Forensic Idiolectometry and Index of Idiolectal Similitude

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

Updated: January 2013

1. Presentation

This research is the result of two projects carried out between 2007 and 2011, and an ongoing research project until 2015:

Phase I

Project: **Idiolectometry applied to forensic linguistics** (EXPLORA-INGENIO HUM2007-29140-E)
PI: Dr. M. Teresa Turell
Funding entity: Ministerio de Educación y Cultura
1) Period: 2007-2008

Phase II

Project: **Forensic idiolectometry and index of idiolectal similitude** (FII2008-03583/FILO)
PI: Dr. M. Teresa Turell
Funding entity: Ministerio de Ciencia e Innovación
Period: 2008-2012

Phase III

Project: **Hacia la consolidación de un Índice de Similitud/Distancia Idiolectal (IS/DI) en Idiolectometría Forense** (FFI2012-34601)
PI: Dr. M. Teresa Turell
Funding entity: Ministerio de Economía y Competitividad
Period: 2012-2015

1.1 Aim

This project aims to study the speakers' **idiolectal style** in its application to **forensic linguistics (**in particular to **forensic phonetics**, in order to identify speakers and establish linguistic profiles, on the one hand, and to **authorship determination/attribution** of written texts, on the other). A speaker's **idiolectal style** can be defined as the set of options that he/she takes from the linguistic repertoire (phonological, morpho-syntactic, pragmatics forms) available to him/her as a speaker/writer of a specific language (Nolan 1994: 331). Thus, a speaker's idiolectal style is individual and unique.

Idiolectometry is the emerging discipline which studies the idiolect**.** So far this discipline has measured the linguistic distance between speakers and has established the borderline between different idiolects; a borderline which by definition keeps a person separated from the rest of speakers. By contrary, what this project will explore and develop is the possibility of measuring the linguistic differences existing between idiolects and each individual's idiolectal distance, so that an **Index of Idiolectal Similitude** (IIS), which will compare several linguistic

samples and calculate the linguistic distance between them, can be obtained. More specifically, it is a question of being able to establish what kind of idiolectal similitude one needs to have before one can say that two linguistic samples (spoken or written) have been produced by the same person. The application of the study of the idiolectal style to forensic linguistics is fundamental since this application will allow linguists acting as expert witnesses in court to unmistakenly identify speakers or writers by comparing a disputed recording or written text and a set of non-disputed spoken or written texts from the set of options chosen by each individual speaker or writer. A protocol will de devised in order to set up the above mentioned IIS from which, once different linguistic parameters (phonological, morphosyntactic, discourse-pragmatic) are analysed, it will be possible to decide whether or not two recordings or two written texts have been produced by the same person.

In this project, a) the IIS technique - already designed through the execution of an EXPLORA - INGENIO project ((HUM2007-29140-E); UVAL-ForensicLab group (http://www.iula.upf.edu/forensiclab; http://www.iula.upf.edu/uval; execution period: December 1, 2007 - November 30, 2008) - will be applied to three languages (Spanish, Catalan and English), b) the global design of the protocol will be evaluated and c) the statistical methods for all languages will be devised.

1.2 Background and State of the Art

The theoretical framework underlying this project is grounded on three disciplines: **forensic linguistics** (in particular, on forensic phonetics, in order to identify speakers and establish linguistic profiles, on the one hand, and on authorship determination/attribution of written texts, on the other; **idiolectometry**, and finally, the **theory of language variation and change**, more specifically, sociolinguistic variation.

Forensic Linguistics can be defined as the interface between language and the law. This discipline includes the study of a number of areas, which have to do with the use of linguistic evidence within diverse public and professional contexts (http://www.iafl.org). From a methodological point of view, forensic linguistics expertise and research is implemented by means of a series of tools, software, and quality statistics which allow forensic linguists to show a much more rigorous and scientific performance to be used by the public administration (judicial school, police,) and private institutions and companies, and also by professional people (judges, lawyers, attorneys, solicitors, notaries, psychologists, doctors). In any case, the evidence which is presented at court is usually complementary with other types of evidence and useful to introduce a reasonable doubt.

Idiolectometry is the emerging discipline which studies the idiolect**.** So far this discipline has measured the linguistic distance between speakers and has established the borderline between different idiolects. A speaker's **idiolectal style** can be defined as the set of options that he/she takes from the linguistic repertoire (phonological, morphosyntactic, pragmatics forms) available to him/her as a speaker/writer of a specific language (Nolan 1994: 331). Thus, a speaker's idiolectal style is individual and unique. At the same time, we can envisage a language as formed by the sum of dialects and sociolects, and these in turn are defined as the sum of idiolects (or individual uses of language). Social factors, namely socioeconomic group,

age, educational level, gender, profession, manifest themselves in the configuration of an individual's idiolect. Figure 1 shows this linguistic representation:

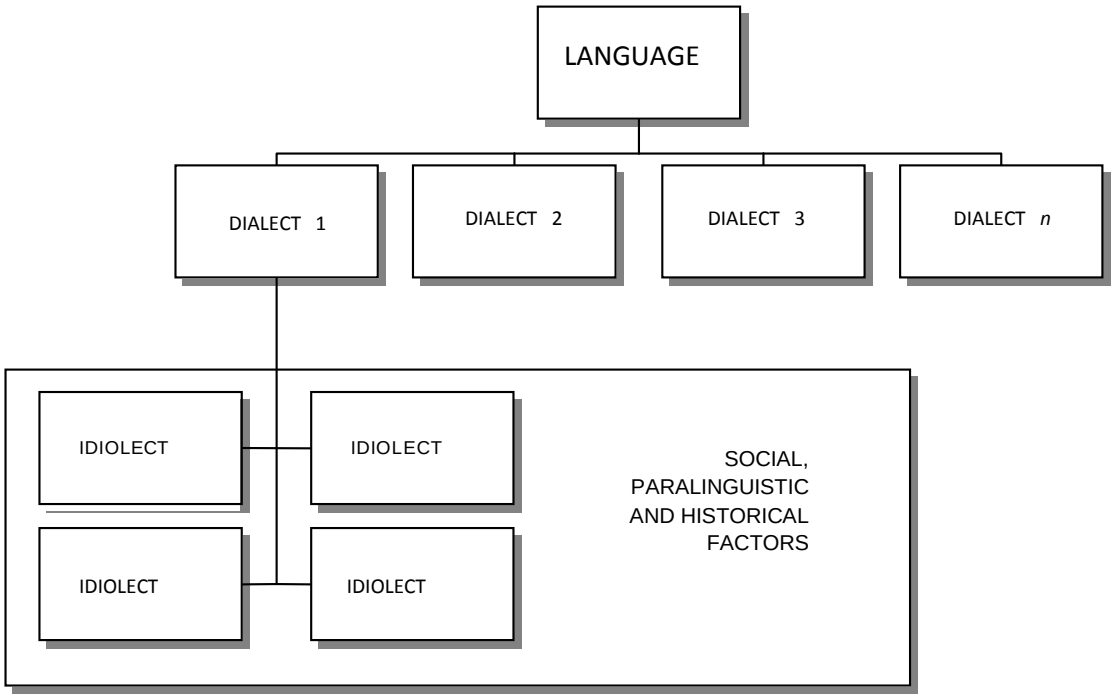

Figure 1. Diagram showing the structure of languages, formed by dialects, and these by idiolects, which are constrained by social, paralinguistic and historical factors.

