Loquens 3(1)
January 2016, e029
eISSN 2386-2637
doi: http://dx.doi.org/10.3989/loquens.2016.029

Cálculo de frecuencias de aparición de fonemas y alófonos en español actual utilizando un transcriptor automático

Iván Arias Rodríguez

Universidad Complutense de Madrid
ivan.arias.rodriguez@gmail.com

Recibido: 10/11/2015. Aceptado: 15/01/2016. Publicado on line: 09/01/2017

Citation / Cómo citar este artículo: Arias Rodríguez, I. (2016). Cálculo de frecuencias de aparición de fonemas y alófonos en español actual utilizando un transcriptor automático. *Loquens, 3*(1), e029. doi: http://dx.doi.org/10.3989/loquens.2016.029

RESUMEN: El cálculo de la frecuencia de aparición de los distintos fonemas de la lengua española es un asunto que se ha abordado previamente en varias ocasiones. Sin embargo, la gran mayoría de los estudios realizados no toma en consideración la variación alofónica o la trata solo parcialmente. Además, en todos estos trabajos los recuentos se han llevado a cabo a partir de palabras aisladas, sin tener en cuenta los fenómenos fonéticos que se producen en la secuencia fónica, especialmente el resilabeo.

En el presente trabajo se realiza un recuento de los porcentajes de aparición de los distintos fonemas y alófonos en el dialecto castellano. Para ello, se ha creado un corpus de novelas en español, que se transcriben automáticamente con ayuda de un programa informático creado *ex profeso*, cuyo diseño se detalla en este artículo. También se realiza un estudio limitado de la estructura silábica en términos de frecuencias de aparición.

Palabras clave: frecuencia de aparición; fonemas; alófonos; español; cadena hablada.

ABSTRACT: *Frequency of occurrence of phonemes and allophones in contemporary Spanish as calculated by an automatic transcription system.*– The frequency of occurrence of the different phonemes of the Spanish language has been the subject of several previous studies. However, most of those studies did not take into account the frequency of the allophones, or they did it only partially (many times there was not even a phonetic transcription, and the plain orthographic transcription was used instead). Moreover, in all those previous works the frequency of occurrence was calculated from the transcription of isolated words, without taking into account the phonetic changes produced by its insertion in the speech chain, especially that of resyllabification.

The present article calculates the frequency of occurrence of the phonemes and allophones of the Castilian Spanish dialect, as they are pronounced in the spoken language. The transcription of a corpus (consisting in 560 novels of modern Spanish writers) is done automatically thanks to a piece of software implemented for the purpose of this study. The article details the contents of the corpus as well as a detailed description of the design of the automatic transcriber. Finally, there is also a limited study of the syllable structure in terms of its type and frequency of occurrence.

Keywords: frequency of occurrence; phonemes; allophones; Spanish; speech chain.

1. INTRODUCCIÓN

El objetivo de este trabajo ha sido doble:

- Por un lado, crear un transcriptor que calcule automáticamente la transcripción fonética estrecha de un corpus de español actual (en su forma ortográfica normativa).
- Por otro, y como objetivo final, realizar un recuento estadístico de las apariciones de los distintos fonemas y alófonos, así como de las estructuras silábicas más frecuentes, del español, en su variante dialectal castellana, del centro-norte peninsular.

El transcriptor fonético que se ha implementado tiene en cuenta los cambios fonéticos y alofónicos que se dan

en la cadena hablada, principalmente debidos al resilabeo. Un transcriptor de este tipo puede ser útil, no solo para la tarea que se expone en este trabajo, sino como primera etapa de un conversor de texto en habla, o como parte de un transcriptor de habla automático.

Los estudios llevados a cabo en relación con la frecuencia de aparición de fonemas (y de ciertas estructuras silábicas) normalmente no contemplan el plano alofónico, lo que no permite hacer recuentos de alófonos ni comprobar cuáles son las realizaciones más frecuentes de los diversos fonemas.

El análisis de las frecuencias silábicas se ha abordado en menor medida, según se desprende de la bibliografía, y sus resultados pueden resultar también relevantes para el campo de la adquisición de la lengua materna o del aprendizaje de L2. Aunque queda fuera del ámbito de este trabajo, sería interesante estudiar si existe alguna correlación entre las sílabas o los alófonos que aparecen más a menudo en la cadena hablada y la facilidad o la velocidad con la que se adquieren o se aprenden.

2. ANTECEDENTES

Los trabajos previos que han realizado algún tipo de recuento de frecuencias de aparición de fonemas aparecen recopilados en la Tabla 1.

De estas obras, solo la de Moreno Sandoval, Torre Toledano, Curto y de la Torre (2006) estudia la estructura de la sílaba en español y su frecuencia de aparición. Otros autores han tratado este aspecto, como Guerra (1983). No obstante, su trabajo tiene una perspectiva diacrónica, con un corpus formado por 125 000 palabras pertenecientes a fragmentos de obras literarias de los siglos xiii al xx. También trataron el tema Álvarez, Carreiras y de Vega (1992), en un estudio que utilizó 25 000 palabras. En este caso, la transcripción utilizada en el análisis fue ortográfica y no fonética.

En un artículo posterior a los últimos citados, Alameda y Cuetos (1995) utilizaron un corpus de 2 000 000 de palabras para calcular su frecuencia de aparición, además de la de sus sílabas, bigramas y letras. Se consideraron las palabras de forma aislada y, de nuevo, se utilizó únicamente la transcripción ortográfica, como se hace también en la base de datos LEXESP (Sebastián-Gallés, Martí, Carreiras y Cuetos, 2000), que analiza 5 000 000 de palabras en términos de sílabas y de vecindad ortográfica.

