



Idiolectometria Forense i Índex de Similitud Idiolectal

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

Actualitzat: gener 2013

1. Presentació

Aquesta recerca s'ha portat a terme amb l'ajut dels finançaments següents:

Fase I

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

(EXPLORA-INGENIO HUM2007-29140-E)

PI: Dr. M. Teresa Turell

Entitat: Ministerio de Educación y Cultura

Període: 2007-2008

Fase II

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

PI: Dr. M. Teresa Turell

Entitat: Ministerio de Ciencia e Innovación

Període: 2008-2012

Fase III

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

PI: Dr. M. Teresa Turell

Entitat: Ministerio de Economía y Competitividad

Període: 2012-2015

1.1 Finalitat

Aquest projecte es proposa estudiar l'estil idiolectal dels parlants i escriptors d'una determinada llengua des d'una perspectiva aplicada a la **lingüística forense** (disciplina que considera la interfície entre llengua i dret), en especial a la **fonètica forense**, per a la identificació de parlants i l'establiment de perfils lingüístics, d'una banda, i a la **determinació/atribució d' autoria** de textos escrits, d'una altra. L'estil idiolectal es pot definir com l'elecció individual per part de cada parlant de determinades formes lingüístiques (fonètiques, fonològiques, morfològiques, sintàctiques, pragmàtiques, discursives) que li ofereix la llengua (Nolan 1994: 331). Per tant, l'estil idiolectal és individual i únic.

La **idiolectometria** té com a objecte d'estudi l'**estil idiolectal**. Fins ara, aquesta disciplina s'havia ocupat d'analitzar la distància lingüística entre parlants/escriptors i establir fronteres entre idiolectes diferents, fronteres que per definició separen a cada persona de la resta de parlants/escriptors. En canvi, aquest projecte es proposa explorar i desenvolupar la possibilitat de mesurar la diferència lingüística entre diversos idiolectes i la distància idiolectal de cada individu, de manera que al final es pugui aconseguir un **Índex de Similitud Idiolectal (ISI)**, una

mesura numèrica que compari diverses mostres lingüístiques i calculi la distància lingüística entre elles. Més específicament, es tracta de poder establir el nivell de similitud idiolectal a partir del qual es pot considerar que dues mostres lingüístiques (orals o escrites) han estat produïdes per una mateixa persona. L'aplicació de l'estudi de l'estil idiolectal al camp de la lingüística forense és fonamental, ja que aquesta aplicació hauria de permetre identificar de manera inequívoca als parlants o escriptors d'una determinada llengua, comparant una gravació o un text escrit **dubitat** (l'autoria del qual presenta dubtes) i un conjunt de gravacions o textos escrits **indubitats** (l'autoria dels quals és coneguda), a partir de l'anàlisi de les formes lingüístiques triades per part de cada parlant/escriptor. Així doncs, en aquest projecte es proposa la creació d'un protocol que permeti establir aquest **Índex de Similitud Idiolectal (ISI)** i, un cop analitzats diversos paràmetres lingüístics (fonològics, morfosintàctics, discursiu-pragmàtics), concloure si dues gravacions o dos textos escrits han estat produïts per una mateixa persona o no.

En aquest projecte a) s'aplicarà la tècnica a quatre llengües (espanyol, català, anglès i àrab), b) s'avaluarà el disseny global del protocol, i c) es dissenyaran les proves estadístiques per a totes les llengües.

1.2 Antecedents i estat actual dels coneixements científico-tècnics

El marc conceptual del tema que es pretén explorar en aquest projecte està basat en tres disciplines lingüístiques: la **lingüística forense** (en particular, la **fonètica forense**, per a la identificació de parlants i l'establiment de perfils lingüístics, d'una banda, i la **determinació/atribució d'autoria** de textos escrits, de l'altra); la **idiolectrometria**, i finalment, la **teoria de la variació i el canvi lingüístic**, i més específicament, la **variació sociolingüística**. Per tal de poder situar conceptualment i metodològicament aquest projecte, de forma molt general es defineix la **lingüística forense** com la interfície entre llengua i dret (per a una definició més detallada, consulteu: <http://www.iafl.org>). Des del punt de vista metodològic cal comentar que en la majoria de casos la lingüística forense utilitza tècniques derivades de les noves tecnologies, és a dir, que el desenvolupament de la recerca i de servei públic en lingüística forense es realitza habitualment a partir de tota una sèrie d'eines informàtiques i estadístiques que possibiliten l'assoliment d'opinions, actuacions i assessorament científic i professional d'alta competència i fiabilitat per part dels lingüistes forenses, per al seu posterior ús des de diferents institucions i organismes públics (escola judicial, ministeri fiscal, policia) i privats (empreses), i per part de professionals diversos (jutges, advocats defensors, psicòlegs, metges). En qualsevol cas, les proves pericials que es presenten no són en general proves definitives ni concloents, sinó més aviat complementàries d'altres tipus d'evidència, o bé permeten introduir un dubte raonable.

La **idiolectometria**, disciplina lingüística que es troba en una fase inicial de desenvolupament, té com a objecte d'estudi l'**estil idiolectal** dels parlants i escriptors d'una determinada llengua. Com ja s'ha esmentat, l'estil idiolectal es pot definir com l'elecció, entre les diferents opcions que ofereix la llengua, que els parlants i escriptors fan d'algunes formes lingüístiques (fonètiques, fonològiques, morfològiques, sintàctiques) determinades. Per tant, podem

entendre que una llengua està formada per la suma de dialectes i sociolectes, i aquests alhora es defineixen com la suma d'idiolectes (o estils idiolectals de la llengua). Així doncs, els factors socials (grup socioeconòmic, grup d'edat, nivell educatiu, gènere, professió) es manifesten en la configuració dels idiolectes. La figura 1 presenta una representació esquemàtica del que s'acaba de comentar:

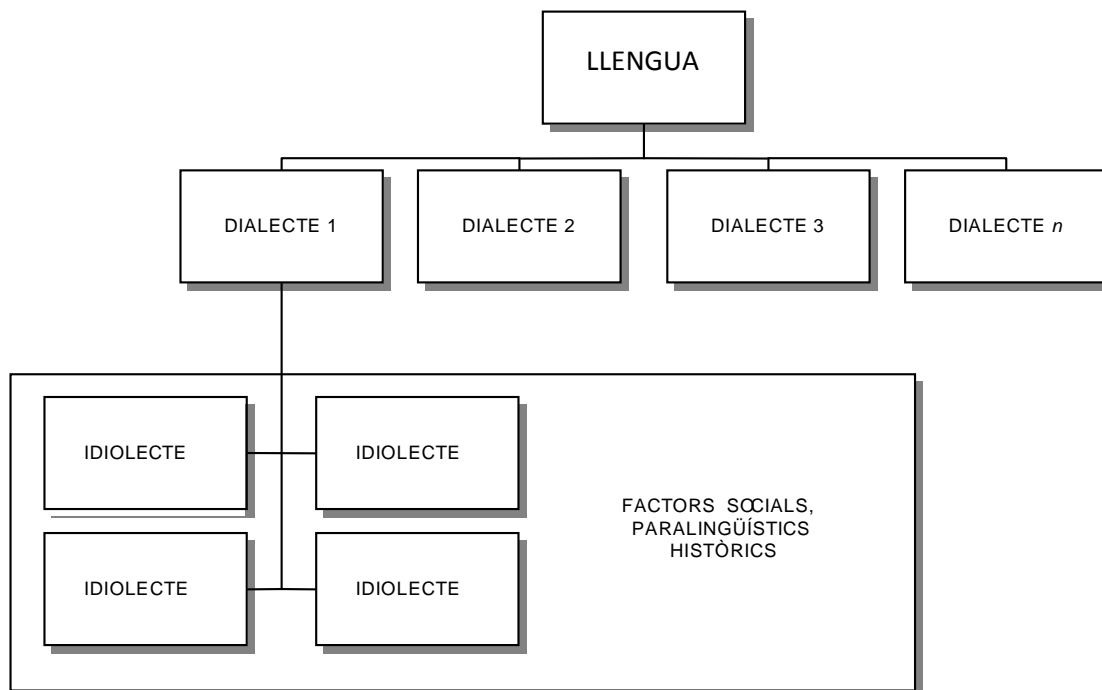


Figura 1. Esquema de l'estructura de les llengües, formades per dialectes, i aquests formats pel conjunt d'idiolectes. Els idiolectes estan condicionats per factors socials, paralingüístics i històrics.