Studies analysing the effect of external linguistic factors on linguistic forms are based on the **theory of language variation and change** (Labov 1994, 2001; Turell 1995), which postulates that variation is inherent to all languages and affects all linguistic levels: phonetics and phonology, morphology, syntax, semantics, discourse, and pragmatics. Sociolinguistic variation studies the way in which linguistic variation is structured, by taking into account those internal (linguistic) and external (social, stylistic) factors that intervene in structured variation. Several studies have shown the effect of style (Schilling-Estes 2002), gender, age, ethnicity, socioeconomic class (Labov 1966, 1994, 2001), or that of social groups or networks (Milroy 1987, Eckert 2000) on linguistic productions, at the phonological, morphological, syntactic and semantic levels. Thus, sociolinguistics considers the structure and nature of variation at different levels and from different perspectives, as is illustrated in Figure 2 below:

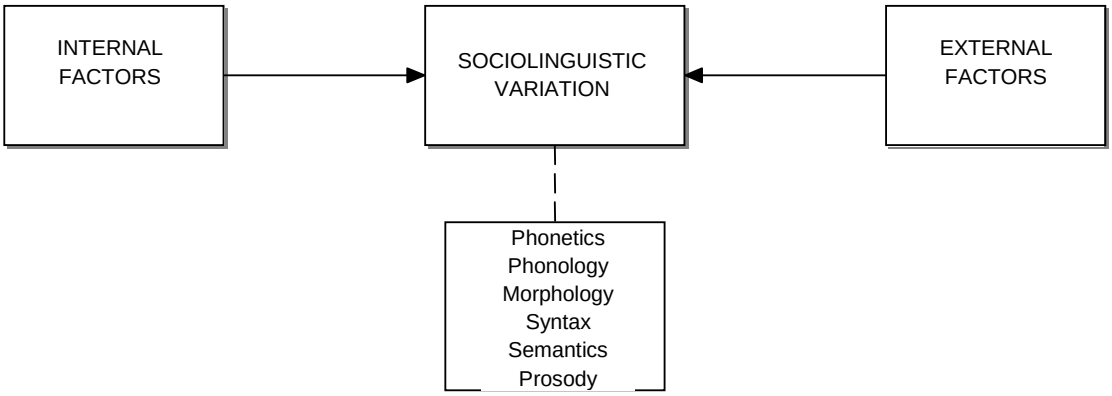

Figure 2. Factors structuring variation

However, sociolinguistic variation has only very rarely been interested in the study of an individual's idiolect, except for quite general studies such as Guy (1980), Abercrombie (1969) or Biber (1988, 1995).

1.3 Members and modules considered

- **Catalan:**
  - Phonological module of Catalan: Dr. Jordi Cicres – UPF and UdG.
  - Phonological module of Spanish: Dr. Fernanda López – UPF and UNAM.
  - Phonological module of English: Núria Gavaldà – UPF.
- **Spanish:**
  - Morpho-syntactic module of Catalan: Dr. Montse Forcadell – UPF and UB.
  - Morpho-syntactc module of Spanish: Dr. Maria Spassova – UPF and NBU.
  - Morpho-syntactic module of English: Dr. M. Teresa Turell – UPF.
- **English:**
  - Discourse-pragmatic module of Catalan: Sheila Queralt – UPF, from 2010.
  - Discourse-pragmatic module of Spanish: Dr. Raquel Casesnoves – UPF, until 2010.
  - Discourse-pragmatic module of English: David García Barrero – UPF, from 2011.

2. Objectives and Hypotheses

2.1. Originality and interest

This is clearly a 'problem-based' project: society needs expertise and experts in order to identify speakers, to attribute authorship, and to detect plagiarism in a more rigorous and reliable way. The application of the study of the idiolectal style to forensic linguistics is fundamental since this application will allow linguists acting as expert witness in court to unmistakenly identify speakers or writers by comparing linguistic forms/parameters occurring in a disputed recording or written text and those occurring in a set of non-disputed spoken or written texts. While at present there is no such linguistic model which would account for all forensic needs, this projects seeks to answer a number of key questions involved in forensic linguistic analysis. The results obtained through the execution of this project could be used in real cases where linguistic expertise is needed in order to (among other actions):

- Identify speakers and come up with linguistic profiles from voice recordings,
- Determine and attribute authorship of written texts such as suicide notes or threats, etc.,
- Detect plagiarism.

And this would also involve a better administration of justice.

2.2. Hypotheses

1. It is hypothesised that each individual has an idiolectal style which is unique and unreproduceable.
2. It is hypothesised that there will be more inter-speaker/writer than intra-speaker/
3. writer variation.
4. It is hypothesised that a speaker's idiolectal style will not change according to genre o context, in particular as to its phonological and syntactic patterns. Another issue is vocabulary which can be, and usually is, constrained by register and genre.
5. It is hypothesised that an individual's idiolectal style varies very slightly throughout time.
6. Once, and if, these hypotheses are confirmed, it is hypothesised that it will be possible to establish an Index of Idiolectal Similitude, which could help experts in speaker identification and authorship attribution contexts.

2.3. Main aims:

1 To show that every speaker/writer makes use of his/her own **idiolect**, that is, a unique and idiosyncratic linguistic style, whose use is unconscious and which changes very slightly throughout time.

2 To apply the study of the idiolectal style to **forensic linguistics**, since this application can help experts to identify producers of an oral text and writers of a written text more reliably.

2.4. Specific objectives (theoretical and methodological):

- To undertake this application by comparing the linguistic forms/parameters used by speakers/writers in the production of **disputed** spoken or written texts and the linguistic forms/parameters used in a set of non-disputed spoken or written texts and thus confirm that there exists more inter-speaker/writer than intra-speaker/writer variation.
- To undertake such application by using **apparent** and **real time** measurements (or two measurement times, MT1 and MT2) in order to confirm that an individual's idiolectal style varies very slightly throughout time.
- To measure the linguistic differences between several idiolects and each individual's idiolectal distance so that an **Index of Idiolectal Similitude** can be obtained. More specifically, it is a question of being able to establish what kind of idiolectal similitude one needs to have before one can say that two linguistic samples (spoken or written) have been produced by the same person.
- To devise a protocol for the setting up of the above mentioned index which will compare several linguistic samples distributed by different text-length and genre, calculate the linguistic distance and help to decide whether or not two recordings or two written texts have been produced by the same person.

3. Experimental design

The project will undertake the following tasks:

Evaluation tasks of the global design of the protocol and of the phonological module for for Spanish and Catalan (by considering the results obtained through the implementation of a previous one-year project (EXPLORA-INGENIO, HUM2007-29140-E).

Execution tasks

- The general protocol will be designed for its application to the three languages under analysis (**Catalan, Spanish and English**).
- The protocol will be specified for the **morphosyntactic** and **discourse-pragmatic modules** in the three languages.
- The statistical methods needed to calculate the **Index of Idiolectal Similitude** will be devised for the morphosyntactic and **discourse-pragmatic modules** in the three languages.
- The three modules (phonological, morphosyntactic, discourse-pragmatic) will be evaluated for the three languages.

