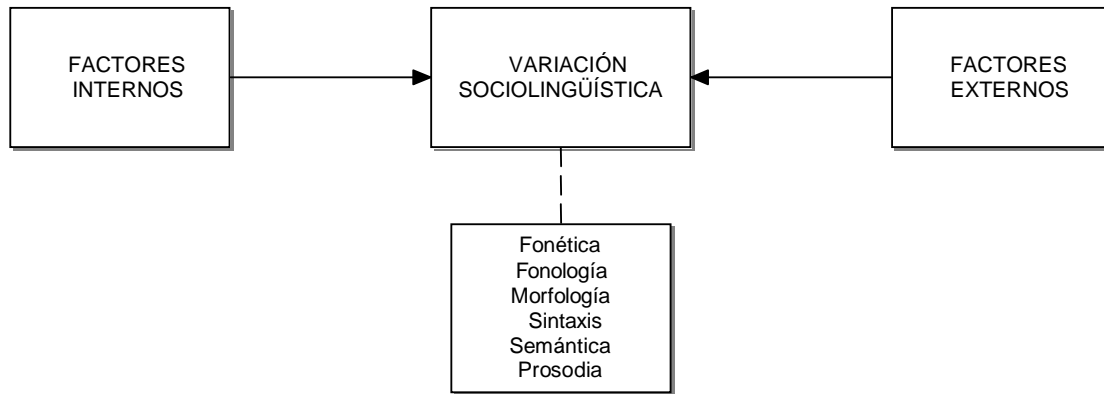


Figura 2. Esquema de los factores que intervienen en la estructuración de la variación lingüística.

No obstante, la variación sociolingüística no se ha centrado de manera prominente en el estudio del idiolecto, a excepción de algunos estudios más bien generalistas como el de Guy (1980), Coulthard (2004), Abercrombie (1969) o Biber (1988, 1995). De ahí el interés que puede suscitar este proyecto, que tiene como objeto de estudio básico el estilo idiolectal en diversas lenguas.

1.3 Miembros y módulos analizados

- **Catalán:**
 - Módulo fonológico del catalán: Dr. Jordi Cicres – UPF y UdG.
 - Módulo fonológico del español: Dra. Fernanda López – UPF y UNAM.
 - Módulo fonológico del inglés: Núria Gavalda – UPF.
- **Español:**
 - Módulo morfosintáctico del catalán: Dra. Montse Forcadell – UPF y UB.
 - Módulo morfosintáctico del español: Dra. Maria Spassova – UPF y NBU.
 - Módulo morfosintáctico del inglés: Dra. M. Teresa Turell – UPF.
- **Inglés:**
 - Módulo discursivo-pragmático del catalán: Sheila Queralt – UPF, desde 2010.
 - Módulo discursivo-pragmático del español: Dra. Raquel Casesnoves – UPF, hasta 2010.
 - Módulo discursivo-pragmático del inglés: David García Barrero – UPF, desde 2011.

2. Objetivos e hipótesis

2.1. *Originalidad e interés*

Este proyecto, que podemos enmarcar en la lingüística aplicada, es claramente ‘problem-based’. Es decir que, a partir de la detección de un problema que existe en nuestra sociedad —la necesidad de identificar hablantes o escritores, y determinar/atribuir la autoría de textos escritos con finalidades forenses— se utilizan una serie de conocimientos lingüísticos, a todos los niveles (fonético, fonológico, morfológico, sintáctico, semántico, pragmático y discursivo) para solucionarlo. Así pues, los resultados de este proyecto podrán usarse en casos reales en que la Justicia necesite la pericia de lingüistas para:

- identificar hablantes,
- construir perfiles lingüísticos a partir de fragmentos de voz o texto,
- determinar la autoría de textos escritos (por ejemplo, notas de suicidio o amenazas).

Los conocimientos actuales en lingüística forense permiten realizar las tareas que se acaban de mencionar con ciertas limitaciones. No obstante, las conclusiones del dictamen pericial siempre han de ser necesariamente conservadoras para evitar la posibilidad de falsas identificaciones. En la actualidad, no hay ningún modelo lingüístico que tenga en cuenta todos los módulos de la gramática con finalidades forenses, aunque hay algunos modelos específicos para las grabaciones, que no entran a valorar aspectos lingüísticos, sino puramente físicos de la voz y otros modelos específicos para textos escritos como el CopyCatch, que ayudan al experto lingüista a detectar el plagio, o incluso otros que analizan el estilo de diversos textos y pueden ser usados como complemento para el análisis forense. La aplicación que se diseñará en el marco de este proyecto permitirá obtener un grado de seguridad mayor que el actual en los dictámenes periciales, revirtiendo en una mejor administración de la Justicia.

2.2. *Hipótesis*

1. Se postula la existencia del idiolecto, como estilo de un hablante y/o escritor único e irrepetible.
2. Se postula la existencia de mayor variación lingüística entre hablantes/escritores (*inter-speaker/writer variation*) que en el discurso de un mismo hablante/escritor (*intra-speaker/writer variation*).
3. Se postula que este estilo idiolectal, sobre todo en relación a sus pautas fonológicas y sintácticas, no varía demasiado en función del género textual o contexto lingüístico. Otra cuestión es el vocabulario, que sí puede tener implicaciones distributivas relacionadas con el registro y el género textual.
4. Se postula que el estilo idiolectal de un individuo se mantiene relativamente estable a lo largo del tiempo.
5. A partir de la confirmación de las hipótesis anteriores, se postula la posibilidad de establecer un índice de Similitud Idiolectal (ISI) que permita identificar hablantes y determinar/atribuir la autoría de textos escritos con mayor fiabilidad.

2.3. *Objetivos generales*

1. Demostrar que todo hablante y escritor tiene su **estilo idiolectal**, es decir, un comportamiento lingüístico diferenciado y único, cuyo uso es inconsciente y que se mantiene estable en cierta medida a lo largo del tiempo.
2. Aplicar el estudio del estilo idiolectal al campo de la **Lingüística forense**, ya que dicha aplicación puede ayudar a identificar de forma más fiable a los productores de un texto oral y/o los escritores de un texto escrito.

2.4. *Objetivos específicos (teóricos y metodológicos):*

1. Llevar a cabo esta aplicación a partir de la comparación entre las formas/parámetros lingüísticos usados por los hablantes y/o escritores en una grabación o texto escrito, para a continuación poder aplicar este mismo protocolo a textos orales y escritos reales, tanto **dubitados** (es decir, cuya autoría presenta dudas) como **indubitados** (es decir, cuya autoría es conocida), y así confirmar la existencia de mayor variación idiolectal inter-hablante/escritor y menor variación idiolectal intra-hablante/escritor.
2. Llevar a cabo esta aplicación a partir de esa misma comparación, planteada en dos tiempos de medición —**tiempo aparente** y **tiempo real**— para confirmar también que el estilo idiolectal se mantiene estable a lo largo del tiempo y con muestras de diversos géneros textuales, para así confirmar que el estilo idiolectal de un mismo individuo varía poco en muestras de diversos géneros textuales.
3. Medir la diferencia lingüística entre diversos estilos idiolectales y la distancia idiolectal de cada individuo, de manera que al final se pueda conseguir un **Índice de Similitud Idiolectal (ISI)**. Más específicamente, se trata de poder establecer el nivel de similitud idiolectal a partir del cual se puede considerar que dos muestras lingüísticas (orales o escritas) han sido producidas o no por una misma persona.
4. Crear un **protocolo** que permita establecer ese **Índice de Similitud Idiolectal (ISI)**, que compare diversas muestras lingüísticas (de diferente longitud y género) y calcule la distancia lingüística, y concluir, una vez analizados diversos parámetros lingüísticos, si dos grabaciones o dos textos escritos han sido producidos por una misma persona.

3. Diseño experimental

En el marco de este proyecto las actividades a desarrollar, derivadas de los objetivos expuestos más arriba, son las siguientes:

Actividades de evaluación (a partir de los resultados del proyecto EXPLORA-INGENIO, HUM2007-29140-E):

- Se evaluará el diseño global del protocolo.
- Se evaluará el módulo fonológico para el español y el catalán.

Actividades de nueva ejecución:

- Se diseñará la técnica para aplicarla a las tres lenguas de análisis (catalán, español e inglés).
- Se elaborará el diseño del protocolo para los módulos morfosintáctico y discursivo-pragmático de las tres lenguas.
- Se diseñarán las pruebas estadísticas que deben permitir el cálculo del ISI para los módulos morfosintáctico y discursivo-pragmático de las tres lenguas.
- Se evaluará evaluarán los tres módulos para las tres lenguas.

La metodología a seguir en este proyecto consiste en el diseño de un protocolo para calcular el **ISI (índice de similitud idiolectal)** entre dos o más textos orales o escritos. Por tanto, es preciso establecer en primer lugar los corpóra de análisis, las variables de estudio, y las técnicas de análisis lingüística que se van a considerar. Así, primero se presentan los corpóra que se van a usar en cada lengua y a continuación se hace referencia a cada uno de los módulos a partir de las variables analizadas y la metodología de análisis usada.

3.1. Corpóra

El proyecto que se presenta pretende estudiar el estilo idiolectal con finalidades forenses, para identificar tanto hablantes como escritores. Por tanto, será preciso utilizar corpóra orales y escritos (y en este último caso, seguiremos las directrices de Coulthard (1994), Eagleson (1994), Turell (1995) y Johnson (1997)).