Els estudis que analitzen la influència dels factors socials en les realitzacions lingüístiques dels parlants i escriptors es centren en la **teoria de la variació i el canvi lingüístics** (Labov 1994, 2001; Turell 1995), que postula l'existència de variació lingüística inherent a totes les llengües i que afecta a tots els nivells lingüístics: fonètica i fonologia, morfologia, sintaxi, semàntica, pragmàtica i discurs. La variació sociolingüística estudia la forma en què s'estructura aquesta variació, és a dir que considera aquells factors interns (lingüístics) i externs (socials) que intervenen en la seva estructuració. Diversos estudis han demostrat la influència de l'estil (Schilling-Estes 2002), del gènere, l'edat, l'etnicitat i la classe socioeconòmica (Labov 1966 [2006], 1994, 2001), o la dels grups o xarxes socials (Milroy 1987, Eckert 2000) en les produccions lingüístiques, tant a nivell fonètic i fonològic, com morfològic, sintàctic i semàntic. En aquest sentit, la sociolingüística variacionista estudia l'estructura de la variació des de diversos nivells i diverses perspectives, tal com queda resumit en el diagrama següent:

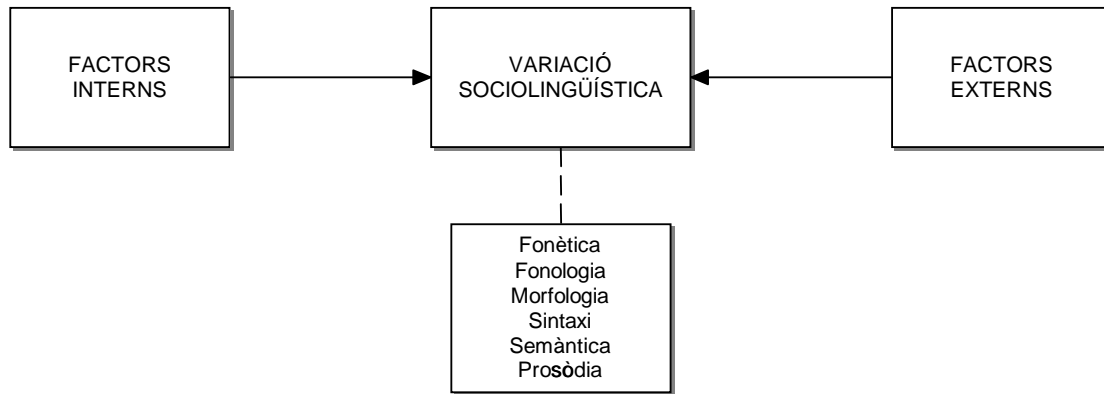


Figura 2. Esquema dels factors que intervenen en l'estructuració de la variació lingüística.

No obstant, la variació sociolingüística no s'ha centrat de manera prominent en l'estudi de l'idiolecte, exceptuant alguns estudis més aviat generalistes com el de Guy (1980), Coulthard (2004), Abercrombie (1969) o Biber (1988, 1995). D'aquí, doncs, l'interès que pot suscitar aquest projecte, que té com a objecte d'estudi bàsic l'estil idiolectal en diverses llengües.

1.3 Membres y mòduls analitzats

- **Català:**
 - Mòdul fonològic del català: Dr. Jordi Cicres – UPF i UdG.
 - Mòdul fonològic de l'espanyol: Dra. Fernanda López – UPF i UNAM.
 - Mòdul fonològic de l'anglès: Núria Gavalda – UPF.
- **Espanyol:**
 - Mòdul morfosintàctic del català: Dra. Montse Forcadell – UPF i UB.
 - Mòdul morfosintàctic de l'espanyol: Dra. Maria Spassova – UPF i NBU.
 - Mòdul morfosintàctic de l'anglès: Dra. M. Teresa Turell – UPF.
- **Anglès:**
 - Mòdul discursiu-pragmàtic del català: Sheila Queralt – UPF, des de 2010.
 - Mòdul discursiu-pragmàtic de l'espanyol: Dra. Raquel Casesnoves – UPF, fins al 2010.
 - Mòdul discursiu-pragmàtic de l'anglès: David García Barrero – UPF, des de 2011.

2. Objectius i hipòtesis

2.1. *Originalitat i interès*

Aquest projecte, que podem emmarcar dins de l'àmbit de la lingüística aplicada, és clarament un projecte 'problem-based'. És a dir que, a partir de la detecció d'un problema que existeix en la nostra societat —la necessitat d'identificar parlants o escriptors, i determinar/atribuir l'autoria de textos escrits amb finalitats forenses— s'empren una sèrie de coneixements lingüístics, a tots els nivells (fonètic, fonològic, morfològic, sintàctic, semàntic, pragmàtic i discursiu) per tal de solucionar-lo. Així doncs, els resultats d'aquest projecte podran ser usats en casos forenses reals en què la Justícia necessiti la perícia de lingüistes per:

- identificar parlants,
- construir perfils lingüístics a partir de fragments de veu o text,
- determinar l'autoria de textos escrits (per exemple, notes de suïcidi o amenaces)

Els coneixements actuals en lingüística forense permeten realitzar les tasques que s'acaben d'esmentar amb certes limitacions. No obstant, les conclusions d'un dictamen pericial forense sempre han de ser necessàriament conservadores per tal d'evitar la possibilitat que es produeixin falses identificacions o atribucions. Actualment, no existeix cap teoria lingüística que tingui en compte tots els models de la gramàtica amb finalitats forenses, tot i que existeixen alguns models específics per a les gravacions, que no entren a valorar els aspectes lingüístics sinó els purament físics de la veu, i altres models específics per a textos escrits, en els quals està basat per exemple el programa CopyCatch, que ajuden a l'expert lingüista a detectar el plagiat, o inclús d'altres que analitzen l'estil de diversos textos i poden ser emprats com a prova complementària en l'anàlisi forense. L'aplicació i protocol que es dissenyaran en el marc d'aquest projecte permetran obtenir un grau de seguretat més elevat que l'actual en els dictàmens pericials, i aquest fet revertirà en una millor administració de la Justícia.

