Salvador Pons Bordería
Shima Salameh Jiménez
Sara Badia Climent

El español coloquial, en cifras

Apuntes cuantitativos sobre la estructura del español hablado

GENERALITAT VALENCIANA

MINISTERIO DE CIENCIA E INNOVACIÓN

AGENCIA ESTATAL DE INVESTIGACIÓN

Esta obra cuenta con Licencia Creative Commons vía: CC BY-NC-ND 4.0

tirant humanidades
Valencia, 2026

Índice

Prólogo: Ceci n'est pas un livre 15

Índice de abreviaturas 17

Capítulo 1.
Introducción 19
1.1. Por qué este libro 19
1.2. Las tres etapas en el estudio del español coloquial 20
1.3. Modelo Val.Es.Co 24
1.4. Segmentación de las conversaciones 26
1.5. Tratamiento de los datos 26
1.5.1. El procesamiento de las conversaciones: 27
1.5.1.1. El proceso de grabación 27
1.5.2.1. El proceso de transcripción 31
1.5.1.3. Convenciones escritas 32
1.5.1.4. El etiquetado 33
1.5.1.5. Alineación audio-video 37
1.5.1.6. Segmentación y análisis en ELAN 40
1.5.7.1. La base de datos 46
1.5.2. Composición del corpus 49
1.5.2.1. Corpus de conversaciones coloquiales actuales 49
1.5.2.2. Corpus de conversaciones de los años 90 52
1.5.3. Trabajo con los datos: recuentos y estadística 53

ANÁLISIS DE LOS DATOS
Introducción al análisis 57

Capítulo 2.
Presentación general de los datos 61
Introducción 61
2.2. Tipología de unidades 63
2.2.1. Tipos de subactos 63
2.2.2. Tipología de intervenciones 68
2.3. Unidades: duración (en segundos) 70
2.4. Unidades: extensión (en palabras) 73
2.5. Unidades: velocidad (en palabras por segundo) 75
2.6. Informantes 77

2.6.1. Sexo ... 77
2.6.2. Nivel sociocultural ... 78
2.6.3. Grupos etarios ... 79
2.6.3.1. Grupo etario 1 ... 79
2.6.3.2. Grupo etario 2 ... 80
2.7. Unidades por hablante ... 81

Capítulo 3.
Análisis de los datos por conversaciones ... 85
3.1. Introducción ... 85
3.2. Unidades por conversación ... 85
3.3. Unidades: duración (en segundos) ... 87
3.4. Unidades: extensión (en palabras) ... 92
3.5. Unidades: velocidad (en palabras por segundo) ... 94
3.6. Informantes por conversación ... 99
3.7. Unidades por hablante y por conversación ... 101

Capítulo 4.
El subacto en la conversación coloquial: descripción y caracterización ... 103
4.1. Introducción ... 103
4.2. Totales de los subactos ... 104
4.2.1. Subactos de nivel 1 y subactos de nivel 2 ... 104
4.2.2. Subactos sustantivos, adyacentes, SS/SA y elementos subestructurales ... 107
4.2.3. Subactos desglosados ... 110
4.2.4. Subactos desglosados sin residuos ... 112
4.2.4.1. Palabras más frecuentes en subactos adyacentes ... 113
4.2.4.2. Bigramas más frecuentes en SAT ... 116
4.2.4.3. Bigramas más frecuentes en SAI ... 117
4.2.4.4. Bigramas más frecuentes en SAM ... 119
4.3. Subactos por conversación ... 121
4.3.1. Subactos colapsados por conversación ... 121
4.3.2. Subactos desglosados por conversación ... 124
4.3.2.1. Tipos de subactos sustantivos vs. subactos adyacentes ... 124
4.3.2.2. Tipos de subactos adyacentes vs. subactos sustantivos ... 127
4.3.2.3. Tipos de elementos subestructurales vs. subactos sustantivos y adyacentes ... 130
4.3.3. Subactos: duración (en segundos) ... 133
4.3.3.1. Duración por tipo de subacto ... 133
4.3.3.2. Duración por subactos colapsados: sustantivos vs. adyacentes ... 137
4.3.4. Subactos: extensión (en palabras) ... 139
4.3.4.1. Extensión por tipo de subacto ... 139

4.3.4.2. Extensión por subactos colapsados: sustantivos vs. adyacentes ... 142
4.3.5. Subactos: velocidad (en palabras por segundo) ... 144
4.3.5.1. Velocidad por tipo de subacto ... 144
4.3.5.2. Velocidad por subactos colapsados ... 148
4.4. Los subactos a la luz de un árbol de decisiones ... 150

