



Searching the PRESEEA Corpus

The sample

The PRESEEA Corpus on this website is composed of a sample of 18 interviews from each community investigated. These interviews take the form of semi-guided conversations and come from men and women of three age groups (20-34 years old / 35-54 years old / 55 years old and over) and three different educational levels (no or primary education / secondary education / higher education) from every community.

The results of the searches made in the PRESEEA Corpus are accompanied by a code that identifies speakers according to their community of origin, sex, age and educational level, according to the following codes:

Identifying codes for each community: e.g., ALCA (Alcalá de Henares), MADR (Madrid), BOGO (Bogotá)...

Sex/gender Code: H (Male) / M (Female)

Age Code: 1: 20-34 years old / 2: 35-54 years old / 3: 55 years old or older

Educational Level Code: 1: no education or primary education / 2: intermediate education / 3: higher education.

Sample of speakers by community on the PRESEEA Corpus website

		EDUCATIONAL LEVEL		
		1	2	3
AGE	1	H / M	H / M	H / M
	2	H / M	H / M	H / M
	3	H / M	H / M	H / M

For example, ALCA_H11_037 is interpreted as follows: interview conducted in Alcalá de Henares, with a male in the first age group and with primary education; the final number indicates that this is the 37th speaker out of the 54 interviewed for the PRESEEA-Alcalá de Henares corpus.

The interviews offered here do not constitute the totality of the PRESEEA Corpus materials, but rather a representative sample of the Spanish spoken in each community extracted for research and educational purposes. For those investigations that require more spoken language materials, interested readers should request them from the coordinators of each team, since all materials are property of the teams that collected them.

The search

In CONSULTA BASICA (Basic Inquiry), you must enter the sequence of characters you wish to search for within the corpus, with the possibility of including the tags you consider appropriate.

In CONSULTA AVANZADA (Advanced Search), the orthographic form being searched for or its lemma can be specified. In the latter case, the lemma incorporates all the morphological variants of a word: for example, the lemma *perro* will allow you to retrieve the forms *perro*, *perra*, *perros*, *perras*. It is also possible to specify the grammatical category of the forms to be searched.

In FILTROS DISPONIBLES (Available Filters), the communities whose materials are being recovered and the sociolinguistic profiles of the speakers with whom the interviews were conducted can be selected. The search system allows to retrieve materials from one, several or all of the communities included in the corpus.

When queries are made, a list of numbered occurrences will be obtained, with each occurrence in concordance and aligned with the corresponding audio, along with the possibility of downloading the transcription of each interview and a 10-minute audio sample. It is also possible to download a CSV file with the result of each search.