Cuando se habla de corpóra orales, es necesario distinguir entre corpóra de la lengua oral (en inglés, *spoken language corpora*) y corpóra orales (Llisterri *et al.* 2005). Los primeros son transcripciones que pretenden representar unos aspectos determinados de la lengua oral, mientras que los segundos están constituidos, como mínimo, por una señal sonora. El grado de segmentación y etiquetaje varía según el propósito de uso del corpus. Así, los corpóra se pueden segmentar por fonemas, difonemas, morfemas, sintagmas, unidades de entonación, oraciones, turnos de palabra, etc. Por otro lado, también pueden incorporar otros tipos de información para representar otros aspectos verbales (tono, velocidad de la elocución, entonación, pausas) o no verbales (cambios de turnos de palabra, ruidos de fondo, gestos).

Para estudiar el efecto del paso del tiempo sobre las variables analizadas, teniendo en cuenta el marco aplicado en lingüística forense en que se encuadra este proyecto, es necesario usar no solo un corpus en un tiempo de medición (que equivaldría al *tiempo aparente* en variación sociolingüística), que consistiría en un corpus con información lingüística de varios informantes obtenido en un punto temporal concreto, sino también un corpus en un segundo tiempo de medición (*tiempo real*), es decir, un corpus formado por material obtenido a partir de los mismos informantes, con un intervalo de tiempo entre las diferentes muestras. La noción de tiempo aparente y tiempo real se ha venido usando, especialmente en el campo de la sociolingüística (Turell 2003), desde los años treinta, y sobre todo desde los años cincuenta del siglo XX.

3.1.1. Corpus para el ISI en catalán

Para todos los módulos del catalán se usará la parte en catalán de un corpus oral en tiempo aparente y tiempo real: el corpus de La Canonja (Pujadas, Pujol Berché, Turell), recogido por el grupo UVAL (Unitat de Variació Lingüística) del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. Inicialmente, este corpus estaba constituido por 29 grabaciones (13 de hablantes catalanohablantes y 16 de hablantes castellanohablantes) realizadas a mediados de los años ochenta del siglo pasado. En el proceso de realización de este proyecto se realizará una nueva tanda de grabaciones, de manera que se podrán estudiar los mismos hablantes con un intervalo temporal de más de 20 años.

3.1.2. Córpora para el ISI en español

Para la identificación de hablantes se usará la parte en español de un corpus oral en tiempo aparente y tiempo real: el corpus de La Canonja, recogido por el grupo UVAL (Unitat de Variació Lingüística) del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. Inicialmente, este corpus estaba constituido por 29 grabaciones (13 de hablantes catalanohablantes y 16 de hablantes castellanohablantes) realizadas a mediados de los años ochenta. En el proceso de realización de este proyecto se realizará una nueva tanda de grabaciones, de manera que se podrán estudiar los mismos hablantes con un intervalo temporal de más de 20 años.

También se usará el corpus en tiempo aparente y real para el español de México que se está recogiendo a partir de una beca CONACIT, concedida a Fernanda López para la realización de su tesis doctoral. Se trata del corpus DIMEx100, compilado por el Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas de la Universidad Nacional Autónoma de México, y que tenía como principal objetivo utilizarse para crear un reconocedor automático de habla. El corpus fue grabado por 100 hablantes que leyeron 10 frases iguales para todos y 50 frases distintas para cada uno. Los hablantes fueron seleccionados según su edad (de 16 a 36 años), su nivel escolar (estudios superiores a secundaria), y su lugar de origen (Ciudad de México).

Para el análisis del módulo morfosintáctico del español se usará el corpus de textos escritos compilado por María Spassova para la realización de su tesis doctoral. Dicho corpus está compuesto por textos escritos en español de hasta 20 autores hispanohablantes nativos de procedencia peninsular y sudamericana, con dos subcórpora: textos de narrativa literaria (N) y textos de artículos de opinión (AO).

Y para el módulo discursivo-pragmático se usará el corpus Preseva (*Proyecto para el Estudio Sociolingüístico del Español de Valencia*, <http://www.uv.es/preseval/ppal.htm>), del que se considerarán informantes de nivel sociocultural medio y alto, y se controlará por género biológico y edad (Gómez Molina, J. R. 2001 y 2005).

3.1.3. Córpora para el ISI en inglés

Para la identificación de hablantes se usará un corpus de 20 hablantes (10 británicos y 10 norteamericanos de EE.UU) y otras muestras orales del BNC (British National Corpus). Se prevé

también el uso de grabaciones reales de entrevistas de radio y TV a personajes del mundo cultural y artístico del Reino Unido

En el caso de la determinación/atribución de autoría se usarán textos escritos procedentes de diversos córpora existentes y otros córpora escritos derivados de casos reales ya sentenciados que nos proporcionará el grupo de investigación del Aston Centre for Forensic Linguistics.

3.2. Variables

Respecto a las variables de estudio, nos centraremos inicialmente en las variables fonológicas. Rose (2002) plantea la complejidad del análisis comparativo de las muestras de voz, y la necesidad de analizar diversas variables, con diversos gráficos que se presentan a continuación. En primer lugar describe una situación ideal en la que las realizaciones de una variable concreta son únicas para cada hablante, tal como muestra la figura 3, a partir de datos inventados para la variable mediana de la F0:

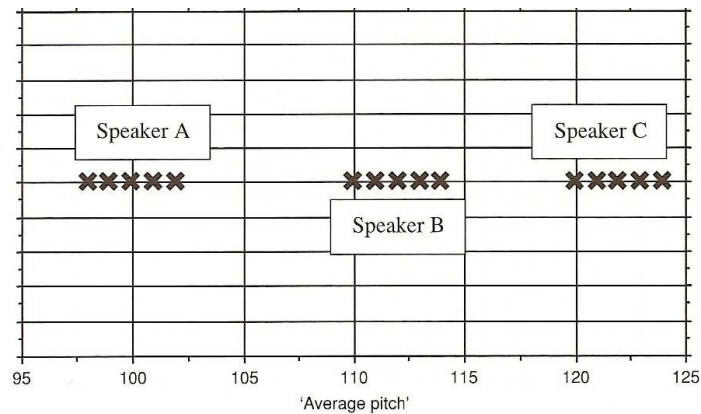


Figura 3. Representación “ideal” de los valores de una sola variable en tres hablantes. Fuente: Rose (2002).

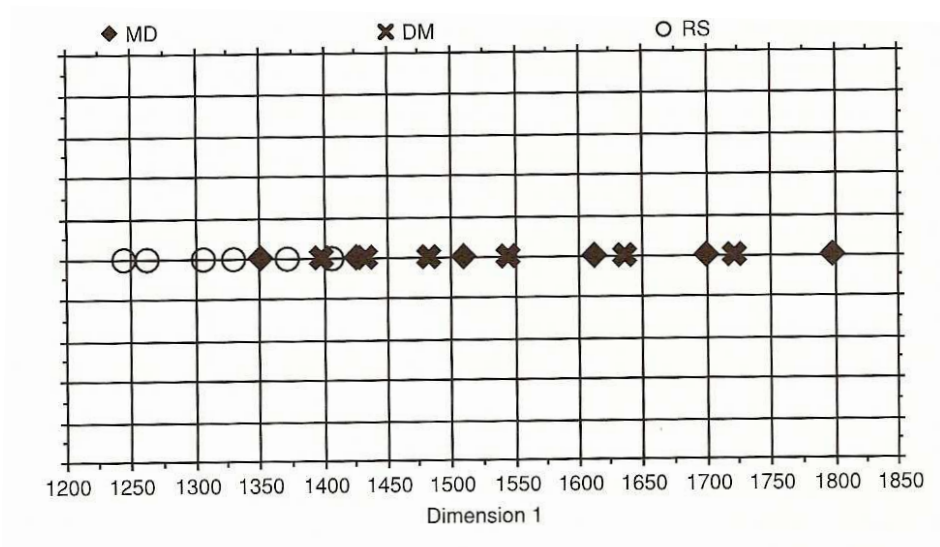


Figura 4. Representación “realista” de los valores de una sola variable en tres hablantes, en la que se produce superposición de valores. Fuente: Rose (2002).

En la figura 4, se observa como los valores de una variable concreta (una dimensión, según terminología de Rose 2002) no son suficientes para discriminar hablantes, ya que la superposición de datos no permite llegar a ninguna conclusión en lo que a atribución de autoría se refiere. Por tanto, se deben introducir nuevas dimensiones en los estudios de identificación de hablantes, para obtener espacios delimitados para cada hablante. Es lo que se muestra en las figuras 5 (2 dimensiones) y 6 (3 dimensiones).

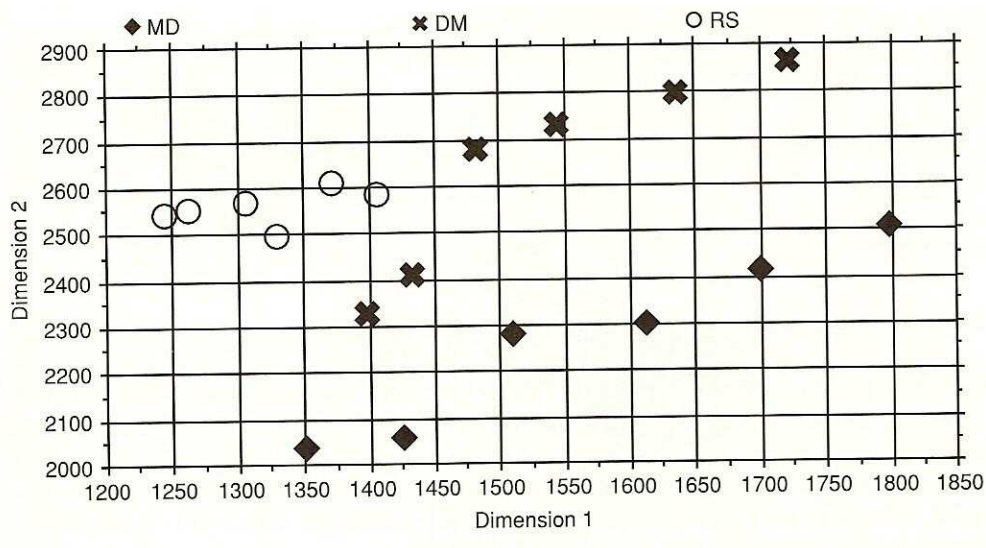


Figura 5. Representación de los valores de dos variables en tres hablantes. Fuente: Rose (2002).