3.1. Corpora

This project aims to study the speakers'/writers' idiolectal style with forensic applications. Thus, both spoken and written corpora have to be used, and in the latter case, we will follow the guidelines proposed by Coulthard (1994), Eagleson (1994), Turell (1995) and Johnson (1997).

When considering spoken corpora, it is useful to distinguish between *spoken language corpora* and *spoken corpora* (Llisterri *et al.* 2005). The former are transcriptions which aim to represent specific aspects of the spoken language, while the latter are composed at least by sound signal. The degree of segmentation and tagging varies according to the purpose of usage attributed to a specific corpus. Thus, corpora can be segmented in phonemes, diphonemes, morphemes, phrases, intonation units, clauses, sentences, turns, etc; furthermore, they can also incorporate other types of information in order to represent other verbal aspects (tone, illocution speed, intonation, pauses) or non verbal (turn changes, noise, gestures).

In order to evaluate the effect of time on the variables under study, considering the forensic linguistic application content within which this project is framed, it will be necessary to use not only a particular corpus compiled in one measurement time (which could be equated to the term *apparent time* in sociolinguistic variation), but also another corpus obtained in a second measurement time (real *time*), that is, a corpus consisting of material elicited from the same informants, with a time span between the two samples. The notion of apparent and real time has been used, especially in sociolinguistics (Turell 2003), since the thirties, but above all, since the fifties of the 20th century.

3.1.1. Corpus for the Catalan IIS

In order to consider all modules in Catalan we will use the part in Catalan of a spoken corpus in apparent and real time: La Canonja corpus (Pujadas, Pujol Berché, Turell), compiled by the UVAL (Unitat de Variació Lingüística) group (Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra). Initially, this corpus consisted of 29 sociolinguistic interviews (13 Catalan-speaking informants and 16 Spanish-speaking informants), which were recorded in the eighties (MT1) and then the same speakers were recorded again in the first decade of the XXI century (MT2 - 2006-2008).

3.1.2. Corpora for the Spanish IIS

For the phonological module of Spanish, we will use the part in Spanish of a spoken corpus in apparent and real time: La Canonja corpus (Pujadas, Pujol Berché, Turell), compiled by the UVAL (Unitat de Variació Lingüística) group (Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra). Initially, this corpus consisted of 29 sociolinguistic interviews (13 Catalan-speaking informants and 16 Spanish-speaking informants), which were recorded in the eighties (MT1) and then the same speakers were recorded again in the first decade of the XXI century (MT2 - 2006-2008).

For this same module, we will also use a corpus in apparent and real time compiled for Mexican Spanish through a CONACYT scholarship awarded to Fernanda López to do her PhD. Thesis: the DIMEx100 corpus, from the Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas at Universidad Nacional Autónoma de México, to be used to create an automatic

speech recognition device. The corpus consists of 100 recordings from 100 speakers who read 10 sentences identical for all and 50 sentences different for all informants. These were selected by age (between 16 and 36 years of age), their educational level (secondary and higher education), and their origin (Ciudad de México).

For the morpho-syntactic module of Spanish, we will use the written corpus compiled by Maria Spassova to do her PhD. thesis, consisting of texts in Spanish written by 20 writers from Spain and several South-American countries, with two sub-corpora: novels (N) and newspaper articles (NA).

For the discourse-pragmatic module of Spanish, we will use the Preseea corpus (*Proyecto para el Estudio Sociolingüístico del Español de Valencia*, http://www.uv.es/preseval/ppal.htm), from which we will consider informants of high and middle sociocultural level and we will control for gender and age (Gómez Molina, J. R. 2001 and 2005).

3.1.3. Corpora for the English IIS

For all modules in English, we will use an English corpus consisting of 16 speakers of British English in two measurement times (MT1 and MT2), extracted from interviews available in Youtube and TV and radio channels.

3.2. Variables

Firstly, we'll base our analysis on phonological variables. We follow Rose (2002) on the difficulty encountered when comparing different voice samples, as the following diagrams show. Figure 3 shows an ideal situation where the variable realizations are unique for each speaker (mean values for F0):

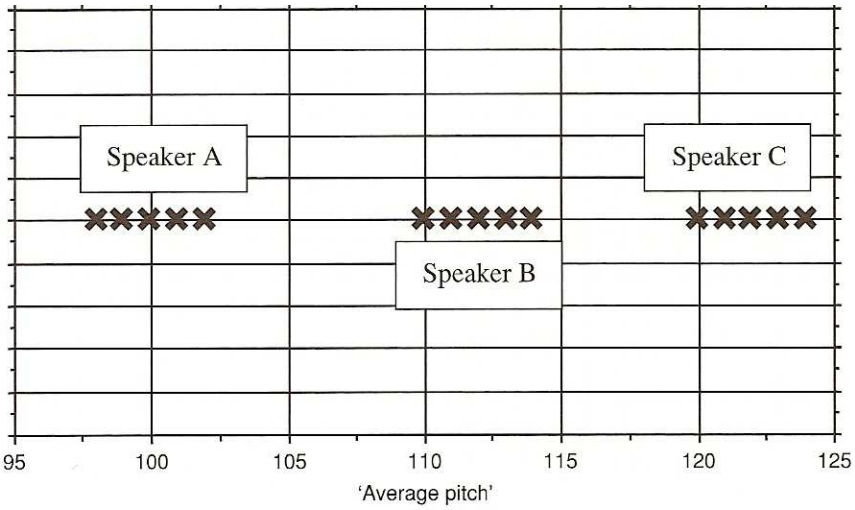

Figure 3. "Ideal" representation of the values of one variable in three speakers. Source: Rose (2002).

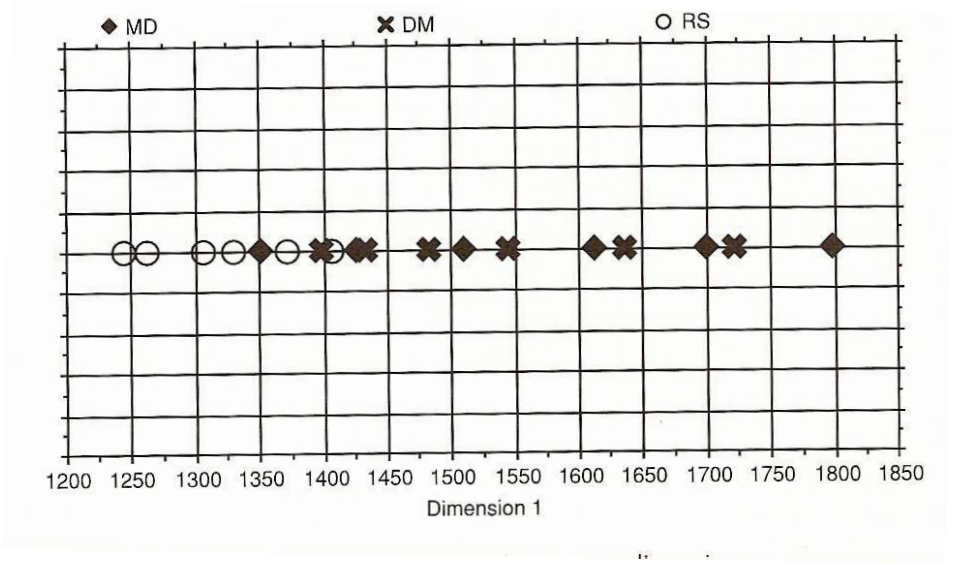

Figure 4. "Realistic" representation of the values of one variable in three speakers, without superposition. Source: Rose (2002).