2.2. *Hipòtesis*

1. Es postula l'existència de l'idiolecte, o estil idiolectal, com aquell estil d'un parlant o escriptor que és únic i irrepetible.
2. Es postula l'existència de més variació lingüística entre parlants/escriptors (*inter-speaker/writer variation*) que dins del discurs d'un mateix parlant/escriptor (*intra-speaker/writer variation*).
3. Es postula que aquest estílo idiolectal, sobretot en relació a les pautes fonològiques i sintàctiques, no varia massa en funció del gènere textual o context lingüístic. Una altra qüestió és el vocabulari que sí pot tenir implicacions distributives relacionades amb el registre i el gènere textual.
4. Es postula que l'estil idiolectal d'un individu es manté relativament estable al llarg del temps.
5. A partir de la confirmació de les hipòtesis anteriors, es postula la possibilitat d'establir un Índex de Similitud Idiolectal (ISI) que permeti identificar parlants i determinar/atribuir l'autoria de textos escrits amb una major fiabilitat.

2.3. *Objectius generals*

1. Demostrar que tot parlant i escriptor té el seu **estil idiolectal**, és a dir, un comportament lingüístic diferenciat i únic, l'ús del qual és inconscient i que en certa mesura es manté estable al llarg del temps.
2. Aplicar l'estudi de l'estil idiolectal al camp de la **Lingüística forense**, ja que aquesta aplicació pot ajudar a identificar de forma més fiable als productors d'un text oral i/o al escriptors d'un text escrit.

2.4. *Objectius específics (teòrics i metodològics)*

1. Portar a terme aquesta aplicació a partir de la comparació entre les formes/paràmetres lingüístics usats pels parlants i/o escriptors en una gravació o text escrit, per a continuació poder aplicar aquest mateix protocol a textos forenses reals, orals i escrits, i tant **dubitats** (l'autoria dels quals presenta dubtes) com **indubitats** (l'autoria dels quals és coneguda), i així confirmar l'existència de major variació idiolectal inter-parlant/escriptor i menor variació idiolectal intra-parlant/escriptor.
2. Portar a terme aquesta aplicació a partir d'aquesta mateixa comparació, plantejada en dos temps de medició —**temps aparent** i **temps real**— per tal de confirmar també que l'estil idiolectal es manté estable al llarg del temps i amb mostres de diversos gèneres textuais, i en aquest sentit poder confirmar que l'estil idiolectal d'un mateix individu varia poc en mostres de diversos gèneres textuais.
3. Mesurar la diferència lingüística entre diversos estils idiolectals i la distància idiolectal de cada individu, de manera que al final es pugui aconseguir un **Índex de Similitud Idiolectal (ISI)**. Més específicament, es tracta de poder establir el nivell de similitud idiolectal a partir del qual es pot considerar que dues mostres lingüístiques (orals o escrites) han estat produïdes o no per una mateixa persona.
4. Crear un **protocol** que permeti establir aquest **Índex de Similitud Idiolectal (ISI)**, que compari diverses mostres lingüístiques (de diferent longitud i gènere textual) i en calculi la distància lingüística, i poder concloure, un cop han estat analitzats diversos paràmetres lingüístics, si dues gravacions o dos textos escrits han estat produïts per una mateixa persona o no.

3. Disseny experimental

En el marc d'aquest projecte les activitats a desenvolupar, derivades dels objectius exposats més amunt, són les següents:

Activitats d'avaluació (a partir dels resultats del projecte EXPLORA-INGENIO, HUM2007-29140-E)

- S'avaluarà el disseny global del protocol.
- S'avaluarà el mòdul fonològic per a l'espanyol i el català.

Activitats de nova execució

- Es dissenyarà la tècnica per aplicar-la a les tres llengües d'anàlisi (català, espanyol i anglès).
- S'elaborarà el disseny del protocol per als mòduls morfosintàctic i discursiu-pragmàtic de les tres llengües.
- Es dissenyaran les proves estadístiques que han de permetre el càlcul de l'ISI per als mòduls morfosintàctic i discursiu-pragmàtic de les tres llengües.
- S'avaluaran els tres mòduls per a les tres llengües.

La metodologia que s'aplicarà en aquest projecte consisteix en el disseny d'un protocol per calcular l'ISI (**Índex de similitud idiolectal**) entre dos o més textos orals o escrits. Per tant, en primer lloc és necessari establir els corpus d'anàlisi, les variables d'estudi, i les tècniques d'anàlisi lingüística que es consideraran. En aquest sentit, primer es presenten els corpus que s'empraran en cada llengua i a continuació es fa referència a cadascun dels mòduls a partir de les variables analitzades i la metodologia d'anàlisi usada.

3.1. Corpus

El projecte que es presenta pretén estudiar l'estil idiolectal amb finalitats forenses, tant per identificar parlants com escriptors. Per tant, serà necessari utilitzar corpus orals i escrits (i en aquest últim cas, se seguiran les directrius de Coulthard (1994), Eagleson (1994), Turell (1995) i Johnson (1997)).

Quan es parla de corpus orals és necessari distingir entre corpus de la llengua oral (en anglès, *spoken language corpora*) i corpus orals (Llisterri *et al.* 2005). Els primers són transcripcions que pretenen representar uns aspectes determinats de la llengua oral, mentre que els segons estan constituïts, com a mínim, per senyal sonora. El grau de segmentació i etiquetatge varia segons el propòsit d'ús del corpus. Així, els corpus es poden segmentar per fonemes, difonemes, morfemes, sintagmes, unitats d'entonació, oracions, torns de paraula, etc. D'altra banda, també poden incorporar altres tipus d'informació per representar altres aspectes verbals (to, velocitat de l'elocució, entonació, pauses) o no verbals (canvis de torns de paraula, soroll de fons, gestos).

Per estudiar l'efecte del pas del temps sobre les variables analitzades, tenint en compte el marc aplicat en lingüística forense en què s'enquadra aquest projecte, és necessari usar no només un corpus en un temps de medició (equivalent al *temps aparent* en variació sociolingüística), que consistiria en un corpus amb informació lingüística de diversos informants obtingut en un punt temporal concret, sinó també un corpus relatiu a un segon temps de medició (*temps real*), és a dir, un corpus format per material obtingut a partir dels mateixos informants, amb un interval de temps entre les diferents mostres. La noció de temps aparent i temps real s'ha anat emprant, especialment en el camp de la sociolingüística de la variació (Turell 2003), des dels anys trenta, i sobretot des dels anys cinquanta del segle XX.

3.1.1. Corpus per a l'ISI en català

Per a tots els corpus s'usarà la part en català d'un corpus oral en temps aparent i temps real: el corpus de La Canonja (Pujadas, Pujol Berché i Turell, 1992), recollit pel grup UVAL (Unitat de Variació Lingüística) de l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. Inicialment, aquest corpus estava constituït per 29 gravacions (13 de parlants catalanoparlants i 16 de parlants castellanoparlants) realitzades a mitjans dels anys vuitanta del segle passat. En el procés de realització d'aquest projecte es realitzarà una nova tanda de gravacions, de manera que es podran estudiar els mateixos parlants amb un interval temporal de més de 20 anys.

3.1.2. Corpus per a l'ISI en espanyol

Per a la identificació de parlants s'usarà la part en espanyol d'un corpus oral en temps aparent i temps real: el corpus de La Canonja, recollit pel grup UVAL (Unitat de Variació Lingüística) de l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. Inicialment, aquest corpus estava constituït per 29 gravacions (13 de parlants catalanoparlants i 16 de parlants castellanoparlants) realitzades a mitjans dels anys vuitanta. En el procés de realització d'aquest projecte es portarà a terme una nova tanda de gravacions, de manera que es podran estudiar els mateixos parlants amb un interval temporal de més de 20 anys.