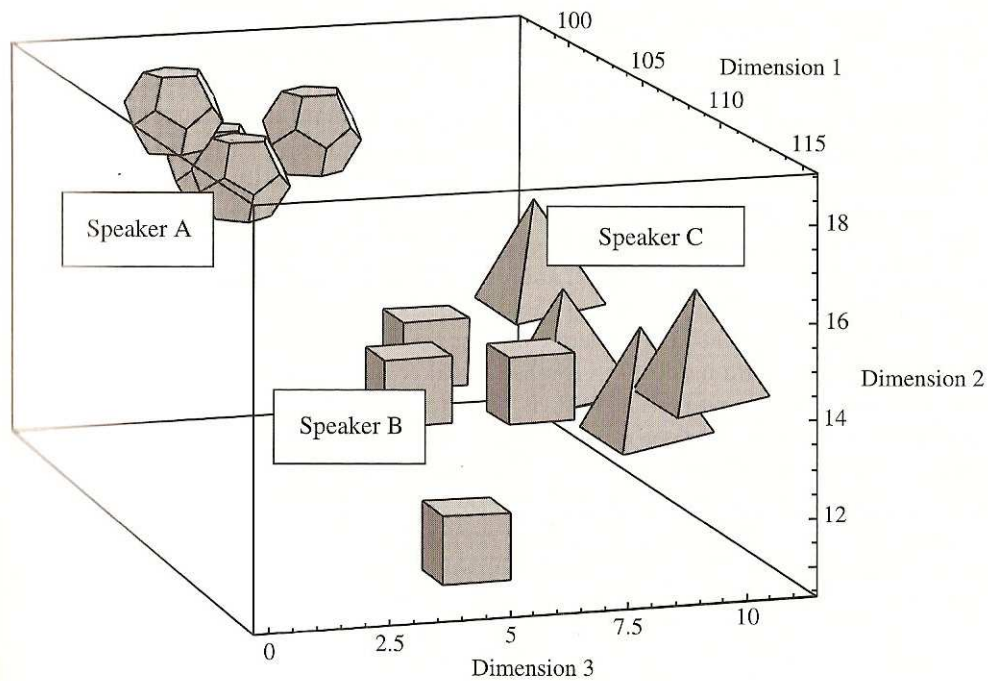


Figura 6. Representación de los valores de tres variables en tres hablantes. Fuente: Rose (2002).

No obstante, tres dimensiones tampoco serán suficientes para poder llegar a conclusiones fiables sobre la autoría de una grabación. Es imprescindible aumentar el número de variables para que la discriminación de los hablantes sea más completa.

Por otro lado, Rose (2002: 33-53) clasifica el análisis forense de la voz en cuatro grandes grupos: auditivo lingüístico, auditivo no lingüístico, acústico lingüístico y, finalmente, acústico no lingüístico:

Tabla 1. Clasificación primaria de los parámetros en el análisis forense según Rose (2002).

	<i>LINGÜÍSTICO</i>	<i>NO LINGÜÍSTICO</i>	
<i>AUDITIVO</i>	AUDITIVO LINGÜÍSTICO	AUDITIVO LINGÜÍSTICO	NO
<i>ACÚSTICO</i>	ACÚSTICO LINGÜÍSTICO	ACÚSTICO LINGÜÍSTICO	NO

En este contexto, ‘lingüístico’ se usa como sinónimo de ‘fonemático’, es decir, que tiene una función dentro de la lengua. Según esta clasificación, los parámetros acústicos o auditivos no lingüísticos corresponderían a los parámetros fonéticos (segmentales y suprasegmentales), ya que el hecho de pronunciar un determinado alófono con unas características espectrales, temporales y de intensidad concretas no influye en su contenido lingüístico. De igual forma, los parámetros del tercer grupo, referidos a la ‘forma de hablar’, no tienen un valor lingüístico en este sentido. En cambio, todo el grupo de parámetros fonológicos se considerarán en el contexto de los grupos auditivo lingüístico y acústico lingüístico, ya que las unidades básicas en las que se basa el análisis son los fonemas (y no las realizaciones acústicas de estos fonemas).

Estas variables han de cumplir las siguientes características, según Nolan (1983: 11):

- Presentar una alta variabilidad *inter-hablantes* y una variabilidad baja *intra-hablante*.
- Ser resistentes a los intentos de imitación.
- Tener una alta frecuencia de ocurrencia en las muestras de análisis.
- Ser robustas en la transmisión (es decir, que la transmisión no altere o elimine las propiedades acústicas).
- Ser relativamente fáciles de extraer y analizar.

Además, Rose (2002: 52) añade una nueva condición:

- Cada parámetro debe ser independiente de los otros

Consúltense la lista completa de variables para cada módulo en el ANEXO: VARIABLES.

4. RESULTADOS

4.1 *Desarrollo de los objetivos planteados*

Objetivo 1

Demostrar que hay más distancia (variación lingüística) entre hablantes y escritores (**inter**) que en el habla o la escritura de un mismo hablante o escritor (**intra**).

Progreso y consecución del objetivo 1

En todos los módulos (fonológico, morfosintáctico y pragmático-discursivo) de cada lengua considerada (catalán, español e inglés) —excepto en el caso del módulo pragmático-discursivo del español con solo un tiempo de medición— y con los tres métodos estadísticos usados, los valores del **ISI** que corresponden a las comparaciones de producciones lingüísticas de un mismo individuo son más altos (**intra: IIS > 0.9 y 0.8**) que los valores obtenidos cuando se comparan dos individuos (**inter: IIS > 0,6 y < 0,8**), con lo que se confirma la hipótesis 1 y se alcanza el objetivo 1. El hecho de que estos valores **inter** del **ISI** no sean más bajos se explica

Objetivo 2

Demostrar que el estilo idiolectal de un hablante/escritor se mantiene relativamente estable a lo largo del tiempo.

Objetivo 3

Demostrar que el estilo idiolectal de una hablante/escritor es menos estable cuando se comparan géneros textuales producidos por un mismo individuo que cuando se comparan producciones de ese mismo individuo en dos tiempos de medición.

metodológicamente, en el sentido de que todos los hablantes/escritores considerados en los tres módulos y lenguas pertenecen al mismo dialecto, ya que no fue posible recoger muestras de hablantes/escritores de más de un dialecto que además estuvieran estratificadas en dos tiempos de medición (**MT1** y **MT2**). Esta limitación metodológica se explicará y justificará en el punto C3.

Progreso y consecución del objetivo 2

En todos los módulos (fonológico, morfosintáctico y pragmático-discursivo) de cada lengua considerada (catalán, español e inglés), y con los tres métodos estadísticos usados, los valores del **ISI** que corresponden a las comparaciones de producciones lingüísticas de un mismo individuo son altos (**intra: IIS > 0.9** y **0.8**), lo cual indica que el estilo idiolectal de un individuo (hablante/escritor), a partir de las variables consideradas en cada módulo para establecer el **ISI**, no varía substancialmente a lo largo del tiempo, se confirma, por tanto, la hipótesis 2 y se alcanza el objetivo 2.

Progreso y consecución del objetivo 3

Como se ha comentado más arriba este objetivo no se ha alcanzado, a causa de una limitación metodológica, en el sentido de que no fue posible estratificar todas las muestras de todos los módulos y lenguas no solo con dos tiempos de medición sino también por al menos dos géneros textuales. Esta limitación metodológica se explicará y justificará en el punto C3.

4.2 Actividades realizadas y resultados alcanzados

El proyecto se ha desarrollado a partir del concepto de **Índice de Similitud (o Distancia) Idiolectal**, formalizado en la **Figura 7**.

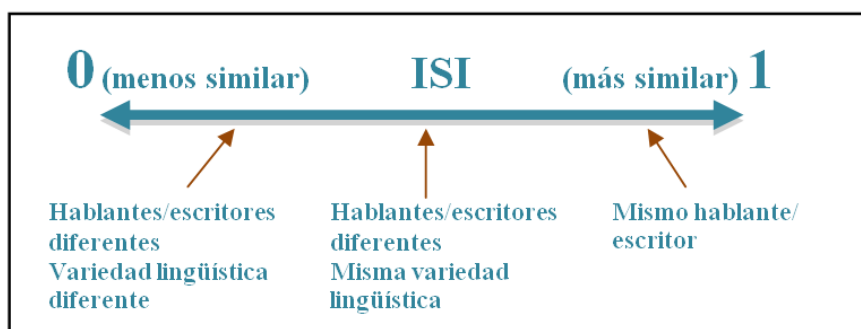


Figura 7. Representación del ISI

5 La recogida de datos de los diferentes **córpora** se ha basado en la técnica laboviana de entrevista sociolingüística o técnicas muy similares, así como en la explotación de córpore institucionales o descargables por Internet, todos ellos de habla semi-espontánea.

6 Los **métodos estadísticos**¹ que se han usado son los siguientes:

Método 1

Comparación del porcentaje de ocurrencias de cada variante: a partir del cálculo y comparación del porcentaje de ocurrencia de las variantes de cada variable para cada hablante.