The previous figure shows that the values of one single variable are not enough to discriminate speakers and so as Rose (2002) claims new parameters and dimensions must be incorporated to forensic speaker identification in order to obtain delimited spaces for each speaker. This is shown in Figures 5 (2 dimensions) and 6 (3 dimensions).

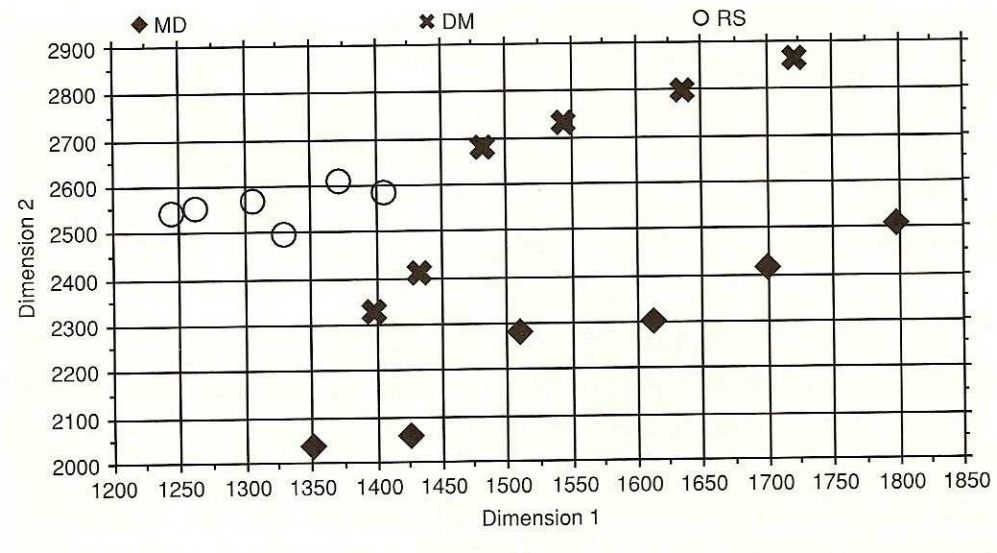

Figure 5. Representation of the values of two variables in three speakers. Source: Rose (2002).

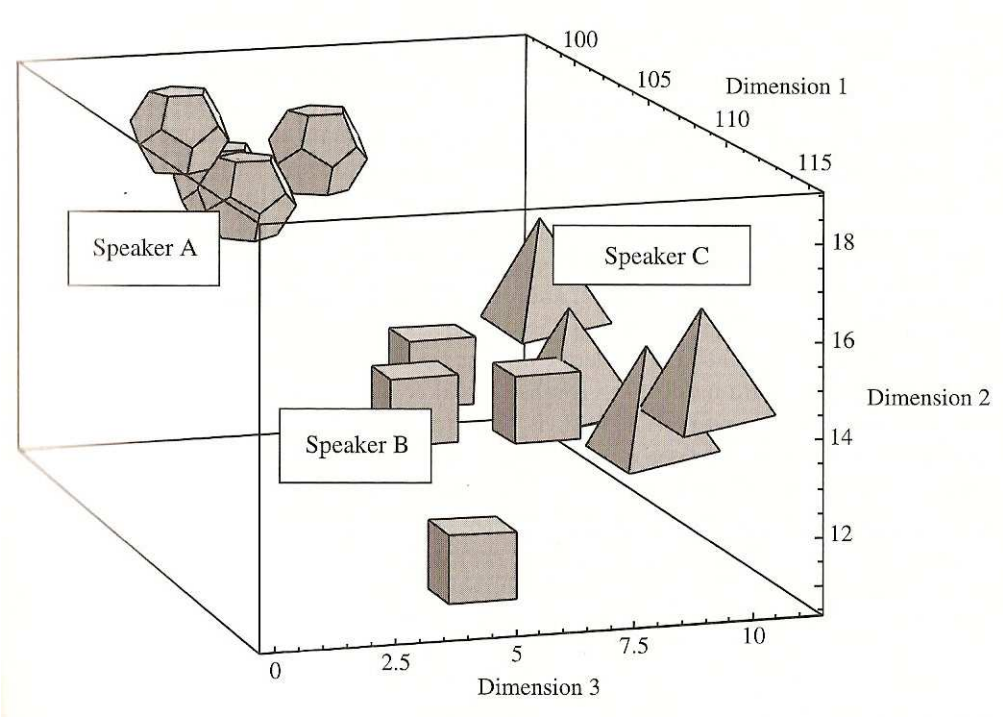

Figure 6. Representation of the values of three variables in three speakers. Source: Rose (2002).

But as it will be shown, the analysis with three variables will still not allow researchers to draw reliable conclusions, so the number of variables will have to be increased.

We also follow Rose (2002: 33-53) for a primary classification of the parameters to be used in forensic analysis. This classification is displayed in Table 1:

Table 1. **Primary classification of parameters in forensic analysis, according to Rose (2002).**

	LINGUÍSTIC	*NON LINGUÍSTIC*
AUDITORY	**AUDITORY-LINGUISTIC**	**AUDITORY- NON LINGUISTIC**
ACOUSTIC	**ACOUSTIC- LINGUISTIC**	**ACOUSTIC-NON LINGUISTIC**

These variables will have to comply with the following requirements, according to Nolan (1983: 11). They should:

- Show high inter-speaker and low intra-speaker variability.
- Be resistant to disguise attempts.
- Exhibit a high frequency of occurrence in the samples under analysis.
- Be robust in the transmission.
- Be easily retrievable.

Rose (2002: 52) adds a new condition:

- Each parameter has to be independent from the others.

Please find a list of variables for all modules and languages in the ANNEX: VARIABLES.

4. RESULTS

4.1 Attainment of objectives

Objective 1	**Progress and attainment of objective 1**
To show that there is more idiolectal distance (linguistic variation) between speakers/writers (**inter variation**) than in the speech or writing of one same individual (**intra variation**).	For all modules and languages —except in the discourse-pragmatic module of Spanish, with one measurement time only— **hypothesis 1** is confirmed (and **objective 1** is attained) in that there is **more variation** and thus more idiolectal distance **between speakers and writers'** samples than between two samples of the same speaker or writer, which show a quite steady idiolectal similitude throughout time. With the three statistical methods used, the IIS values that correspond to the comparisons of linguistic samples produced by the same individual are higher (**intra: IIS > 0.9** and **0.8**) than the values obtained when two different speakers or writers are compared (**inter: IIS > 0.6** and **< 0.8**)**.** The fact that the IIS inter variation values are not as low as expected can be methodologically explained by the fact that all speakers considered in all modules and languages, except for the phonological module of Catalan, belong to the same linguistic variety or dialect, since it was not possible to compile samples for more than one dialect that would in turn be stratified in two measurement times (**MT1** and **MT2**).
Objective 2	**Progress and attainment of objective 2**
To show that a speaker's/writer's idiolectal style remains quite stable throughout time.	For all modules and languages, **hypothesis 2** is confirmed (and **objective 2** is attained) in that an **individual's idiolectal style** (spoken or written) **does nott seem to vary much throughout time.** With the three statistical methods used, the IIS values that correspond to the comparisons of linguistic samples produced by the same individual are high (**intra: IIS > 0.9** and **0.8).**
Objective 3	**Progress and attainement of objective 3**
To show that a speaker's/writer's idiolectal style remains less stable when samples from one individual in different textual genres are compared than with samples in two different measurement times.	This objective has not been attained due to a methodological drawback in the sense that it was not possible to stratify all samples in all modules and languages in two measurement times and at least in two textual genres.