També es podrà comptar amb el corpus en temps aparent i real per a l'espanyol de Mèxic que s'està recollint a partir d'una beca CONACYT, concedida a Fernanda López, membre del projecte, per a la realització de la seva tesi doctoral. Es tracta del corpus DIMEx100, compilat per l'Instituto de Investigaciones en Matemáticas Aplicadas i Sistemas de la Universidad Nacional Autónoma de Mèxic, i que té com a principal objectiu la creació d'un reconeixedor automàtic de parla. El corpus va ser gravat per 100 parlants que van llegir 10 frases iguals per a tots i 50 frases diferents per a cadascú. Els parlants van ser seleccionats segons la seva edat (de 16 a 36 anys), el seu nivell escolar (d'estudis superiors a secundària), i el seu lloc d'origen (Ciudad de México).

Per a l'anàlisi del mòdul morfosintàctic de l'espanyol s'usarà el corpus de textos escrits compilat per Maria Spassova per a la realització de la seva tesi doctoral. Aquest corpus està compost per textos escrits en espanyol fins a 20 autors hispanoparlants nadius, de procedència peninsular i sud-americana, amb dos subcorpus: textos de narrativa literària (N) i textos d'articles d'opinió (AO).

I per al mòdul discursiu-pragmàtic s'emprarà el corpus Preseva (*Proyecto para el Estudio Sociolingüístico del Español de Valencia*, <http://www.uv.es/preseval/ppal.htm>), del qual es consideraran informants de nivell sociocultural mitjà i alt, i es controlarà per gènere biològic i edat (Gómez Molina, J. R. 2001 i 2005).

3.1.3. Corpus per a l'ISI en anglès

Per a la identificació de parlants s'usarà un corpus de 20 parlants (10 britànics i 10 nord-americans dels EE.UU) i altres mostres orals del BNC (British National Corpus). Es preveu també l'ús de gravacions reals d'entrevistes de ràdio i TV a personatges del món cultural i artístic del Regne Unit.

En el cas de la determinació/atribució d'autoria s'usaran textos escrits procedents de diversos corpus existents i altres corpus escrits derivats de casos reals ja sentenciats que seran proporcionats pel grup de recerca de l'Aston Centre for Forensic Linguistics (CFL).

3.2. Variables

Pel que fa a les variables d'estudi, inicialment ens centrarem en les variables fonològiques. Rose (2002) planteja la complexitat de l'anàlisi comparativa de les mostres de veu, i la necessitat d'analitzar diverses variables a partir de diversos gràfics que es presenten a continuació. En primer lloc, aquest fonetista descriu una situació ideal en la qual les realitzacions d'una variable concreta són úniques per a cada parlant, tal com es mostra a la figura 3, a partir de dades inventades per a la variable: mitjana de la F0:

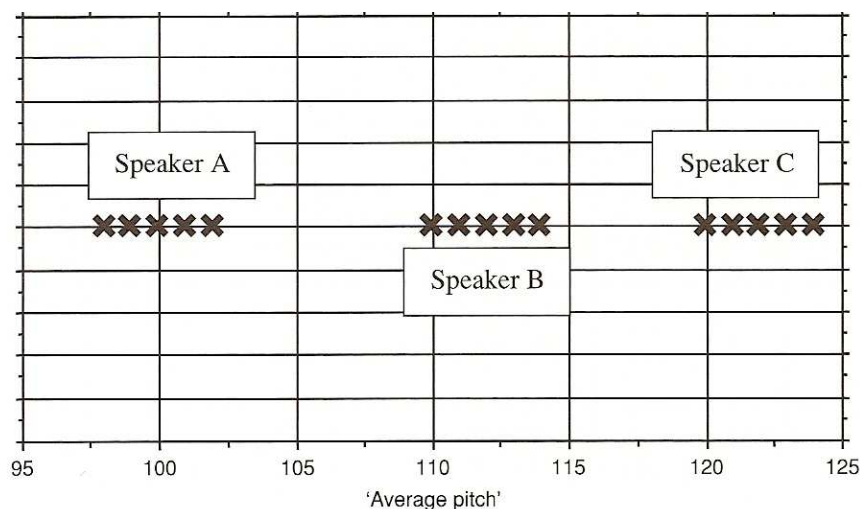


Figura 3. Representació "ideal" dels valors d'una sola variable en tres parlants. Font: Rose (2002).

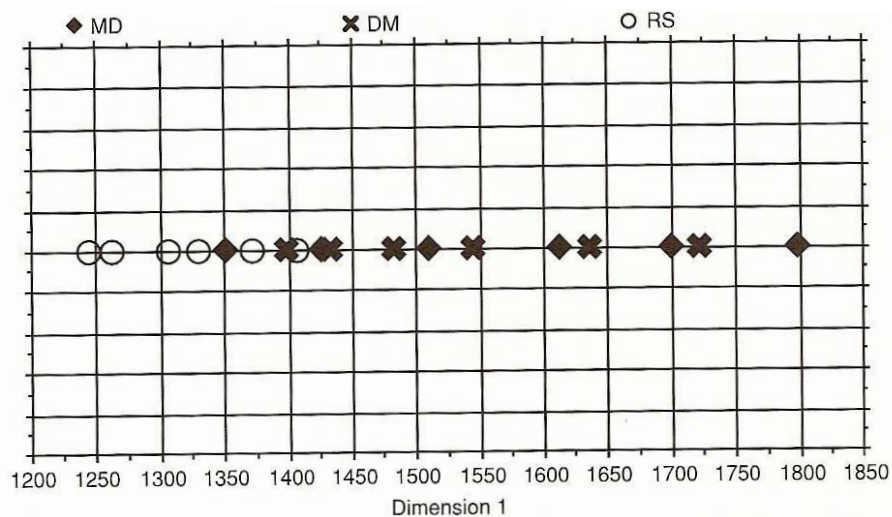


Figura 4. Representació "realista" dels valors d'una sola variable en tres parlants, en la que es produeix per superposició de valors. Font: Rose (2002).

En la figura 4, s'observa com els valors d'una variable concreta (una dimensió, segons terminologia de Rose 2002) no són suficients per discriminar parlants, ja que la superposició de dades no permet arribar a cap conclusió. Per tant, s'han d'introduir noves dimensions en els estudis d'identificació de parlants, per obtenir espais delimitats per a cada parlant. És el que es mostra en les figures 5 (2 dimensions) i 6 (3 dimensions).

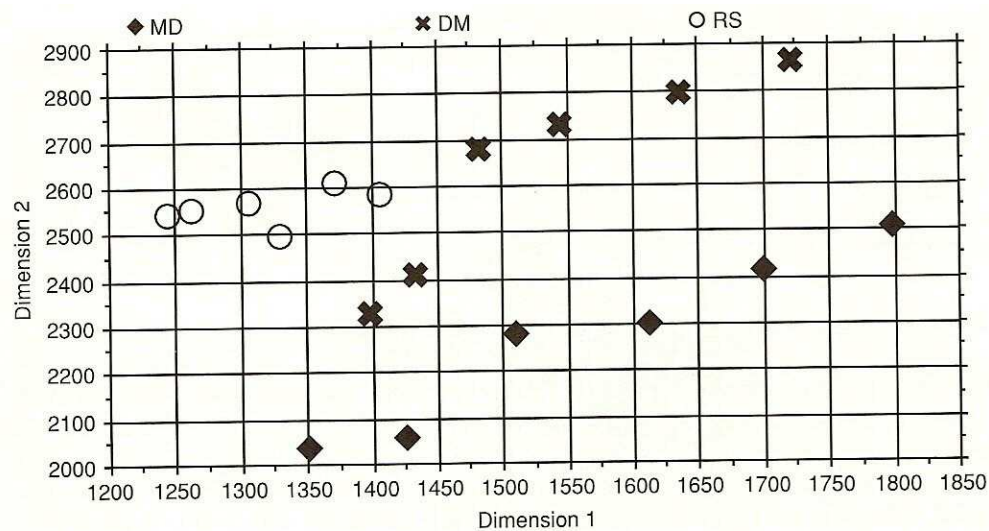


Figura 5. Representació dels valors de dues variables en tres parlants . Font: Rose (2002).