Capítulo 5.
El acto en la conversación coloquial: descripción y caracterización ... 163
5.1. Introducción ... 163
5.2. La estructura interna del acto ... 165
5.3. Tipos de subacto por posición en el acto ... 175
5.3.1. Número de subactos en la posición estructural 1 ... 180
5.3.2. Número de subactos en la posición estructural 2 ... 183
5.3.3. Número de subactos en la posición estructural 3 ... 185
5.3.4. Número de subactos en las posiciones estructurales 4 a 6 ... 186
5.3.5. Número de subactos en primera posición de intervención con los residuos colapsados ... 187
5.3.6. Subactos por posiciones ... 192
5.3.7. Número de subactos en primera posición con residuos colapsados (teniendo en cuenta la muestra de cada estructura) ... 194
5.4. Actos: puntuación de informatividad ... 196
5.5. Línea temporal *(timeline)* ... 204
5.5.1. Proporción entre silencio y lenguaje ... 207
5.5.2. Duración tipología subactos dentro del acto ... 208
5.5.3. Duración conceptual vs. procedimental ... 209
5.6. Composición de los actos: prueba *K-medias* ... 210

Capítulo 6.
La intervención en la conversación coloquial: descripción y caracterización ... 215
6.1. Introducción ... 215
6.2. Datos totales para las intervenciones ... 215
6.3. Normalización de las intervenciones ... 218
6.4. De la intervención al turno: turnos totales ... 220
6.5. Normalización de los turnos ... 222
6.6. Tipología de las intervenciones ... 223
6.7. Normalización de la tipología de las intervenciones ... 226
6.8. Propiedades de las intervenciones ... 229
6.8.1. Intervenciones: duración (en segundos) ... 229
6.8.2. Intervenciones: extensión (en palabras) ... 233
6.8.3. Intervenciones: velocidad (en palabras por segundo) ... 236

6.9. Las propiedades de las intervenciones-turno y no turno. 237
6.9.1. Intervenciones-turno vs. no turno: duración (en segundos) 237
6.9.2. Intervenciones-turno vs. no turno: extensión (en palabras) 240
6.9.3. Intervenciones-turno y no turno: velocidad (en palabras por segundo) 243
6.10. Las intervenciones a la luz de un árbol de decisiones 244

Capítulo 7.

Las conversaciones coloquiales de los años noventa: una mirada al pasado desde la intervención 249
7.1. Introducción 249
7.2. Comparación entre el total intervenciones-turno vs. no turno 250
7.3. Total por tipos de intervenciones 252
7.4. Normalización de los turnos 256
7.5. Normalización de las intervenciones 258
7.6. Las propiedades de las intervenciones de las conversaciones de los 90 261
7.6.1. Intervenciones de los 90: duración (en segundos) 261
7.6.2. Intervenciones de los 90: extensión (en palabras) 265
7.6.3. Intervenciones de los 90: velocidad (en palabras por segundo) 267
7.7. Las propiedades de las intervenciones y su estatuto como turno y no turno en las conversaciones de los 90 269
7.7.1. Intervenciones-turno vs. no turno: duración (en segundos) 269
7.7.2. Intervenciones-turno vs. no turno: extensión (en palabras) 271
7.7.3. Intervenciones-turno vs. no turno: velocidad (en palabras por segundo) 272
7.8. Las intervenciones de los años 90 a la luz de un árbol de decisiones 274

Capítulo 8.

Los rasgos sociolingüísticos de los informantes en la conversación coloquial 279
8.1. Introducción 279
8.2. Grupos etarios 280
8.3. Sexo de los informantes 281
8.3.1. Intervenciones por informante 281
8.3.2. Intervenciones: duración por sexo (en segundos) 282
8.4. Nivel social 284
8.4.1. Relación entre intervenciones y nivel social de los hablantes 284

Capítulo 9.
El estilo directo en la conversación coloquial 287
9.1. Introducción 287
9.2. La estructura interna del acto en el estilo directo 288
9.3. Tipos de subactos por posición en estilo directo 293
9.3.1. Subactos en primera posición 296
9.3.2. Subactos en segunda posición 297
9.3.3. Subactos en tercera posición 299
9.3.4. Subactos en cuarta, quinta, sexta y séptima posición 301
9.4. Actos estilo directo: puntuación de informatividad 302
9.5. Línea temporal (*timeline*) del estilo directo 310
9.6. Intervenciones y estilo directo 311
9.6.1. Intervenciones normalizadas 312
9.6.2. Intervenciones estilo directo: duración (en segundos) 316
9.6.3. Intervenciones estilo directo: extensión (en palabras) 318
9.6.4. Intervenciones estilo directo: velocidad (en palabras por segundo) 320
9.7. Composición de los actos en estilo directo: prueba de K-medias 322
9.8. El estilo directo a la luz de un árbol de decisiones 325
9.8.1. Subactos 326
9.8.2. Intervenciones 329