Método 2

Cálculo del Valor Ajustado residual (en inglés ARV): a) cálculo de tablas de contingencia y pruebas de la Chi-cuadrada; b) obtención del Valor Ajustado Residual (ARV), que indica si los dos hablantes analizados tienen una distribución similar de las variantes o no; cuánto mayor es la diferencia entre los hablantes comparados, mayor es el ARV; c) ya que el ARV puede ser infinito, asignación de un valor a cada rango de ARV hasta a un máximo de 5; y d) el cálculo del **ISI** se obtiene mediante el cálculo de la suma de valores de cada variable dividido entre el número de variables multiplicado por 5 (el valor máximo de conversión ARV), según la fórmula siguiente:

$$IIS = \frac{ARV \text{ (variable 1)} + \dots + ARV \text{ (variable n)}}{(\text{max. ARV conversion}) * n}$$

¹ Durante el proceso experimental, también se ha aplicado el método de distancia Euclidiana, que finalmente se ha descartado en el proceso de evaluación por no haber arrojado resultados positivos en todos los módulos y lenguas.

Método 3

Cálculo del coeficiente de Phi: a) cálculo de las pruebas de la Chi-cuadrada y obtención del coeficiente de Phi, que se sitúa entre 0 y 1 e indica la fuerza de la relación entre las variables. El cuadrado del coeficiente de Phi representa el porcentaje de varianza conjunta; b) el **ISI** se obtiene mediante el cálculo de la media de cuadrados de Phi:

$$ISI_{Total} = 1 - \sqrt{\frac{1}{k} \sum_j^k Phi^2}$$

4.3 Resultados por módulos

4.3.1 Módulos fonológicos

- **Módulo fonológico del catalán:** 6 hablantes (4 del dialecto de La Canonja, comunidad de habla del tarragonés, y 2 del dialecto barcelonés); 18 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Investigador responsable de este módulo: **Dr. Jordi Cicres Bosch**.
- **Módulo fonológico del inglés:** 6 hablantes (todos del dialecto del sur de Inglaterra); 14 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado por **Núria Gavalrà**, uno de los miembros investigadores del proyecto, becaria FPU con tesis doctoral en curso, realizada en el marco del proyecto, y a punto de finalizar.
- **Módulo fonológico del español:** 5 hablantes del mismo dialecto mexicano, 5 en MT1 y 5 en MT2; 13 variables y 3 métodos estadísticos aplicados. Analizado por la **Dra. Fernanda López**, miembro del proyecto de investigación.

Los resultados del **ISI** obtenidos a partir de la aplicación de los 3 métodos a este conjunto de módulos fonológicos confirman la hipótesis 1, en el sentido de que los valores del **ISI** son bajos en el caso de la variación inter-hablante, como puede observarse en la **Figura 8**, a modo de ejemplo de la aplicación del método 3. En este gráfico los valores oscilan entre **0.2 y 0.80**; los más bajos son los obtenidos en el módulo fonológico del catalán, lo que indica distancia idiolectal. Además, la aplicación de este método al módulo fonológico del catalán ha permitido observar que hay diferencias entre hablantes de diferentes dialectos o variedades, resultado enormemente útil en casos forenses reales.

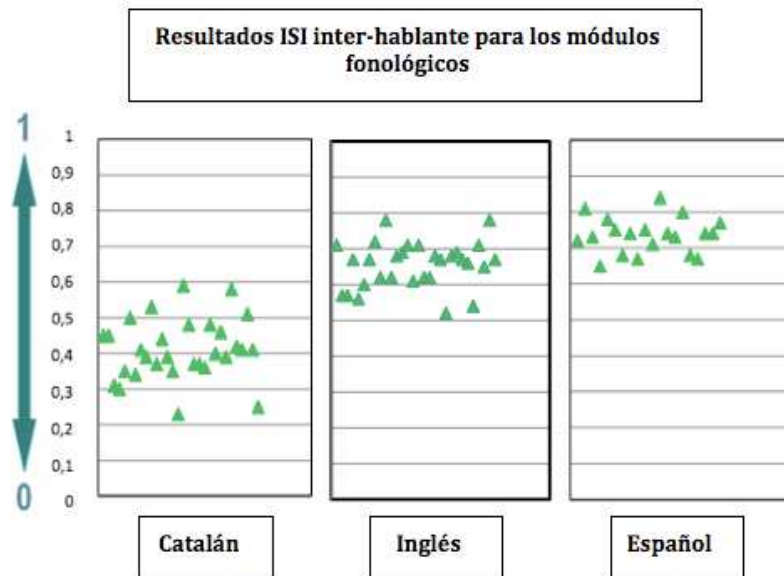


Figura 8. ISI de los módulos fonológicos – Variación inter-hablante

La **Figura 9**, a modo de ejemplo de la aplicación del método 3, muestra que la hipótesis 2 para estos módulos fonológicos se confirma, en el sentido de que no hay grandes diferencias en el estilo idiolectal de los hablantes (variación intra-hablante/escritor) de las tres lenguas analizadas a lo largo del tiempo, ya que los valores oscilan entre **0.9 y 0.8**, como muestran los puntos rojos del gráfico, que indican el cálculo del **ISI** de cada hablante en MT1 y MT2.

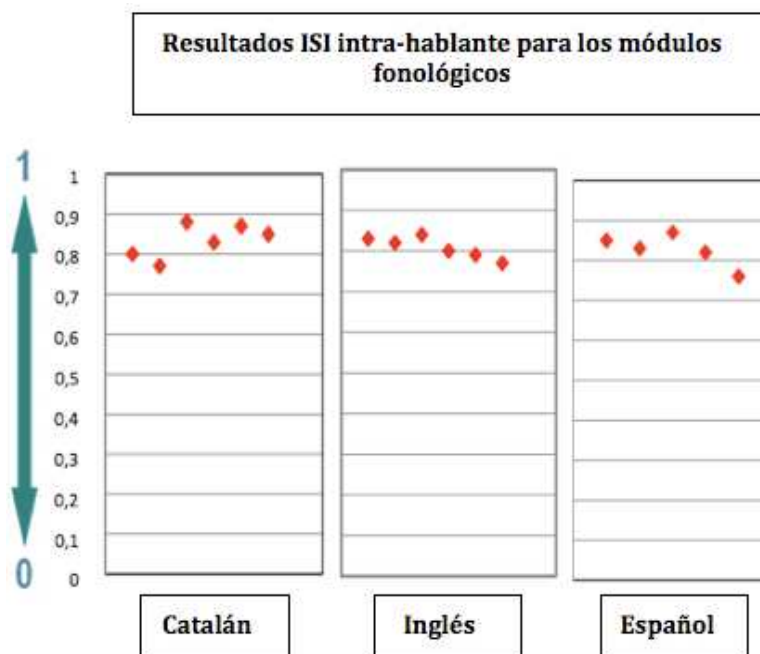


Figura 9. ISI – Fonológicos – Variación intra-hablante

4.3.2 Módulos morfosintácticos

- **Módulo morfosintáctico del catalán:** 6 hablantes (todos del dialecto tarragonés de La Canonja); 7 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado por la Dra. **Montserrat Forcadell**, uno de los miembros investigadores del proyecto.
- **Módulo morfosintáctico del inglés:** 6 hablantes (todos del dialecto del sur de Inglaterra); 7 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado por la Dra. **M. Teresa Turell**, investigadora principal del proyecto.
- **Módulo morfosintáctico del español:** 6 escritores (todos del mismo dialecto del español peninsular); 10 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado por la Dra. **Maria Spassova**, una de los miembros investigadores del proyecto.

Los resultados del **ISI** obtenidos a partir de la aplicación de los 3 métodos a este conjunto de módulos morfosintácticos confirman la hipótesis 1, en el sentido de que los valores del **ISI** son relativamente bajos en el caso de la variación inter-hablante, ya que la mayoría oscila entre **0.8 y 0.6**, tal y como se esperaba, y como puede observarse en la **Figura 10**, a modo de ejemplo de la aplicación del método 3.

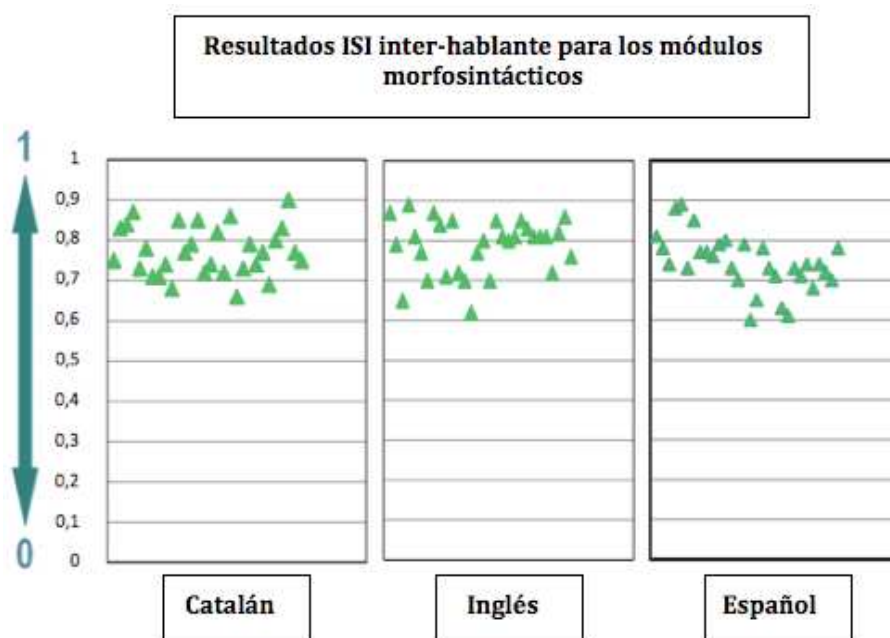


Figura 10. ISI – Morfosintácticos – Inter-hablante/escritor

La **Figura 11**, a modo de ejemplo de la aplicación del método 3, muestra que la hipótesis 2 para estos módulos morfosintácticos se confirma, en el sentido de que no hay grandes diferencias en el estilo idiolectal de los hablantes de las tres lenguas analizadas a lo largo del tiempo, como muestran los puntos rojos del gráfico, que indican el cálculo del **ISI** de cada hablante en MT1 y MT2, ya que los valores la mayoría de los hablantes oscilan entre **0.9 y 0.8**, tal y como se esperaba.