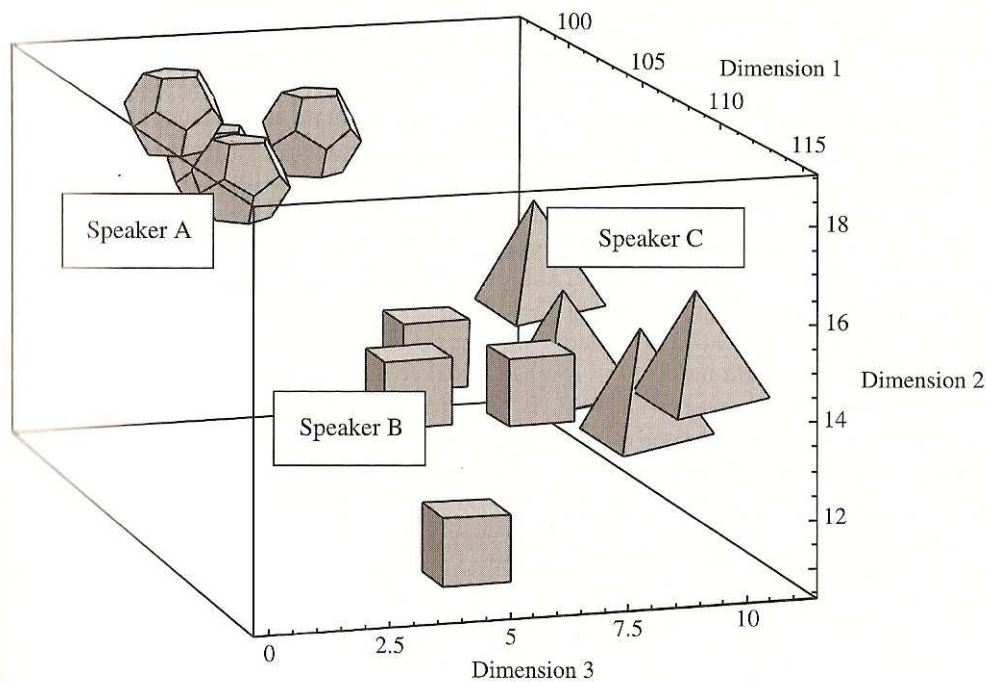


Figura 6. Representació dels valors de tres variables en tres parlants . Font: Rose (2002).

No obstant, tres dimensions tampoc no seran suficients per poder extraure conclusions fiables sobre l'autoria d'una gravació. És imprescindible augmentar el nombre de variables per tal que la discriminació dels parlants sigui més completa.

D'altra banda, Rose (2002: 33-53) classifica l'anàlisi forense de la veu en quatre grans grups: auditiu lingüístic, auditiu no lingüístic, acústic lingüístic i, finalment, acústic no lingüístic:

Taula 1. Classificació primària dels paràmetres en l'anàlisi forense segons Rose (2002).

	<i>LINGÜÍSTIC</i>	<i>NO LINGÜÍSTIC</i>
<i>AUDITIU</i>	AUDITIU LINGÜÍSTIC	AUDITIU NO LINGÜÍSTIC
<i>ACÚSTIC</i>	ACÚSTIC LINGÜÍSTIC	ACÚSTIC NO LINGÜÍSTIC

En aquest context, 'lingüístic' s'usa com a sinònim de 'fonemàtic', és a dir, que té una funció dins del sistema de la llengua. Segons aquesta classificació els paràmetres acústics o auditius no lingüístics correspondrien als paràmetres fonètics (segmentals i suprasegmentals), ja que el fet de pronunciar un determinat al·lòfon amb unes característiques espectrals, temporals i d'intensitat concretes no influeix en el seu contingut lingüístic. Igualment, els paràmetres del tercer grup, referits com la 'forma de parlar', no tenen un valor lingüístic en aquest sentit. En canvi, el conjunt del grup de paràmetres fonològics es considerarà en el context dels grups auditiu lingüístic i acústic lingüístic, ja que les unitats bàsiques en les quals es basa l'anàlisi són els fonemes (i no les realitzacions acústiques d'aquests fonemes).

Segons Nolan (1983: 11), aquest tipus de variables han de complir els següents requisits,:

- Presentar una variabilitat *inter-parlants* alta i una variabilitat *intra-parlant* baixa.
- Ser resistents als intents d'imitació.
- Tenir una alta freqüència d'ocurrència en les mostres d'anàlisi.
- Ser robustes en la transmissió (és a dir, que la transmissió no alteri o elimini les propietats acústiques).
 - Ser relativament fàcils d'extraure i analitzar.

A més a més, Rose (2002: 52) afegeix una nova condició:

- Cada paràmetre ha de ser independent dels altres.

Consulteu la llista completa de variables per a cada mòdul en el ANNEX: VARIABLES.

4. RESULTATS

4.1 Desenvolupament dels objectius plantejats

Objectiu 1

Demostrar que hi ha més distància (variació lingüística) entre parlants i escriptors (**inter**) que no en la parla o l'escriptura d'un mateix (**intra**) parlant o escriptor.

Progrés i consecució de l'objectiu 1

En tots els mòduls (fonològic, morfosintàctic i pragmàtico-discursiu) de cada llengua considerada (català, espanyol, anglès) —excepte en el cas del mòdul pragmàtico-discursiu de l'espanyol amb només un temps de medició— i amb els tres mètodes estadístics usats, els valors de l'**ISI** que corresponen a les comparacions de produccions lingüístiques d'un mateix individu són més alts (**intra: IIS > 0.9 i 0.8**) que els valors obtinguts quan es comparen dos individus (**inter: IIS > 0,6 i < 0,8**), **resultat** que confirma la hipòtesi 1 i s'assoleix l'objectiu 1. El fet que aquests valors **inter** de l'**ISI** no siguin més baixos s'explica metodològicament, en el sentit que tots els parlants /escriptors considerats en els tres mòduls i llengües pertanyen al mateix dialecte, ja que no va ser possible recollir mostres de parlants /escriptors de més d'un dialecte que a més a més estiguessin estratificades en dos temps de medició (**MT1 i MT2**).

Objectiu 2

Demostrar que l'estil idiolectal d'un parlant/escriptor es manté relativament estable al llarg del temps.

Progrés i consecució de l'objectiu 2

En tots els mòduls (fonològic, morfosintàctic i discursiu-pragmàtic) de cada llengua considerada (català, espanyol, anglès), i amb els tres mètodes estadístics usats, els valors de l'**ISI** que corresponen a les comparacions de produccions lingüístiques d'un mateix individu són més alts (**intra: IIS > 0.9 i 0.8**), la qual cosa indica que l'estil idiolectal d'un individu (parlant/escriptor), a partir de les variables considerades en cada mòdul per establir l'**ISI**, no varia substancialment al llarg del temps, i es confirma, per tant, la hipòtesi 2 i s'assoleix l'objectiu 2.