4.2 Activities undertaken and results obtained

This research project has been conducted in terms of an **Index of Idiolectal Similitude (or Distance)**, shown in **Figure 7:**

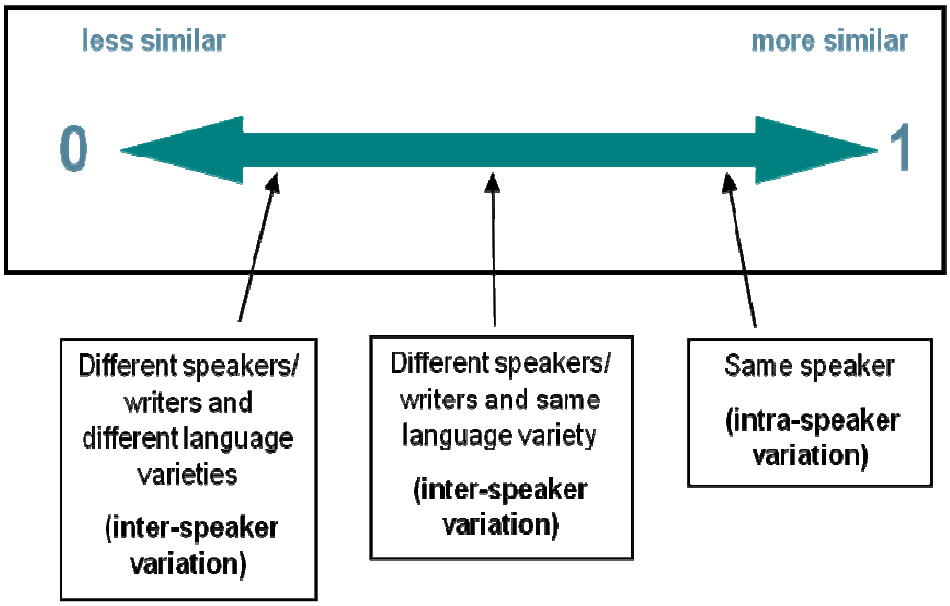

Figure 7. Formalization of the IIS

Data collection was based on the Labovian sociolinguistic interview, or very similar techniques, as well as the exploitaiton of institutional corpora, some unloadable from the Internet, all of them reflecting semi-spontaneous speech.

The statistical methods[1] used are the following:

Method 1

Estimate and comparison of the percentatge of occurrences of each variable and variants considered for each pair of analysed speakers/writers. Application value variant. % of variant realization /each variant and calculate difference between the 2 speakers. Sum up all the % differences and calculate the average. Divide this figure by 100. Substract 1 to the final figure.

[1] During the experimental stage, the Euclidian distance method was also used, but was finally disregarded in the evaluation stage because the results drawn were not very positive in either languages and modules.

Method 2

Calculation of the Adjusted Residual Value. Cross-tabulation running with SPSS. Calculation of the Adjusted Residual Values (ARV)) for each variable. Assigment of a value to each A range:

- ARV <1 → 0
- ARV >1&<2 → 1
- ARV >2&<3 → 2
- ARV > 5 → 5

(max. ARV conversion)

The **IIS** is obtained by the calculation of the following formula:

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

Method 3

Calculation of the Phi coefficient. Calculation of χ^2 and obtention of the Phi coefficient, which is located between 0 and 1, indicating the relation between variables. The square of the Phi coefficient (Phi^2) represents the percentage of overall variance. The IIS is obtained through the calculation of the average of Phi^2 :

$$ISI_{Total} = 1 - \sqrt[2]{\frac{1}{k}\sum_{j}^{k} Phi^2}$$

4.3 Results by modules

4.3.1. Phonological modules

- **Phonological module of Catalan**: 6 speakers (4 from La Canonja variety, *tarragoní*, and 2 from *barceloní*); 18 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Principal investigator: **Dr. Jordi Cicres Bosch.**
- **Phonological module of English**: 6 speakers (all from the Southern British English variety); 14 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Analyzed by **Núria Gavaldà**, one of the members of the research team, holding an FPU

grant to do her PhD., which is almost completed.

- **Phonological module of Spanish**: 5 speakers from the same Mexican Spanish variety, 5 in MT1 and 5 in MT2; 13 variables and 3 statistical methods. Analyzed by **Dr. Fernanda López**, a member of this research project.

Hypothesis 1 is confirmed with all 3 methods for the phonological modules in all 3 languages. With **M3** (Figure 8), all **IIS** values are relatively low in general (between **0.2** and **0.8**), which is an expected result considering that, except for the Catalan corpus, all speakers belong to the same dielactal area**. M3** has proven useful in the case of the phonological module of Catalan in order to observe that, when the **IIS** is calculated between speakers of different varieties, the inter-speaker IIS values are lower than when the speakers compared belong to the same dialectal area, a result which is very relevant in real forensic cases concerned with linguistic profiling.

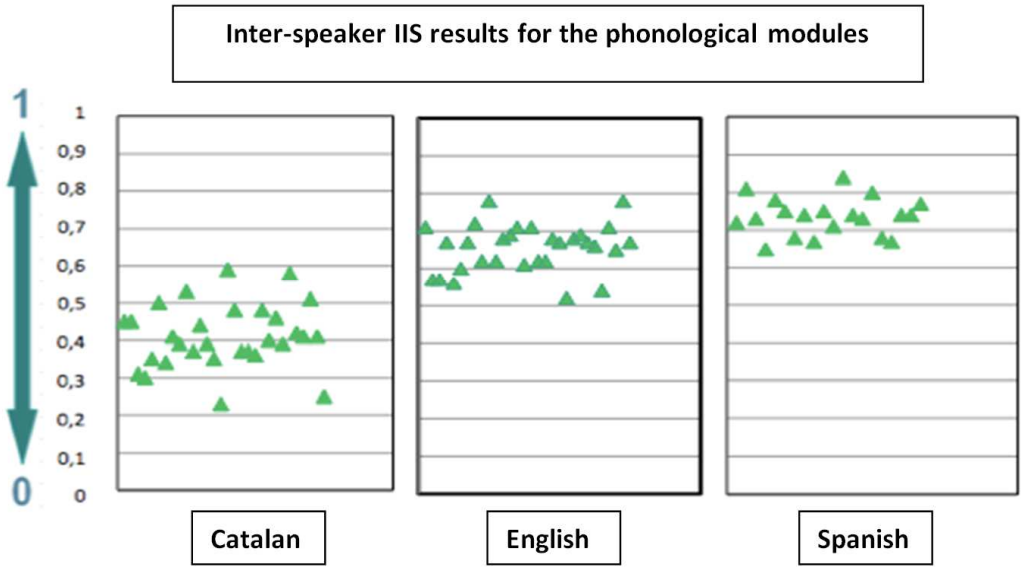

Figure 8. IIS for the phonological modules – Inter-speaker variation

Hypothesis 2, on intra-speaker variation throughout time, is confirmed with all 3 methods for the phonological modules in all 3 languages. With **M3**, all **IIS** values are quite high, as expected, ranging between **0.9** and **0.8, as shown in Figure 9:**