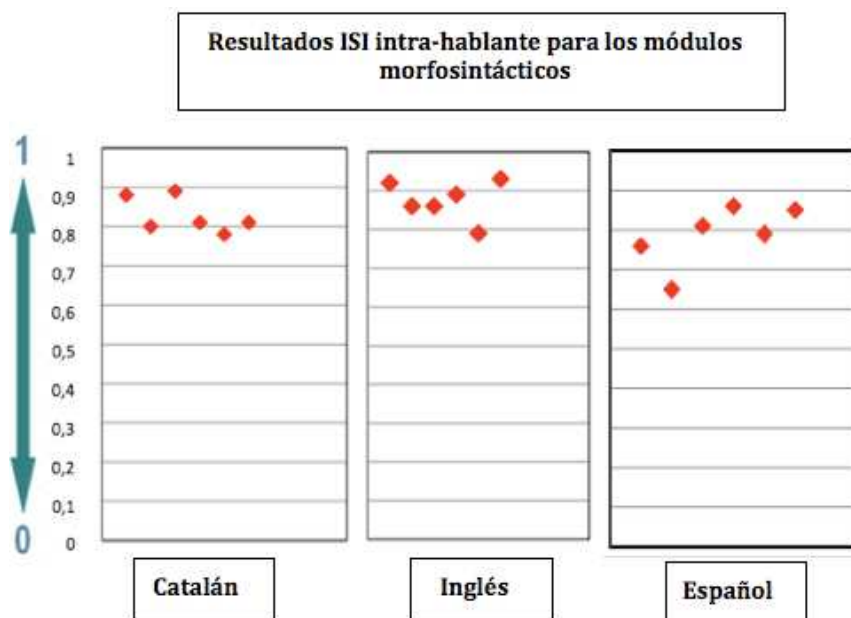


Figura 11. ISI – Morfosintácticos – Intra-hablante/escritor

4.3.3 Módulos pragmático-discursivos

- **Módulo pragmático-discursivo del catalán:** 6 hablantes (todos del dialecto tarragonés de La Canonja; 8 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado por la Dra. **M. Teresa Turell**, investigadora principal del proyecto, con el apoyo de **Sheila Queralt**, una de las investigadoras contratadas del ForensicLab (Institut Universitari de Lingüística Aplicada, UPF).
- **Módulo pragmático-discursivo del inglés:** 6 hablantes (todos del dialecto del sur de Inglaterra); 9 variables, 2 tiempos de medición (MT1 y MT2) y 3 métodos estadísticos aplicados. Analizado durante los últimos 3 meses por **David García Barrero**, nuevo becario FPU del ForensicLab (Institut Universitari de Lingüística Aplicada, UPF), que va a aplicar el **ISI** al árabe, tal como se ha explicitado en la memoria de solicitud de un nuevo proyecto en la convocatoria de 2012.
- **Módulo pragmático-discursivo del español:** 6 hablantes (todos del mismo dialecto del español peninsular); 10 variables, 1 solo tiempo de medición (MT1) y 3 métodos estadísticos aplicados. Analizado por la Dra. **Raquel Casesnoves**, una de los miembros

investigadores del proyecto hasta que causó baja en 2010 al serle concedido un proyecto del Plan Nacional.

Los resultados del **ISI** obtenidos a partir de la aplicación de los 3 métodos a este conjunto de módulos pragmático-discursivos confirman la hipótesis 1, en el sentido de que los valores del **ISI** son bajos en el caso de la variación inter-hablante, oscilando la mayoría entre 0.6 y 0.8, como puede observarse en la **Figura 12**, a modo de ejemplo de la aplicación del método 3.

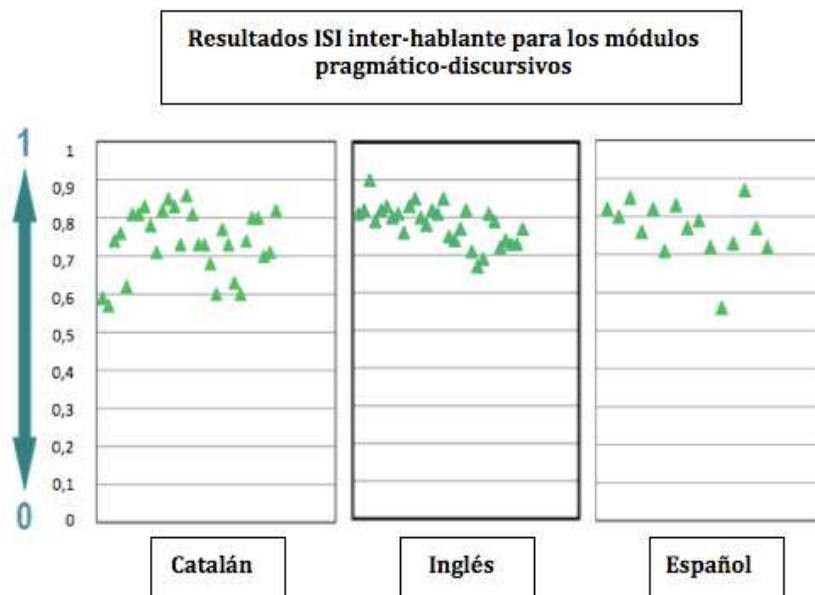


Figura 12. ISI – Pragmático-discursivos – Variación inter-hablante

La **Figura 13**, a modo de ejemplo de la aplicación del método 3, muestra que la hipótesis 2 para estos módulos pragmático-discursivos se confirma, en el sentido de que no hay grandes diferencias en el estilo idiolectal de los hablantes del inglés y del catalán a lo largo del tiempo, como muestran los puntos rojos del gráfico, que indican el cálculo del **ISI** de cada hablante en MT1 y MT2, ya que los valores oscilan entre 0.9 y 0.7. Recuérdese que no se puede calcular el **ISI** intra-hablante para el módulo pragmático-discursivo del español porque no hay datos para el tiempo de medición MT2.

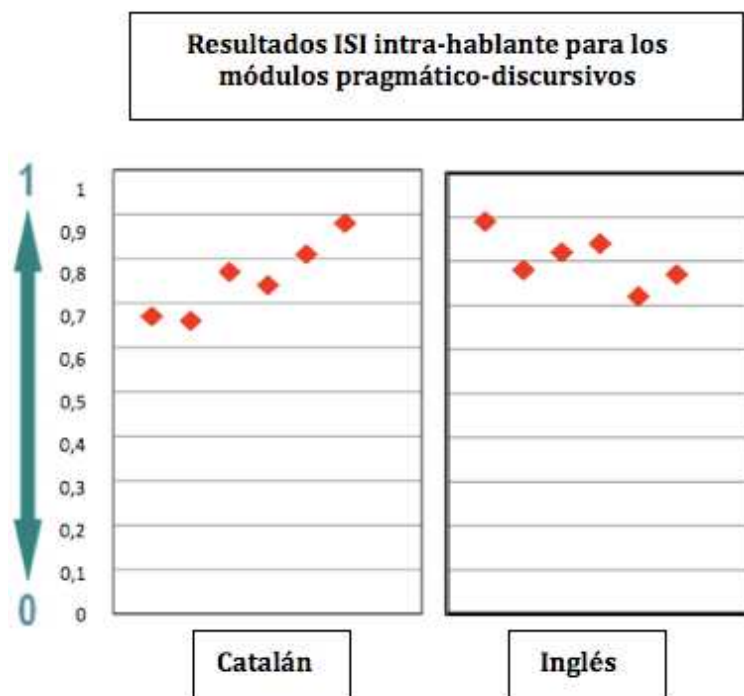


Figura 13. ISI – Pragmático-discursivos – Variación intra-hablante

En el caso del módulo pragmático-discursivo del español, cuyos resultados presentan muchos ceros (0), parece interesante resaltar que el método 3, basado en el cálculo de los valores de la Chi-cuadrada y del valor Phi, funciona especialmente bien, porque a pesar de esa distribución desigual de los datos el análisis puede continuar sin problemas.

4.4 Conclusiones finales

1. El método 3 resulta ser el más fiable y el que permite obtener resultados más concluyentes.
2. Para todos los módulos y lenguas se confirma la hipótesis 1, en el sentido de que hay más variación, y por tanto mayor distancia idiolectal, entre hablantes y escritores (variación inter) que en el estilo idiolectal de un mismo individuo (variación intra).
3. Para todos los módulos y lenguas se confirma la hipótesis 2, en el sentido de que el estilo idiolectal de un individuo no varía demasiado a lo largo del tiempo.
4. Los valores del ISI inter-hablante/escritor pueden parecer demasiado altos, pero hay que tener en cuenta que, excepto en el caso del módulo fonológico del catalán, en todos los módulos se han considerado hablantes y escritores del mismo dialecto.