Objectiu 3

Demostrar que l'estil idiolectal d'un parlant/escriptor és menys estable quan es comparen gèneres textuais produïts per un mateix individu que quan es comparen produccions d'aquest mateix individu en dos temps de medició.

Progrés i consecució de l'objectiu 3

Com s'ha comentat més amunt aquest objectiu no s'ha assolit, a causa d'una limitació metodològica, en el sentit que no va ser possible estratificar totes les mostres de tots els mòduls i llengües no només en dos temps de medició sinó també per al menys dos gèneres textuais.

4.2 Activitats realitzades i resultats assolits

El projecte s'ha desenvolupat a partir del concepte d'**Índex de Similitud (o Distància) Idiolectal**, formalitzat en la **Figura 7**.

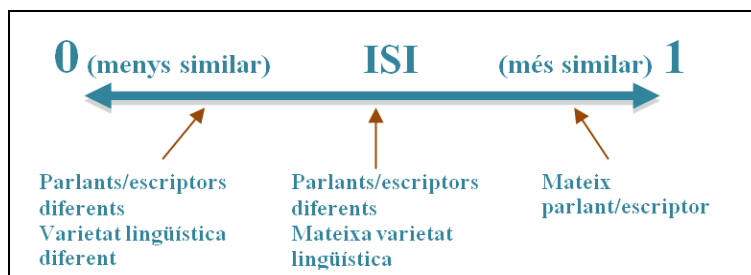


Figura 7. Representació general de l'ISI

La recollida de dades dels diferents **corpus** s'ha basat en la tècnica laboviana d'entrevista sociolingüística, o tècniques molt similars, així com en l'explotació de corpus institucionals o descarregables per Internet, tots ells de parla semi-spontània.

Els **mètodes estadístics**¹ que s'han usat són els següents:

Mètode 1

Comparació del percentatge d'ocurrències de cada variant: a partir del càlcul i la comparació del percentatge d'ocurrència de les variants de cada variable per als dos parlants de la parella de parlants comparada.

Mètode 2

Càlcul del Valor Ajustat Residual (en anglès ARV): a) càlcul de taules de contingència i proves de la Chi-quadrada; b) obtenció del Valor Ajustat Residual (ARV), que indica si els dos parlants analitzats tenen una distribució similar de les variants o no; quan més gran és la diferència entre els parlants comparats més gran és aquest ARV; c) ja que aquest ARV pot ser infinit, es fa l'assignació d'un valor a cada rang d'ARV fins a un màxim de 5; i d) el càlcul de l'ISI s'obté mitjançant el càlcul de la suma de valors de cada variable partit per la multiplicació del nombre de variables per 5 (el valor màxim de conversió d'ARV), segons la fórmula següent:

¹ Durant el procés experimental, també s'ha aplicat el mètode de distància Euclidiana, que finalment s'ha descartat en el procés d'avaluació per no haver donat resultats positius en tots els mòduls i llengües.

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

Mètode 3

Càlcul del coeficient de Phi: a) càlcul de les proves de la chi-quadrada i obtenció del coeficient de Phi, que es situa entre 0 i 1 i indica la força de la relació entre les variables. El quadrat del coeficient de Phi representa el percentatge de variància conjunta; b) l'**ISI** s'obté mitjançant el càlcul de la mitjana de quadrats de Phi:

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

4.3 Resultats per mòduls

4.3.1 Mòduls fonològics

- **Mòdul fonològic del català:** 6 parlants (4 del dialecte de La Canonja, comunitat de parla del tarragonès, i 2 del dialecte barceloní); 18 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Investigador responsable d'aquest mòdul: **Dr. Jordi Cicres Bosch**.
- **Mòdul fonològic de l'anglès:** 6 parlants (tots del dialecte del sud d'Anglaterra); 14 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Analitzat per **Núria Gavalrà**, un dels membres investigadors del projecte, becària FPU amb tesi doctoral en curs, realitzada en el marc del projecte, i a punt de finalitzar (estiu 2013).
- **Mòdul fonològic de l'espanyol:** 5 parlants del mateix dialecte mexicà, 5 en MT1 i 5 en MT2; 13 variables i 3 mètodes estadístics aplicats. Analitzat per la **Dra. Fernanda López**, membre del projecte de recerca.

Els resultats de l'**ISI** obtinguts a partir de l'aplicació dels 3 mètodes a aquest conjunt de mòduls fonològics confirmen la hipòtesi 1, en el sentit que els valors de l'**ISI** són baixos (propers al 0), en el cas de la variació inter-parlant, tal i com es pot observar a la **Figura 8**, que reflecteix els resultats de l'aplicació del mètode 3. En aquest gràfic els valors oscil·len entre **0.2 i 0.80**, sent els més baixos els obtinguts en el mòdul fonològic del català, indicant, distància idiolectal. A més a més, l'aplicació d'aquest mètode al mòdul fonològic del català ha permès observar que hi ha diferències entre parlants de diferents dialectes o varietats, resultant enormement útil en casos forenses reals.

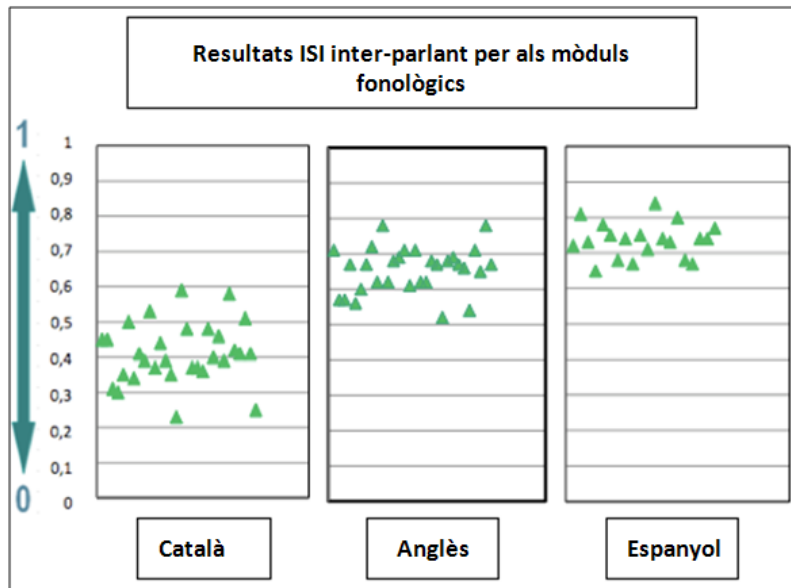


Figura 8. ISI dels mòduls fonològics – Variació inter-parlant

La **Figura 9**, que reflecteix els resultats de l'aplicació del mètode 3, mostra que la hipòtesi 2 per aquests mòduls fonològics també es confirma, en el sentit que no hi ha grans diferències al llarg del temps en l'estil idiolectal dels parlants (variació intra-parlant) de les tres llengües analitzats, ja que els valors oscil·len entre **0.9** i **0.8**, com mostren els punts vermells del gràfic, que indiquen el càlcul de l'ISI de cada parlant en MT1 i MT2,