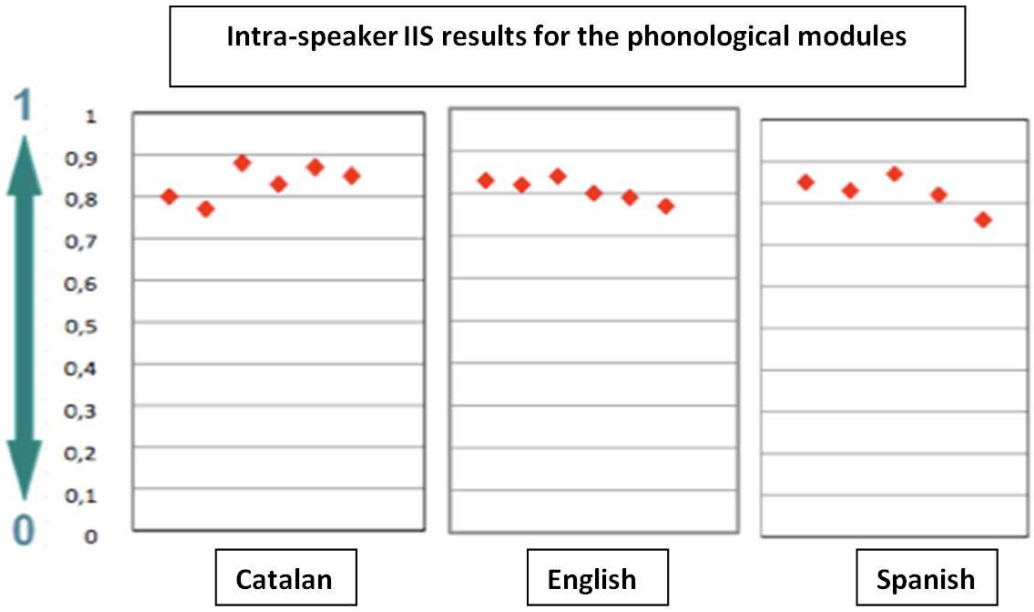

Figure 9. IIS – Phonological modules – Intra-speaker variation

4.3.2. Morpho-syntactic modules

- **Morpho-syntactic module of Catalan:** 6 speakers (all from La Canonja variety, *tarragoní*); 7 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Principal investigator: Dr. **Montserrat Forcadell.**
- **Morpho-syntactic module of English**: 6 speakers (all from the Southern British English variety); 7 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Analyzed by Dr. **M. Teresa Turell,** Principal investigator of the project.
- **Morpho-syntactic module of Spanish**: 6 writers (all from the same dialect of Peninsular Spanish); 10 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Analyzed by Dr. **Maria Spassova**, one of the members of the project.

Hypothesis 1 is confirmed with all 3 methods for the morphosyntactic modules in all 3 languages. As shown in Figure 10, with **M3**, all **IIS** values are relatively low in general (although they range between **0.6** and **0.8**), which is an expected result considering that all speakers belong to the same dielactal area**.**

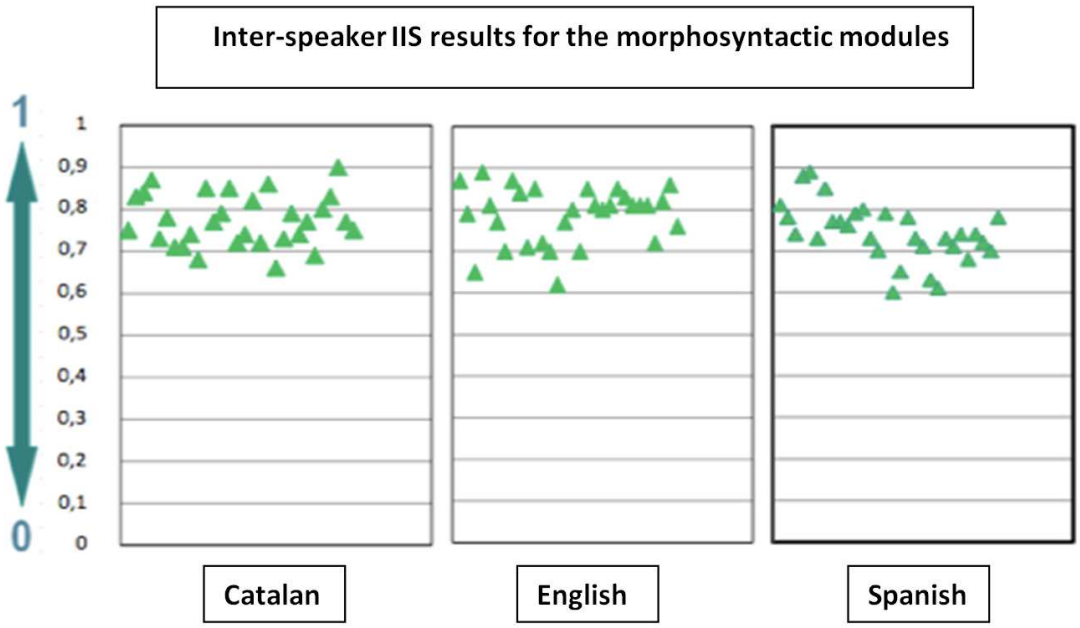

Figure 10. IIS – Morpho-syntactic modules – Inter-speaker/writer variation

Figure 11, shows that hypothesis 2, on intra-speaker/writer variation throughout time, is confirmed with all 3 methods for the morphosyntactic modules in all 3 languages. With M3, all IIS values are quite high as expected (the majority ranging between **0.9** and **0.8**), which is an indication as well that the threshold level at which it is possible to say that two spoken/written samples are from the same speaker/writer has to be located above 0.9.