Referencias bibliográficas 335

Anexos
1. Signos y convenciones de transcripción del grupo Val.Es.Co. 341
2. Notación para la segmentación en unidades discursivas 344
3. Glosario de términos estadísticos 345
4. Palabras que componen subactos adyacentes 349
5. Árboles de decisiones de subactos 360
6. Rasgos prosódicos en la conversación coloquial: el tonema 365
6.1. Introducción 365
6.2. Frecuencia de los tipos de tonema por conversación normalizados 365
7. Datos descartados en los análisis: el análisis sociolingüístico 367
7.1. Grupos etarios 368
7.1.1. Intervenciones por hablante 368
7.1.2. Duración intervenciones por grupo etario (en segundos) 369
7.1.3. Extensión intervenciones por grupo etario (en palabras) 370
7.1.4. Velocidad intervenciones por grupo etario (en palabras por segundo) 371
7.2. Sexo 372
7.2.1. Total de intervenciones 372
7.2.2. Duración intervenciones por nivel social e informante (en segundos) 373

7.2.3. Extensión de las intervenciones: hombres vs. mujeres (en palabras) ... 374
7.2.4. Velocidad intervenciones: hombres vs. mujeres (palabras por segundo) ... 374
7.3. Nivel educativo ... 375
7.3.1. Intervenciones por nivel educativo ... 375
7.3.2. Duración intervenciones por nivel educativo (en segundos) ... 376
7.3.3. Extensión intervenciones por nivel educativo (en palabras) ... 377
7.3.4. Velocidad intervenciones por nivel educativo (en palabras por segundos) ... 377
7.4. Cruce de características de los informantes: sexo por edad ... 378
7.4.1. Velocidad de subactos ... 378
7.4.2. Velocidad de los actos ... 379
7.4.3. Velocidad de intervenciones ... 380
7.5. Cruce de características de los informantes: edad por nivel educativo ... 382
7.5.1. Velocidad de subactos ... 382
7.5.2. Velocidad de los actos ... 383
7.5.3. Velocidad de las intervenciones ... 384
7.6. Cruce de características de los informantes: edad por clase social ... 386
7.6.1. Velocidad de subactos ... 386
7.6.2. Velocidad de los actos ... 387
7.6.3. Velocidad de las intervenciones ... 388

Prólogo: Ceci n'est pas un livre

Estimado lector:

El libro que tienes en la mano se presenta como tal porque, según la RAE, para entrar dentro de dicha definición solo ha de cumplir con dos características fácilmente atingibles: "[tener] muchas hojas de papel" y "formar un volumen". Y este, a fe que las cumple. Así, será computado como libro por el Ministerio correspondiente cuando tenga que calcular nuestros méritos investigadores, servirá como mérito para las oposiciones pertinentes y, en fin, cubrirá su hueco dentro del *cursus honorum* universitario.

Sin embargo, todo libro viene con una promesa de legibilidad que no estamos seguros de que esta obra cumpla de forma adecuada, aun descontando a dicha cualidad el plus de indegestibilidad propio de toda literatura científica. Las más de ciento cincuenta tablas y otras tantas gráficas que habitan sus páginas requieren de comentarios específicos, repetidos en ocasiones, pero necesarios para comprender el dato concreto al que se refieren. Los criterios de normalización o la descripción de las pruebas estadísticas deben volverse a explicar, aun con su proporción de *variatio*, para no quedar enterrados en una página ignota. Y, aunque se incluya un pequeño glosario con los términos más técnicos, su introducción se asocia siempre a una pequeña explicación que, con toda seguridad, resultará superflua para los expertos y excesivamente escueta para los no iniciados. Estas y otras características de la publicación pueden convertir la lectura ordenada y sistemática de estos *Apuntes...* en una experiencia poco ligera y hasta exasperante: lo sabemos porque lo hemos probado –repetidas veces– en el proceso de escritura. *Caveat lector!*

En realidad, el trabajo que presentamos a la comunidad científica se asemeja más a un anuario ("libro [...] con información, direcciones y otros datos de actualidad") o a un glosario ("catálogo de palabras de una misma disciplina [...] definidas o comentadas"), ya que contiene datos –muchos– y comentarios –también muchos– y se ofrece como un catálogo de consulta al que el usuario puede recurrir cuando trate de relacionar los datos de sus investigaciones con los datos de la normalidad procesados en esta muestra. Porque, no nos olvidemos, el regis-

tro coloquial, en su variedad conversacional, es el género no marcado, la opción por defecto, el lugar al que volvemos, el refugio seguro de todo hablante, y su descripción y cuantificación permite un término de comparación con el resto de las variedades, registros o géneros de un idioma.

Este, pues, es un primer, incompleto, diminuto y parcial intento de descripción global y sistemática que, esperemos, venga acompañado de estudios posteriores, más concienzudos, realizados por otros investigadores, tal vez en un futuro (cercano). Desde este presente hacia el futuro, *valete collegae!*