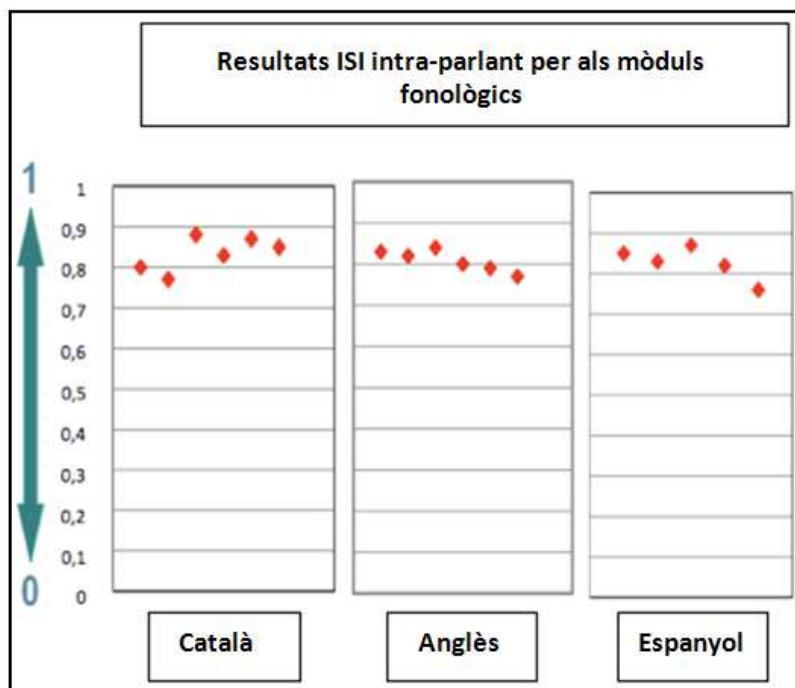


Figura 9. ISI – Fonològics – Variació Intra-parlant

4.3.2 Mòduls morfosintàctics

- **Mòdul morfosintàctic del català:** 6 parlants (tots del dialecte tarragoní de La Canonja); 7 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Mòdul analitzat per la Dra. **Montserrat Forcadell**, un dels membres investigadors del projecte.
- **Mòdul morfosintàctic de l'anglès:** 6 parlants (tots del dialecte del sud d'Anglaterra); 7 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Mòdul analitzat per la Dra. **M. Teresa Turell**, investigadora principal del projecte .
- **Mòdul morfosintàctic de l'espanyol:** 6 escriptors (tots del mateix dialecte de l'espanyol peninsular); 10 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Mòdul analitzat per la Dra. **Maria Spassova**, un dels membres investigadors del projecte

Els resultats de l'ISI obtinguts a partir de l'aplicació dels 3 mètodes a aquest conjunt de mòduls morfosintàctics confirmen la hipòtesi 1, en el sentit que els valors de l'ISI són relativament baixos en el cas de la variació inter-parlant, oscil·lant la majoria entre **0.8 i 0.6**, tal i com a s'esperava, i com es pot observar en la **Figura 10**, que reflecteix els resultats de l'aplicació del mètode 3.

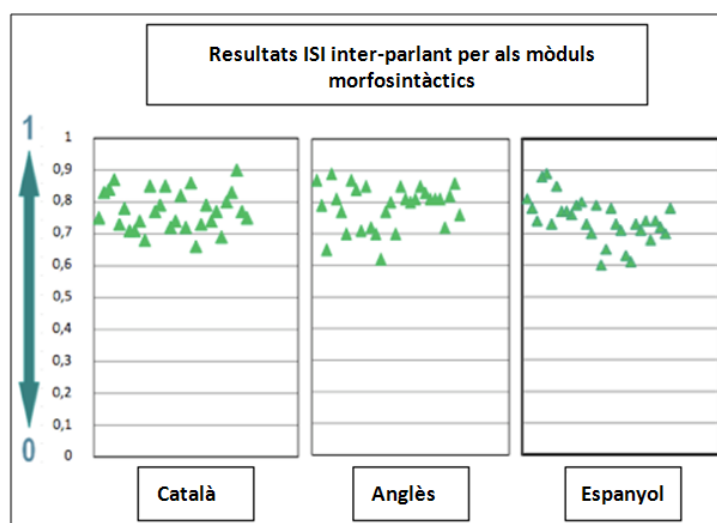


Figura 10. ISI – Morfosintàctics – Inter-parlant/escriptor

La **Figura 11**, que reflecteix els resultats de l'aplicació del mètode 3, mostra que la hipòtesi 2 per a aquests mòduls morfosintàctics es confirma, en el sentit que no hi ha grans diferències al llarg del temps en l'estil idiolectal dels parlants de les tres llengües analitzats, com mostren els punts vermells del gràfic, que indiquen el càlcul de l'ISI de cada parlant/escriptor /escriptor en MT1 i MT2, ja que els valors de la majoria dels parlants oscil·len entre **0.9 i 0.8**, tal i com s'esperava.

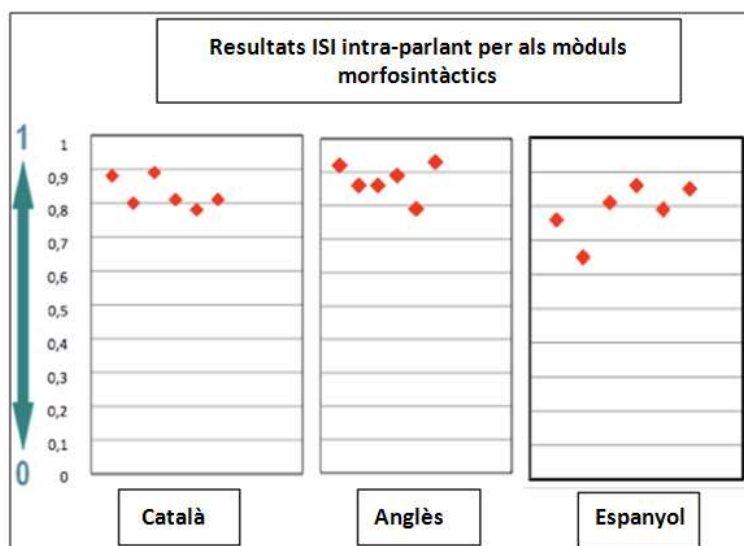


Figura 11. ISI – Morfosintàctics – Intra-parlant/escriptor

4.3.3 Mòduls discursius-pragmàtics

- **Mòdul discursiu-pragmàtic del català:** 6 parlants (tots del dialecte tarragoní de La Canonja); 8 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Mòdul analitzat per la Dra. **M. Teresa Turell**, amb el suport de la **Sheila Queralt**, una de les investigadores contractades del ForensicLab (Institut Universitari de Lingüística Aplicada, UPF).
- **Mòdul discursiu-pragmàtic de l'anglès:** 6 parlants (tots del dialecte del sud d'Anglaterra); 9 variables, 2 temps de medició (MT1 i MT2) i 3 mètodes estadístics aplicats. Mòdul analitzat durant els últims 3 mesos de l'execució del projecte pel **David García Barrero**, nou becari FPU del ForensicLab (Institut Universitari de Lingüística Aplicada, UPF), que aplicarà el protocol de l'ISI a l'àrab, tal i com s'ha explicat en la memòria de sol·licitud d'un nou projecte en la convocatòria del Plan Nacional de 2012.
- **Mòdul discursiu-pragmàtic de l'espanyol:** 6 parlants (tots del mateix dialecte de l'espanyol peninsular); 10 variables, 1 únic temps de medició (MT1) i 3 mètodes estadístics aplicats. Mòdul analitzat per la Dra. **Raquel Casesnoves**, un dels membres investigadors del projecte fins que va ser baixa d'aquest projecte al 2012, en ser-li concedit un projecte del Plan Nacional.