Referencias

- Abercrombie, D. 1969. 'Voice qualities'. In N. N. Markel (ed.) *Psycholinguistics: an Introduction to the Study of Speech and Personality*. London: The Dorsey Press.
- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP.
- Biber, D. 1995. *Dimensions of Register Variation: a Cross-linguistic Comparison*. Cambridge: CUP.
- Coulthard, M. 2004. 'Author Identification, idiolectal style and Linguistic Uniqueness'. *Applied linguistics* 2004, vol. 25, 4, 431-447.
- Eagleson, R. 1994. Forensic analysis of personal written text: a case study. In J Gibbons (ed.). *Language and the Law*, London: Longman, 362-373.
- Eckert, P. 2000 *Linguistic Variation as Social Practice*. Oxford:Blackwell.
- Guy, G. 1980. Variation in the group and the individual. In W. Labov (ed.), *Locating language in time and space*, New York: Academic Press. 1-36.
- Gómez Molina, J. R. (coord.) 2001. El español hablado de Valencia. Materiales para su estudio. I. Nivel sociocultural alto. Anejo XLVI de Cuadernos de Filología. Universitat de València.
- Gómez Molina, J. R. (coord.) 2005. El español hablado de Valencia. Materiales para su estudio. II. Nivel sociocultural medio. Anejo LVIII de Cuadernos de Filología. València: Universitat de València.
- Johnson, A. 1997. Textual kidnapping – a case of plagiarism among three student texts, *Forensic Linguistics: The International Journal of Speech, Language and the Law* 4: 210-25.
- Labov, W. 2006 [1966]. *The Social Stratification of English in New York City*. Cambridge, U.K.: Cambridge University Press.
- Labov, W. 1994. *Principles of Linguistic Change, I: Internal Factors*. Oxford: Blackwell.
- Labov, W. 2001. *Principles of Linguistic Change: External Factors*. Vol 2. Oxford UK and Cambridge USA: Blackwell.
- Llisterri *et al.* 2005. "Corpus orales para el desarrollo de las tecnologías del habla en español". *Oralia. Anàlisis del discurso oral* 8.
- McMahon, A.M.S. and Foulkes, P. 1995. Sound change, phonological rules, and Articulatory Phonology. *Belgian Journal of Linguistics* 9: 1-20.
- Milroy, L. 1987. *Language and social networks*. Second ed. Oxford: Blackwell.
- Nolan, F. J. 1994. Auditory and acoustic analysis in speaker recognition. In Gibbons, J. (ed.), *Language and the law*. London/New York: Longman. 326-345.
- Rose, Ph. 2002. *Forensic Speaker Identification*. Londres: Taylor & Francis.
- Schilling-Estes, Natalie. 2002. Investigating stylistic variation. In J.K. Chambers, P. and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 375-401). Malden: Blackwell.
- Turell, M.T. (ed.) 1995. *La sociolingüística de la variació*. Barcelona: Promociones y Publicaciones Universitarias.
- Turell, M. T. 2003. El temps aparent i el temps real en estudis de variació i canvi lingüístic. *Noves de Sociolingüística Catalana*, Autumn 2003. http://www6.gencat.net/llengcat/noves/hm03tardor/turell1_4.htm

ANEXO: VARIABLES

Tabla 1: Variables para el módulo fonológico del catalán

VARIABLE	VARIANTES	EJEMPLOS
1 Cambio de /ə/ a [i] ante fricativa palatal	1. [ə] 2. [i]	<i>seixanta, deixar</i>
2 Pérdida de la oclusiva en el grupo /ks/	1. [ks] 2. [s]	<i>expressament,</i>
3 Pérdida de la oclusiva en el grupo /ɬ/	1. [ɬ] 2. [z]	<i>organitzar,</i>
4 Pérdida de oclusiva en los grupos /rt(s), lt(s), st/ a final de palabra	1. [rt, lt, st] 2. [r, l, s]	<i>verd, malalt, vist, part, dalt, just</i>
5 Pérdida del grupo /tʃ/ correspondiente al morfema de primera persona del presente de indicativo de algunos verbos de la segunda conjugación	1. [tʃ] 2. ∅	<i>veig, faig</i>
6 Pérdida de /s/ en la palabra aquesta	1. aque[st]a 2. aque[t]a	<i>aquesta</i>
7 Confusión de /b,v/ en /b/	1. distinción entre /b/ y /v/ 2. confusión en /b/	<i>(cualquier palabra con estos fonemas)</i>
8 Yodización en grupos de palabras (jo, ja, vull)	1. Yodización 2. No yodización	<i>jo, ja, vull</i>
9 Sensibilización de /r/ finales en sustantivos	1. [r] 2. ∅	<i>por, carnisser,</i>
10 Geminación del grupo /bl/ intervocálico	1. [bbl] 2. [βl]	<i>poble, cable</i>
11 Africación de /ʒ/ intervocálica	1. [ʒ] 2. [ʒ̥]	<i>ajuda, apujar, ajuntament,</i>
12 Africación del grupo /rʃ/	1. [rʃ] 2. [rʃ̥]	<i>arxiu, parrís,</i>
13 Africación /tʃ/	1. [tʃ] 2. [tʃ̥]	<i>xai, xocolata, xerrada,</i>
14 Ensordecimiento de la fricativa alveolar sonora /z/	1. [z] 2. [s]	<i>casino</i>
15 Africación de las africadas sonoras	1. [ʒ] 2. [ʒ̥]	<i>gent</i>

Tabla 1: Variables para el módulo fonológico del catalán

VARIABLE	VARIANTES	EJEMPLOS
16 Cambio del grupo /gz/ a [ʒ]	1. [gz] 2. [ʒ]	<i>examen,</i>
17 Presencia de /n/ en el grupo de plurales arcaicos	1. home[ns] 2. home[s]	<i>hòmens, jóvens</i>
18 Sonorización de /s/ por ultracorrección	1. [s] 2. [z]	<i>expressar</i>

Tabla 2: Variables para el módulo fonológico del español

VARIABLE	VARIANTES	EJEMPLOS
1 Pérdida de aproximante dental sonora /ð̞/ intervocálica de palabra	1. [ø] 2. [ð̞]	<i>Acostumbrada, todo, acostumbrada.</i>
2 Pérdida de la aproximante bilabial sonora /β̞/ intervocálica	1. [ø] 2. [β̞]	<i>Quedaba, sabes.</i>
3 Pérdida de la primera oclusiva sorda en /ks/, /pt/, /kt/	1. [s][t][t] 2. [ks][pt][kt]	<i>Octavo, acción.</i>
4 Pérdida de /r/ implosiva	1. [ø] 2. [r]	<i>Parte, porque</i>
5 Pérdida de /r/ intervocálica	1. [ø] 2. [r]	<i>Para, aceptaron, quiere</i>
6 Adición de oclusiva dental sorda /t/ al inicio de monosílabos que	1. [ts] 2. [s]	<i>Sí, se, sin</i>
7 Aspiración de /s/ implosiva	1. [h] 2. [s]	<i>Mismo, más</i>
8 Asibilamiento de /r/ implosiva	1. [ʃ] 2. [r]	<i>Persona, parte, mayor</i>
9 Simplificación de la vibrante múltiple intervocálica [r] a vibrante simple [r]	1. [r] 2. [r]	<i>Tierra, agarrar, torre</i>
10 Cierre vocálico de /o/ átona a [u] en sílaba libre /n/ seguida de /o/	1. [nu] 2. [no]	<i>Bueno, humano</i>

Tabla 2: Variables para el módulo fonológico del español

VARIABLE	VARIANTES	EJEMPLOS
11 Pérdida de la vocal /e/ átona al interior de la palabra /entonses/ à [entons]	1. [en'tonses] 2. [en'tons]	
12 Reducción del hiato /ea/ a /a/ en la palabra /osea/	1. [o'sa] 2. [o'sea]	
13 Cambio de la palabra /entonces/ por /tons/	1. [en'tonses] 2. [tons]	

Tabla 3: Variables para el módulo fonológico del inglés

VARIABLE	VARIANTES	EJEMPLOS
----------	-----------	----------

[Análisis y datos derivados de una tesis doctoral en curso.]

Tabla 4: Variables para el módulo morfosintáctico del catalán

VARIABLE	VARIANTES	EJEMPLOS
1 Negació amb “pas”	1. pas	“ <i>Jo que no hi és pas aquest problema de...</i> ” (VC) -RT
	2. \emptyset	“ <i>sembla que no puguin anar aquí o allà</i> ” (DC) -RT
2 Realització del pronom del subjecte	1. subjecte	“ <i>jo ho vai descobrir no fa gaire, eh.</i> ” (DC) -RT
	2. \emptyset	“ <i>No érem aquí, érem a Vila-seca.</i> ” (VC) -RT
3 Concordança nom-verb col·lectius	1. singular	“ <i>No hi ha massa problemes de convivència</i> ” (VC) -RT
	2. plural	“ <i>però veig que vénen gent de molts de pobles d'aquí al voltant eh</i> ” (DC) -RT
4 Caiguda del clíctic	1. \emptyset	“ <i>Jo penso que ho hi ha massa problemes de convivència entre lo que és comunitat catalana-castellana</i> ” (VC) -RT
	2. clíctic	“ <i>... al poble pràcticament hi vaig a dormir.</i> ” (VC) -RT
5 Dislocació a la dreta	1. Accent	
	2. dislocació a la dreta	
6 Dislocació	1. dislocació a l'esquerra	“ <i>bandera, home, sí que la tenen, no?</i> ” (VC) -RT