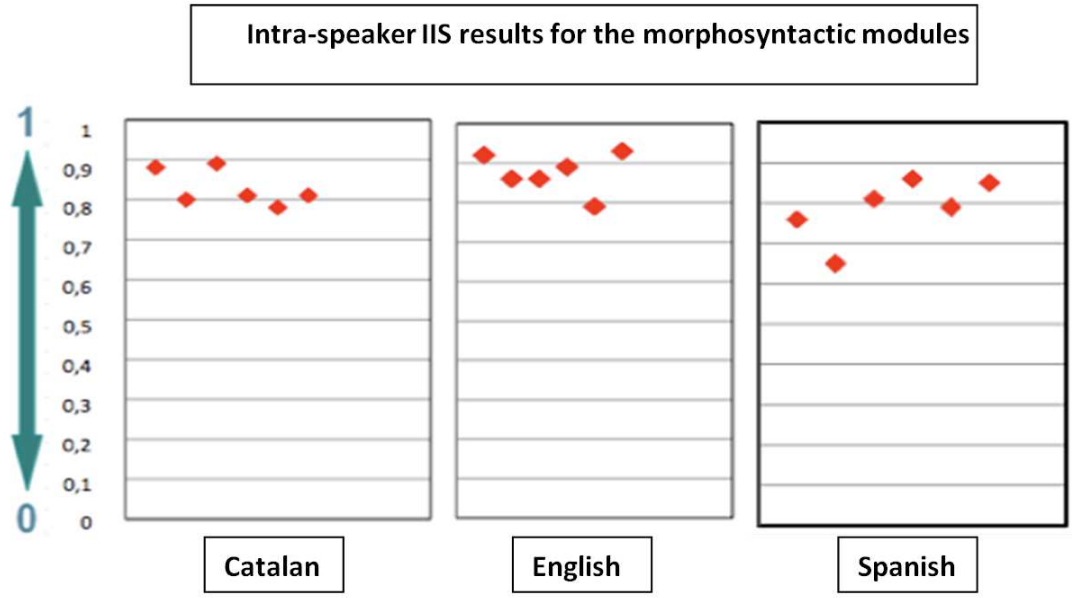

Figure 11. IIS – Morpho-syntactic modules – Intra-speaker/writer variation

4.3.3. Discourse-pragmatic modules

- **Discourse-pragmatic module of Catalan:** 6 speakers (all from La Canonja variety, *tarragoní*), 8 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Analyzed by Dr. **M. Teresa Turell, Principal** investigator of the project, with the suport of Sheila Queralt, one of the researchers appointed by ForensicLab (Institut Universitari de Lingüística Aplicada, UPF).

- **Discourse-pragmatic module of English**: 6 speakers (all from the Southern British English variety); 9 variables, 2 measurement times (MT1 and MT2) and 3 statistical methods. Analyzed by **David García Barrero,** holding a FPU scholarship to do his PhD. on forensic authorship attribution of Arabic.

- **Discourse-pragmatic module of Spanish**: 6 speakers (all from the same Peninsular Spanish variety); 10 variables, 1 measurement time only (MT1) and 3 statistical methods. Analyzed by Dr. **Raquel Casesnoves**, one of the members of this research project until she resigned in 2010 to become a Principal investigator of another Plan Nacional Project.

Hypothesis 1 is confirmed with all 3 methods for the discourse-pragmatic modules in all 3 languages. With **M3** (Figure 12), all **IIS** values are relatively low (the majority range between **0.6** and **0.8**), which is an expected result considering that all speakers belong to the same dielactal area**.**

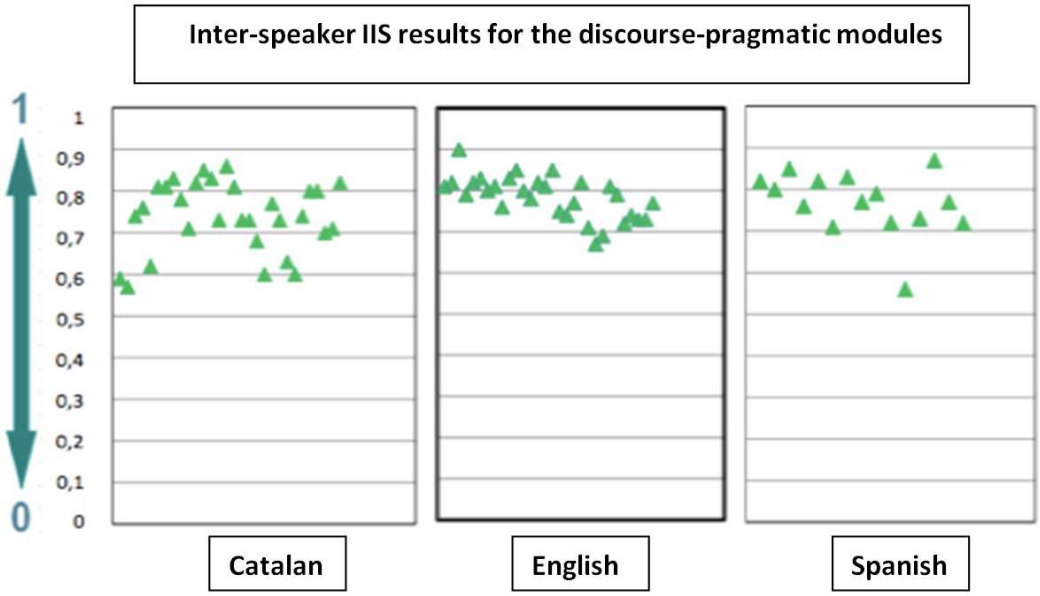

Figure 12. IIS – Discourse-pragmatic modules – Inter speaker variation

Figure 13 shows that hypothesis 2, on intra-speaker variation throughout time, is confirmed

with all 3 methods for the discourse-pragmatic modules of Catalan and English. With M3, all IIS values are quite high as expected; the majority ranging between **0.9** and **0.7.** Intra-speaker variation cannot be calculated because we could count on one measurement time only.

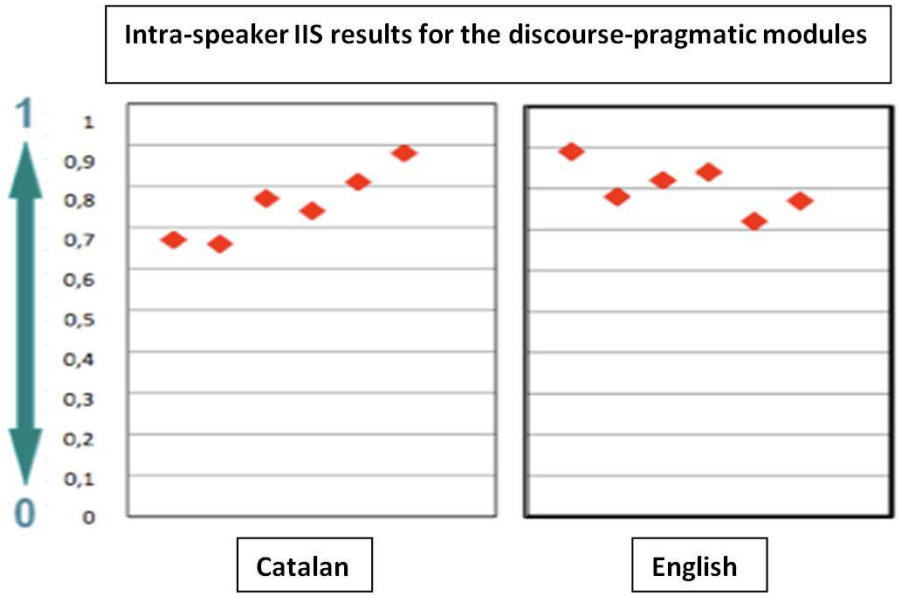

Figure 13. IIS – Discourse-pragmatic modules of Spanish – Intra speaker variation

4.4 Final conclusions

1 **Method 3** has turned out to be the most reliable method and has triggered the most robust results for both intra and inter speaker/writer variation, and in particular in the phonological modules, with some exceptions.

2 For all modules and languages, **hypothesis 1** is confirmed in that there is **more variation** and thus more idiolectal distance **between speakers and writers'** samples than between two samples of the same speaker or writer, which show a quite steady idiolectal similitude throughout time.