Els resultats de l'ISI obtinguts a partir de l'aplicació dels 3 mètodes a aquest conjunt de mòduls discursiu-pragmàtics confirmen la hipòtesi 1, en el sentit que els valors de l'ISI són baixos en el cas de la variació inter-parlant, oscil·lant la majoria entre 0.6 i 0.8, tal i com es pot observar a la **Figura 12**, que reflecteix els resultats de l'aplicació del mètode 3.

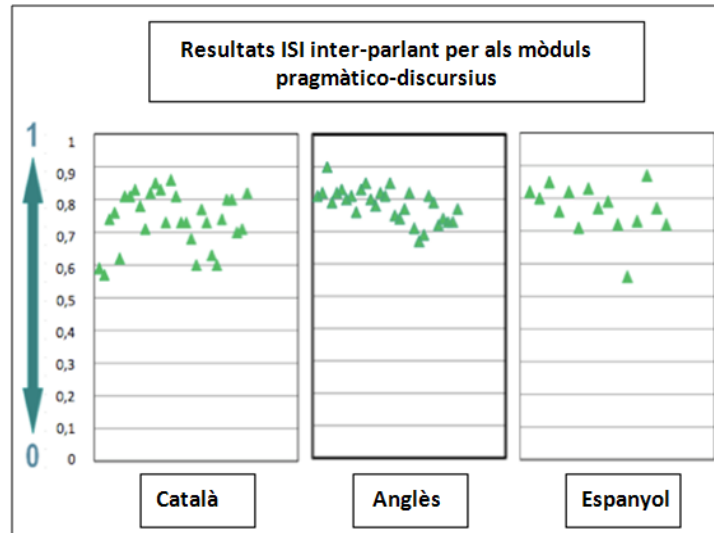


Figura 12. ISI – Discursiu-pragmàtic - Inter

La **Figura 13**, que reflecteix els resultats de l'aplicació del mètode 3, mostra que la hipòtesi 2 per a aquests mòduls discursiu-pragmàtics es confirma, en el sentit que no hi ha grans diferències en l'estil idiolectal dels parlants de l'anglès i del català al llarg del temps, tal i com mostren els punts vermells del gràfic, que indiquen el càlcul de l'ISI de cada parlant en MT1 i MT2, ja que els valors oscil·len entre 0.9 i 0.7. Cal recordar que no es pot calcular l'ISI intra-parlant per al mòdul discursiu-pragmàtic de l'espanyol perquè no hi ha dades del temps de medició MT2.

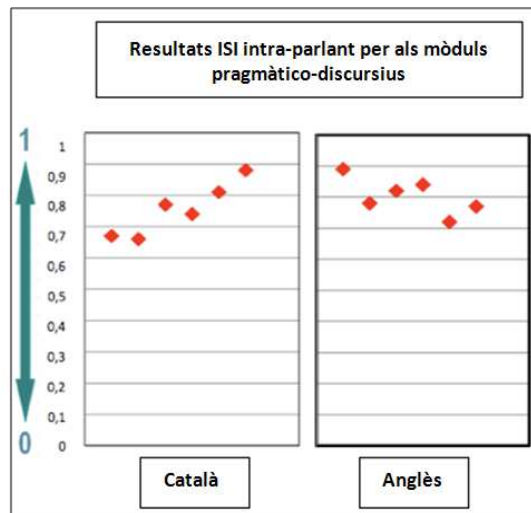


Figura 13. ISI – Discursiu-pragmàtics - Intra

En el cas del mòdul discursiu-pragmàtic de l'espanyol, els resultats del qual presenten molt zeros (0), sembla interessant remarcar que el mètode 3, basat en el càlcul dels valors de la

Chi-quadrada i del valor Phi, funciona especialment bé, perquè malgrat aquesta distribució desigual de les dades l'anàlisi pot continuar sense problemes.

4.4 Conclusions finals

1. El mètode 3 resulta ser el més fiable i el que permet obtenir resultats més concloents.
2. Per a tots els mòduls i llengües es confirma la hipòtesi 1 en el sentit que hi ha més variació i, per tant, major distància idiolectal entre parlants i escriptors (variació **inter**) que en l'estil idiolectal d'un mateix individu (variació **intra**).
3. Per a tots els mòduls i llengües es confirma la hipòtesi 2 en el sentit que l'estil idiolectal d'un individu no varia massa al llarg del temps.
4. Els valors de l'**ISI** inter-parlant/escriptor poden semblar massa alts, però s'ha de tenir en compte que, exceptuant el mòdul fonològic del català, en tots els mòduls s'han considerat parlants i escriptors del mateix dialecte.

Referències

- 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

ANNEX: VARIABLES

Taula 1: Variables per al mòdul fonològic del català

VARIABLE	VARIANTES	EXEMPLES
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>

Taula 1: Variables per al mòdul fonològic del català

VARIABLE	VARIANTES	EXEMPLES
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>

Taula 2: Variables per al mòdul fonològic del espanyol

VARIABLE	VARIANTS	EXEMPLES
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>

Taula 2: Variables per al mòdul fonològic del espanyol

VARIABLE	VARIANTS	EXEMPLES
11 Pèrdua de la vocal /e/ àtona al interior de la paraula /entonses/ à [entons]	1. [en'tonses] 2. [en'tons]	
12 Reducció del hiato /ea/ a /a/ en la paraula /osea/	1. [o'sa] 2. [o'sea]	
13 Canvi de la paraula /entonces/ per /tons/	1. [en'tonses] 2. [tons]	

Taula 3: Variables per al mòdul fonològic del anglès

VARIABLE	VARIANTS	EXEMPLES
----------	----------	----------

[Anàlisi i dades derivats d'una tesi doctoral en curs.]

Taula 4: Variables per al mòdul morfosintàctic del català

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

Taula 4: Variables per al mòdul morfosintàctic del català

VARIABLE	VARIANTS	EXEMPLES
	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 canvia , 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

Taula 5: Variables per al mòdul morfosintàctic del espanyol

VARIABLE	VARIANTS	EXEMPLES
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>

Taula 5: Variables per al mòdul morfosintàctic del espanyol

VARIABLE	VARIANTS	EXEMPLES
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 advertírselo..."</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)

Taula 6: Variables per al mòdul morfosintàctic del anglès

VARIABLE	VARIANTS	EXEMPLES
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>

Taula 7: Variables per al mòdul discursiu-pragmàtic del català

VARIABLE	VARIANTS	EXEMPLES
1 Resposta a pregunta oberta amb marcadors discursiu	1. Presència de marcadors 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>

Taula 7: Variables per al mòdul discursiu-pragmàtic del català

VARIABLE	VARIANTS	EXEMPLES
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>

Taula 7: Variables per al mòdul discursiu-pragmàtic del català

VARIABLE	VARIANTS	EXEMPLES
	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- Llavors 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>

Taula 8: Variables per al mòdul discursiu-pragmàtic del espanyol

VARIABLE	VARIANTS	EXEMPLES
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/ h́ablanos 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>

Taula 8: Variables per al mòdul discursiu-pragmàtic del espanyol

VARIABLE	VARIANTS	EXEMPLES
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	

Taula 9: Variables per al mòdul discursiu-pragmàtic del anglès

VARIABLE	VARIANTS	EXEMPLES
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]

Taula 9: Variables per al mòdul discursiu-pragmàtic del anglès

VARIABLE	VARIANTS	EXEMPLES
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>