Tabla 4: Variables para el módulo morfosintáctico del catalán

VARIABLE	VARIANTES	EJEMPLOS
	2. dislocació a la dreta	“Últimament n’hem fet moltes de coses eh.” (DC)] -RT
7	Morfologia de clítics	1. me, te, ne “ me sembla a mi” (DC)
		2. em, et, en “Aquestes, no, em sembla que no” (VC)
8	Relatiu de lloc	1. que “vam anar l’altre dia aquella botiga que tenen animals” (DC)
		2. on “cap al Sector Nord en concret que és allà on es feia aquesta festa...” (VC)
9	Article masculí	1. lo, lu “ni que ho reguis no és lo mateix” (VC).
		2. el “ el nostre fill per exemple va estar a la Canonja” (VC)
10	Caiguda de preposició	1. \emptyset “Però fins i tot jo veig companys del meu fill \emptyset que [qui] ell es relaciona” (VC)
		2. preposició
11	Alternança d’adverbi	1. en aquí, en allà “I durant el matí fins a les tres i llavors en allà en veiem” (VC)
		2. aquí, allà I ara canbia , ara es posarà allà , es veu que ho farà més gran (DC)
12	Tria de preposició	1. amb “Sí i el vai trucar corrents amb ell” (DC)
		2. a, en “Naltros tenim gent de la que veus en fotos d’aquestes” (VC)
13	“dequeisme” [sense espai]	1. de que “contactes que he tingut amb altra gent pos parlen de que hi ha molta gent” (VC)
		2. que “Sí doncs per allà, i diu que ell és especialista (DC)
14	Alternança de funció de pronom	1. dative “que jo em sembla de que” (VC)
		2. subjective “ a mi m’és igual” (DC)
15	Nosaltres-Naltros	1. naltros 2. nosaltres
16	Venir-Vindre	1. venir 2. vindre
17	Voler-Volguer	1. voler 2. volguer

Tabla 5: Variables para el módulo morfosintáctico del español

VARIABLE	VARIANTES	EJEMPLOS
1	Uso de las dos formas del pretérito imperfecto de subjuntivo	1. -ra <i>a. “La suerte había hecho que el pobre bizzo volara en mi lugar;...” (RM)</i>
		2. -se <i>b. “...esa idea hizo que Julia sonriese para sus adentros.” (AP)</i>

Tabla 5: Variables para el módulo morfosintáctico del español

VARIABLE	VARIANTES	EJEMPLOS
2 Posición de los pronombres dativo y acusativo	1. Pospuesta al verbo	<i>a. "... supongo que necesitaba contárselo a alguien."</i> (RM)
	2. Antepuesta	<i>b. "Yo se lo voy a decir."</i> (AP)
3 Modos de expresión de acciones futuras descritas en presente y pasado	1. forma verbal de futuro	<i>a. "Durruti pensó que un niño como yo no llamaría la atención."</i> (RM)
	2. perífrasis verbal de infinitivo (ir + a+ inf.)	<i>b. "Julia estaba muy lejos de imaginar hasta qué punto ese gesto iba a cambiar su vida."</i> (AP)
4 Pronombre relativo "que" vs. "quien" referente a persona, con preposición, en adjetiva especificativa	1. quien	<i>a. "Me sentía como el ciego a quien un día cambian los muebles de lugar sin advertírsele..."</i> (RM)
	2. que	<i>b. "... contra el jugador al que se enfrentaba hace un rato allá arriba."</i> (AP)
5 Pronombre relativo "que" vs. "cual", con preposición, en adjetiva	1. cual	<i>a. "... una chabola que nos habían prestado y de la cual apenas si se nos permitía salir ..."</i> (RM)
	2. que	<i>b. "al cabo de un rato, durante el que nadie dijo una palabra..."</i> (AP)
6 Uso de pronombres indefinidos	1. ningún(o/a) - pronombre indefinido en función de adjetivo	<i>a. "No hay ninguna sospecha..."</i> (RM)
	2. algún(o/a) - pronombre indefinido postpuesto al nombre y en función de adjetivo	<i>b. "no había duda alguna..."</i> (AP)
7 Posición del YA	1. Postpuesta	<i>a. "en realidad, conocía ya el contenido del sobre."</i> (AP)
	2. Antepuesta	<i>b. "Ya no me cruzaba con él en el cuarto de baño por las noches..."</i> (RM)
8 El pronombre "Yo" en función de sujeto	1. expresión del pronombre	<i>a. "Yo creo que deberías ir."</i>
	2. omisión del pronombre	<i>b. "ØQuiero decir ganas de jugar..."</i>
9 Cuando/Al	1. Cuando + finito	<i>a. "Era uno de esos tipos que, cuando entran en un cuarto, impregnan de inmediato el aire de tensión."</i> (RM)
	2. Al + infinitivo	<i>b. "sonrió al pensar en César."</i> (AP)
10 Uso de adverbio relativos	1. Adverbio relativo (donde, como, cuando)	<i>a. "... Conozco una cueva cercana donde podemos guarecernos."</i> (RM)
	2. en (el) que	<i>b. "... un espacio en el que ella misma había llegado a sentir vértigo..."</i> (AP)

Tabla 6: Variables para el módulo morfosintáctico del inglés

VARIABLE	VARIANTES	EJEMPLOS
1 Subordinate defining RC non-	1. non-defining RC	<i>I was born in Redhill, which is in Surrey (J)</i>
	2. new sentence	<i>Every day I had to take the train, I still remember the train (J)</i>
2 Position of thematic Adjunct	1. Initial	<i>That year I had a good year (J)</i>
	2. Final	<i>I had a good year that year (J)</i>
3 Adjunct duplication	1. No duplication	<i>I didn't know any Catalan then (J) and after 8 o'clock then (J)</i>
	2. Continuous duplication	<i>now, we are all very good friends, after the years (N)</i>
4 Negative preposing	1. canonical	<i>They never rarely speak (J)</i>
	2. preposing	<i>Never (do they) rarely speak (J)</i>
5 Argument preposing preposing	1. canonical	<i>I wasn't blamed for something I didn't do (N)</i>
	2. preposing	<i>For something I didn't do, I wasn't blamed (N)</i>
6 Pronoun/full subject dropping in coordinate clauses (and, but, or)	1. dropping (\emptyset)	<i>We always travel and (\emptyset) get away from the coast (N)</i>
	2. non-dropping	<i>I got a scholarship and I went to school in Croydon (J)</i>
7 Relative pronoun use/dropping in defining RC (object):	1. dropping (\emptyset)	<i>the sort of person (\emptyset) I would be friends with (J)</i>
	2. that	<i>(blamed) for something that [Vb] I didn't do (N)</i>
8 that dropping in that-clauses	1. dropping (\emptyset)	<i>I thought (\emptyset) it was, it was really interesting (J)</i>
	2. non-dropping	<i>I thought it was better that she should learn at least one (N)</i>
9 Emphasis marking	1. auxiliary	<i>But I do have a lot of friends (N)</i>
	2. adverb	<i>which is a town very near Reus (J)</i>
10 Use of intervening Inf-clause	1. Inf-clause	<i>I do consider the teachers to be my friends (J)</i>
	2. \emptyset	<i>I consider myself (\emptyset) very lucky (N)</i>
11 Anaphora	1. they	<i>Somebody in English they have to be (J)</i>
	2. he/she	<i>Somebody in English he have to be (J)</i>

Tabla 7: Variables para el módulo pragmático-discursivo del catalán

VARIABLE	VARIANTES	EJEMPLOS
1 Resposta a pregunta oberta amb marcador discursiu	1. Presència de marcador discursiu	<i>E -Del 89 fins ara? — A -A veure , les, una entitat històrica com el Casino</i>
	2. Una altra estratègia i/o \emptyset	<i>E-I els Paral·lels? — B- No no</i>

Tabla 7: Variables para el módulo pragmático-discursivo del catalán

VARIABLE	VARIANTES	EJEMPLOS
2 Resposta a pregunta oberta amb repetició de la pregunta	1. Repetició	<i>E-O sigui que hi ha hagut una única festa al poble — A- Hi ha hagut una única festa al poble...passa que</i>
	2. Una altra estratègia i/o \emptyset	<i>E-La reforma? — B- \emptysetEl que no es vegi</i>
3 Resposta a pregunta oberta amb exclamació	1. Exclamació	<i>E-I la Maria què diu? — A- Ui! La Maria</i>
	2. Una altra estratègia i/o \emptyset	<i>E-sembla que el poble ha anat millor? — B- \emptysetamb aspect de fet doncs</i>
4 Posició del marcador discursiu pues.	1. Inicial en resposta a pregunta oberta	<i>E-la teva filla on va estudiar? — A- pues va anar a Barcelona</i>
	2. Una altra: intermèdia o final	<i>E-Tú d'on ets? — B- de Tortosa</i>
5 Ús de eh? i no?	1. Eh?	<i>A- de fet perquè sóc bastant perucot eh?, però vull dir no se m'hauria ocorregut mai en general</i>
	2. No?	<i>B-de fet pràcticament això ja començava a bullir no? però jo recordo les coses més típiques</i>
6 Posició discursiva eh?	1. Final de torn o absoluta (funció apel·lativa, espera resposta)	<i>A-Què vols dir amb això,eh?</i>
	2. Intermèdia (final d'acte) (funció fàtica, no espera resposta)	<i>B- jo aquest poble no l'havia sentit ni anomenar eh?, per començar i bueno vai vindre un any</i>
7 Posició discursiva no?	1. Final de torn o absoluta (funció apel·lativa, espera resposta)	<i>A-Vindreu demà,no?</i>
	2. Intermèdia (final d'acte) (funció fàtica, no espera resposta)	<i>B- jo penso en l'altre extrem, no? saps i doncs jo no sé si les opinions que jo te doni són massa d'allò</i>
8 Intensificació morfològica vs. Lèxica de l'enunciat	1. Ús de prefixos i quantificadors	<i>A-la meva filla està súper feliç en aquest nou col·legi.</i>
	2. Ús de la repetició	<i>B- ja està bé, ja està bé.</i>
9 Intensificació del subjecte de primera persona singular	1. 1a persona singular JO	<i>A-Jo era sempre la que ho feia tot. .</i>
	2. \emptyset	<i>B- Cada any \emptysetvinc a les festes</i>
10 Atenuació del subjecte per impersonalització del JO	1. JO+ \emptyset	<i>A-Jo sempre li dic a la nena que no faci cas.</i>
	2. Es, un/una, 2a p. sing	<i>B-Perquè clar, tú vens i t'agrada, i no saps que dir</i>
11 Dislocació a la dreta	1. Accent	<i>A-jo n'he sentit a parlar del tema perquè ho havien venut o no sé què</i>