3 For all modules and languages, **hypothesis 2** is confirmed in that an **individual's idiolectal style** (spoken or written) **doesn't seem to vary much throughout time.**

4 **Inter-speaker/writer IIS** values may seem to be too high, but it has to be borne in mind that, except for the phonological module of Catalan, all the speakers or writers considered are from the same language variety.

References cited:

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

Table 1: Variables for the phonological module of Catalan

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

Table 1: Variables for the phonological module of Catalan

	VARIABLE	VARIANTS	EXAMPLES
16	Cambio del grupo /gz/ a [ʒ]	1. [gz] 2. [ʒ]	*examen,*
17	Presencia de /n/ en el grupo de plurales arcaicos	1. home[ns] 2. home[s]	*hómens, jóvens*
18	Sonoriación de /s/ por ultracorrección	1. [s] 2. [z]	*expressar*

Table 2: Variables for the phonoligic module of Spanish

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

Table 2: Variables for the phonoligic module of Spanish

	VARIABLE	VARIANTS	EXAMPLES
11	Pérdida de la vocal /e/ átona al interior de la palabra /entonses/ à [entons]	1. [enˈtonses]	
		2. [enˈtons]	
12	Reducción del hiato /ea/ a /a/ en la palabra /osea/	1. [oˈsa]	
		2. [oˈsea]	
13	Cambio de la palabra /entonces/ por /tons/	1. [enˈtonses]	
		2. [tons]	

Table 3: Variables for the phonological module of English

VARIABLE	VARIANTS	EXAMPLES

[Analysis and data derived from an on-going PhD. thesis]

Table 4: Variables for the morpho-syntactic module of Catalan

	VARIABLE	VARIANTS	EXAMPLES
1	Negació amb "pas"	1. pas	*"Jo que **no** hi és **pas** aquest problema de..." (VC) -RT*
		2. ø	*"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. ø	***ø** 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 clític	1. ø	*["Jo penso que ho hi ha massa problemes de convivència entre lo que és comunitat catalana-castellana" (VC) - RT*
		2. clitic	*[". . . 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*

Table 4: Variables for the morpho-syntactic module of Catalan

	VARIABLE	VARIANTS	EXAMPLES
		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. ø	*"Però fins i tot jo veig companys del meu fill **ø** 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 càmbia , 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	

Table 5: Variables for the morpho-syntactic module of Spanish

	VARIABLE	VARIANTS	EXAMPLES
1	Uso de las dos formas del pretérito imperfecto de subjuntivo	1. -ra	***a.*** *"La suerte había hecho que el pobre bizco **volara** en mi lugar;..." (RM)*
		2. -se	***b.*** *"...esa idea hizo que Julia **sonriese** para sus adentros." (AP)*

Table 5: Variables for the morpho-syntactic module of Spanish

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

Table 6: Variables for the morpho-syntactic module of English

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

Table 7: Variables for the discourse-pragmatic module of Catalan

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

Table 7: Variables for the discourse-pragmatic module of Catalan

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

Table 7: Variables for the discourse-pragmatic module of Catalan

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

Table 8: Variables for the discourse-pragmatic module of Spanish

	VARIABLE	VARIANTS	EXAMPLES
1	Respuesta a pregunta abierta con presencia de marcador discursivo	1. Presencia de marcador discursivo	*E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: **pues** por mediación de trabajo*
		2. Otra estrategia y/o ø	*E: y ¿cuáles son las fiestas de Aldaya? — J: las fiestas de Aldaya pues hay mm/ mm celebran muu- casi todos*
2	Respuesta a pregunta abierta con repetición de pregunta	1. Repetición de pregunta	*E: ¿Y tu barrio? — J: **¿mi barrio?** ¡uff! Lo mejor*
		2. Otra estrategia y/o ø	*E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: **pues** por mediación de trabajo*
3	Respuesta a pregunta abierta con exclamación	1. Exclamación	*E: imagínate que recibes una herencia inesperada — J: **¡uy!** También haría muchas cosas/ peroo// lo primero/ no sé...*
		2. Otra estrategia y/o ø	*E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: **pues** por mediación de trabajo*
4	Respuesta simple o compuesta	1. Simple	*E: bien/ háblanos de Cullera — V: **¡uy!** Cullera/ puees/ demasia(d)o ajetreo/ a mí me gusta más ..*
		2. Compuesta	*E: dinos, 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*
5	Posición del marcador discursivo pues.	1. Inicial en respuesta a pregunta abierta	*E: Y tus padres, ¿cómo vinieron a vivir aquí? — V: **pues** por mediación de trabajo*
		2. Otra: Intermedia o final	*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 ...*

Table 8: Variables for the discourse-pragmatic module of Spanish

	VARIABLE	VARIANTS	EXAMPLES
6	Marcadores discursivos ordenadores de la información	1. Simple: luego (y luego)	*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 compraar/* ***luego*** *hemos comido//* ***y luego*** *me he ido…*
		2. Compuesta: luego (y luego) pues	*J: digo pues yo que sé (risas)/ lue- yy naada/* ***y luego pues*** *las Fallas se celebran muuchoo/ yy- y las fiestas de Aldaya*
7	Uso de ¿eh? Vs. ¿no?	1. Eh?	*J. pues no sé cómo la organizaría* ***¿eh?****/ supongo quee sacaría casi todo lo que tengo en el comedoor (risas)/// pondría unos tableroos yy- yy- y haría pues no sé algo sencillo* ***¿eh?****/ algo sencillo pa- para noo formaar/ mucho follón/ y invitaríamos a la familia*
		2. No?	*J: luego yaa- ya empezaron aa- a po- ya pusieron/ la LUUZ pero me acuerdo de alumbrarnos con VEElas/ ir de vacaciones a- de niña/ a eso* ***¿no?****/ a padecer más que a- que a pasártelo bien*
8	Posició discursiva ¿eh?	1. Final de turno o absoluta (función apelativa, espera respuesta)	*V: y las fiestas pues nada en casa// preparohacemos pocas fiestas en casa ¿no?// porque casi siempre nos vamos fuera/ porque la casaa hay quee … es- estar poquito* ***¿eh?***
		2. Intermedia (final de acto) (función fática, no espera respuesta)	*V: yo era muy tranquiila/ yo ahora soy PURO nervio// me expresoo/ 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*
9	Posició discursiva ¿no?	1. Final de turno o absoluta (función apelativa, espera respuesta)	*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*
		2. Intermedia (final de acto) (función fática, no espera respuesta)	*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*
10	Intensificación morfológica vs. Léxica del enunciado	1. Uso de prefijos y cuantificadores	*J: es una niña que no tiene problemas/* ***súper*** *abierta*
		2. Uso de la repetición	*J: yo veo un niño maleducado y* ***no lo soporto/ no lo soporto***
11	Intensificación del sujeto de primera persona singular	1. 1a persona singular YO	*J:* ***yo*** *veo un niño maleducado y no lo soporto/ no lo soporto*
		2. ø	*E: ¿cómo organizarías una fiesta familiar en casa? — V: pues mira/ muy fácil/ encargo unos bocadillos…*
12	Atenuación del sujeto por impersonalización del YO	1. YO+ ø	*V: me gustaría muchísimo ir al cine/ porque me encanta el cine*
		2. Se,uno/una, 2ª p. sing	

Table 9: Variables for the discourse-pragmatic module of English

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

Table 9: Variables for the discourse-pragmatic module of English

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