Tabla 7: Variables para el módulo pragmático-discursivo del catalán

VARIABLE	VARIANTES	EJEMPLOS
	2. dislocació a la dreta	<i>B-hi ha pobles que m'atrau algo</i>
12 Dequeisme	1. de que	<i>A-també influeix de que vinguin o no ells</i>
	2. øque	<i>B- Llavons és una manera que molta gent agafi l'autobús</i>
13 Canvi funció sintàctica	1. datiu	<i>A-Me sembla que vindrà molta gent aquest any.</i>
	2. subjecte	<i>B-Penso que la veïna era de Portobou.</i>
14 Estil comunicatiu	1. control monotrització	<i>A-No sé, eh! Es que saps? cada vegada que et trobes un</i>
	2. confirmació	<i>B-Els nens ara van més contents als, oi que sí?</i>

Tabla 8: Variables para el módulo pragmático-discursivo del español

VARIABLE	VARIANTES	EJEMPLOS
1 Respuesta a pregunta abierta con presencia de marcador discursivo	1. Presencia de marcador discursivo	<i>E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: pues por mediación de trabajo</i>
	2. Otra estrategia y/o ø	<i>E: y ¿cuáles son las fiestas de Aldaya? — J: las fiestas de Aldaya pues hay mm/ mm celebran muu- casi todos</i>
2 Respuesta a pregunta abierta con repetición de pregunta	1. Repetición de pregunta	<i>E: ¿Y tu barrio? — J: ¿mi barrio? ¡uff! Lo mejor</i>
	2. Otra estrategia y/o ø	<i>E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: pues por mediación de trabajo</i>
3 Respuesta a pregunta abierta con exclamación	1. Exclamación	<i>E: imagínate que recibes una herencia inesperada — J: ¡uy! También haría muchas cosas/ pero// lo primero/ no sé...</i>
	2. Otra estrategia y/o ø	<i>E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: pues por mediación de trabajo</i>
4 Respuesta simple o compuesta	1. Simple	<i>E: bien/ hálbanos de Cullera — V: ¡uy! Cullera/ puees/ demasia(d)o ajetreio/ a mí me gusta más ..</i>
	2. Compuesta	<i>E: dínos, por ejemplo, ¿qué has hecho hoy? — V: ¿qué he hecho hoy? // pues/ he ido y he llevado a mi perra por ejemplo aa- el veterinario</i>
5 Posición del marcador discursivo pues.	1. Inicial en respuesta a pregunta abierta	<i>E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: pues por mediación de trabajo</i>
	2. Otra: Intermedia o final	<i>E: ¿Te da tranquilidad interior? — V: Sí, sinceramente sí, porque ahora tengo un poco de preocupación, y entonces pues he ido a pedirle/ pues a ver si me ayuda ...</i>

Tabla 8: Variables para el módulo pragmático-discursivo del español

VARIABLE	VARIANTES	EJEMPLOS
6 Marcadores discursivos ordenadores de la información	1. Simple: luego (y luego)	<i>V: ¿qué he hecho hoy? // pues/ he ido y he llevado a mi perra por ejemplo aa- el veterinario/ que no estaba muy bien// después me he marchado a comprar/ luego hemos comido// y luego me he ido...</i>
	2. Compuesta: luego (y luego) pues	<i>J: digo pues yo que sé (risas)/ lue- yy naada/ y luego pues las Fallas se celebran muuchoo/ yy- y las fiestas de Aldaya</i>
7 Uso de ¿eh? Vs. ¿no?	1. Eh?	<i>J. pues no sé cómo la organizaría ¿eh?/ supongo quee sacaría casi todo lo que tengo en el comedor (risas)/// pondría unos tableroos yy- yy- y haría pues no sé algo sencillo ¿eh?/ algo sencillo pa- para noo formar/ mucho follón/ y invitaríamos a la familia</i>
	2. No?	<i>J: luego yaa- ya empezaron aa- a po- ya pusieron/ la LUUZ pero me acuerdo de alumbarnos con VEElas/ ir de vacaciones a- de niña/ a eso ¿no?/ a padecer más que a- que a pasártelo bien</i>
8 Posició discursiva ¿eh?	1. Final de turno o absoluta (función apelativa, espera respuesta)	<i>V: y las fiestas pues nada en casa// preparo- hacemos pocas fiestas en casa ¿no?// porque casi siempre nos vamos fuera/ porque la casaa hay quee ... es- estar poquito ¿eh?</i>
	2. Intermedia (final de acto) (función fática, no espera respuesta)	<i>V: yo era muy tranquila/ yo ahora soy PURO nervio// me expreso/ me- no me callo/ y antes mm- era diferente// igual me lo ha pega(d)o mi marido ¿eh?/ que mi marido no se calla</i>
9 Posició discursiva ¿no?	1. Final de turno o absoluta (función apelativa, espera respuesta)	<i>V: yo pienso que sí/ ee por ejemplo vi un documental en la televisión el otro día/ que hablaban dee/ (chasquido) de- del Antártico/ o sea dee// de Groenlandia y todo esto/ ¿no? — E: sí del Ártico</i>
	2. Intermedia (final de acto) (función fática, no espera respuesta)	<i>J: y entonces me tenían como una pava/ como la tonta de la clase ¿no?/ pero fue/ pasar a segunda etapa/ y empezar yo a ponerme ya- así yaa- a ser ya una mujercilla y</i>
10 Intensificación morfológica vs. Léxica del enunciado	1. Uso de prefijos y cuantificadores	<i>J: es una niña que no tiene problemas/ súper abierta</i>
	2. Uso de la repetición	<i>J: yo veo un niño maleducado y no lo soporto/ no lo soporto</i>
11 Intensificación del sujeto de primera persona singular	1. 1a persona singular YO	<i>J: yo veo un niño maleducado y no lo soporto/ no lo soporto</i>
	2. ∅	<i>E: ¿cómo organizarías una fiesta familiar en casa? — V: pues mira/ muy fácil/ encargo unos bocadillos...</i>
12 Atenuación del sujeto por impersonalización del YO	1. YO+ ∅	<i>V: me gustaría muchísimo ir al cine/ porque me encanta el cine</i>
	2. Se,uno/una, 2ª p. sing	

Tabla 9: Variables para el módulo pragmático-discursivo del inglés

VARIABLE	VARIANTES	EJEMPLOS
1 Presence of discursive empty marker ‘Well’, ‘actually’	1. Presence 2. Another strategy or \emptyset	<i>Well I start with an ‘uu’ sound and [...]</i>
2 Position of empty marker	1. Initial after Open Q. 2. another: Intermediate / Final	<i>Well because the period that “The remains of the day” takes place in was absolutely during the pre-war and it covers in many ways the rise of, well it covers de rise of the Nazis obviously</i>
3 Position of modesty marker	1. Initial another: Intermediate/Final	<i>I think when you actually start reading [...] at the same time I think he may have regretted saying it because [...]</i>
4 Composed marking	1. Presence of several markers in sentence 2. Presence of a unique marker	<i>Oh yeah, I mean I think there are enormous differences. Well actually my dad did it first [...]</i>
5 Question/topic repetition	1. Repetition 2. another strategy or \emptyset	
6 Subject repetition with pronoun	1. subject + pronoun 2. Subject	<i>my father... he was an architect’ My father</i>
7 Reformulations / interpolated clause	1. with repetition (words or any information: morphological, etc.) 2. without repetition	<i>And went on... Andrea and I went on to this... into this box ring and sang the song, and it was an overnight success. [...] that I’ve looked at in my... so, all about.. spirituality [...].</i>
8 Excourses	1. Excourse given in paragraph 2. No excourse in paragraph	<i>Olivier actually always had... that’s why Ken always found the comparison with him and Olivier so laughable, because Olivier had this extraordinarily patrician glamour.</i>
9 Adverb repetition for intensity purposes	1. repetition 2. no repetition	<i>(Very, really, more, many...)</i>
10 Subject repetition in coordinate sentences for emphasis	1. Coordination with subject repetition 2. Coordination without subject repetition	<i>we had a flute, we had a piano...</i>
11 ANY repetition (words or any information: morphological, etc.)	1. Sentence with repetition 2. Sentence without repetition	<i>Well, no, not really, to be honest, not really [...]</i> [HL]

Tabla 9: Variables para el módulo pragmático-discursivo del inglés

VARIABLE	VARIANTES	EJEMPLOS
12 Ordinal use	1. double 2. simple	<i>(First, second... last, next, another, other...)</i>
13 Presence of intensifier 'really' vs. others	1. sentence with 'really' 2. sentence with other intensifier ('completely', 'absolutely') or \emptyset	<i>It's the opposite thing, really, I have this thing called the curse of Costello.</i>
14 Use of impersonalisation	1. Paragraph with impersonal 'you' 2. Paragraph without impersonal 'you'	<i>There's various forms of lying, you can lie at once at the same time [...]</i>