En cualquier caso, el tamaño del corpus utilizado en estudios previos es relativamente pequeño. Otra característica importante de dichos trabajos es que todos ellos tratan las palabras aisladamente, sin tener en cuenta los efectos que produce su inserción en la cadena hablada.

Apenas se ha abordado en la bibliografía el asunto de la realización alofónica de cada fonema. Habitualmente, las transcripciones utilizadas solo se ciñen al nivel del fonema —salvo los trabajos de Llisterri y Mariño (1993) y Pineda, Villaseñor, Cuétara, Castellanos y López (2004), que llevan a cabo un estudio alofónico parcial—. Además, la práctica totalidad de los trabajos previos agrupa las vocales cerradas junto con sus respectivas variantes semiconsonantes y semivocales (esto es, las conocidas actualmente como 'paravocales'). Sus resultados se muestran en la Tabla A1 y en la Tabla A2.

En los trabajos de este tipo realizados a partir de los años 90 del pasado siglo se ha utilizado siempre algún

Tabla 1. Trabajos previos sobre la frecuencia de los fonemas del español, precisando el tipo y tamaño del corpus, y la variedad dialectal utilizada en la transcripción.

Autores	Tipo de corpus	Tamaño del corpus	Variedad dialectal
Zipf y Rogers (1939)	escrito	5 000 fonemas	castellana
Navarro Tomás (1966)	escrito	20 000 fonemas	castellana
Alarcos Llorach (1965)	escrito	25 cartas	castellana
Delattre (1965)	oral	(sin datos)	castellana
Lloyd y Schnitzer (1967)	escrito (listado)	252 404 sílabas	castellana
Guirao y Borzone (1972)	escrito y oral	62 980 fonemas	rioplatense
Quilis y Esgueva (1980)	oral	160 000 fonemas	castellana
Mosterín (1981)	escrito	58 620 fonemas	castellana
Rojo Sánchez (1991)	escrito	3 641 915 fonemas	castellana/hispanoam.
Guirao y García (1993)	oral	74 460 sílabas	rioplatense
Llisterri y Mariño (1993)	oral	> 100 000 fonemas	castellana
Pérez (2003)	oral	75 269 fonemas	chilena
Pineda *et al.* (2004)	oral	6 000 oraciones	mexicana
Moreno Sandoval *et al.* (2006)	escrito y oral	1 244 411 fonemas	castellana
González Rátiva y Mejía Escobar (2011)	escrito	5 682 417 fonemas	colombiana

tipo de herramienta informática para producir automáticamente la transcripción fonológica de un texto a partir de su transcripción ortográfica. La creación de transcriptores fonéticos/fonológicos en español se ha tratado en la bibliografía de forma independiente, y se han identificado multitud de problemas de difícil solución. El artículo de Ríos (1993) muestra que no es posible obtener una transcripción fonética exacta partiendo únicamente de la representación ortográfica, sino que se necesita información lingüística acerca de cada palabra en cuestión para poder desambiguar si existen ciertos diptongos o hiatos, o para asignar los acentos secundarios.

En otro artículo de Bonaventura, Giuliani, Garrido y Ortín (1998) se presenta un sistema basado en reglas que realiza una transcripción fonética automática, si bien las limitaciones informáticas del momento en que se publicó dificultaban el procesado correcto de las vocales con tilde. También es destacable un trabajo de Castro, España, Salvador y Marzalz (2001) que utiliza el lenguaje de programación Python,[1] aunque solo realiza una transcripción fonológica y no utiliza los símbolos del Alfabeto Fonético Internacional (AFI).

Posiblemente el transcriptor fonético para el español más completo creado hasta el momento sea TexAFon (Garrido, Laplaza, Marquina, Schoenfelder y Rustullet, 2012). Esta herramienta funciona tanto para el español como para el catalán, y además de hacer una transcripción fonética, analiza la modalidad de la oración y otros elementos de carácter entonativo. También implementado en Python, transcribe los efectos fonéticos que se producen como resultado de los contactos entre palabras, pero sin realizar el resilabeo.

En cualquier caso, un transcriptor fonético basado únicamente en reglas nunca puede ser perfecto, puesto que existen vocablos cuya pronunciación no se atiene a dichas reglas (especialmente los préstamos de otras lenguas). Para remediar estas carencias, se puede crear un diccionario de excepciones, que, no obstante, nunca abarcará todas, puesto que, para ello, debería estar en permanente actualización.

Con todo, conviene valorar el hecho de que las transcripciones realizadas de forma automática tienen la ventaja de que no presentan los inevitables errores humanos. Además, un transcriptor automático sigue siempre el mismo criterio de transcripción, con lo que sus transcripciones pueden llegar a ser incluso más correctas que las realizadas manualmente, especialmente si el corpus es grande y el trabajo se divide entre varias personas.

3. METODOLOGÍA DE TRABAJO

Para poder llevar a cabo este trabajo se necesitan los siguientes tres elementos:

- Un corpus de texto en español que sea adecuado para su transcripción y análisis.
- Un transcriptor fonético/fonológico automático que transcriba a partir de la representación ortográfica del corpus.
- Un sistema que elabore recuentos y calcule estadísticas a partir de la transcripción fonética.

Para diseñar el transcriptor fonético, se ha utilizado una arquitectura en capas que calcula la transcripción final pasando por una sucesión de etapas intermedias. Dichas etapas son las siguientes:

- Segmentado del texto en oraciones.
- Segmentado de las oraciones en palabras y expansión ortográfica de caracteres no alfabéticos.
- Discriminación entre palabras tónicas y átonas.
- Segmentado de la palabra en grafemas y desambiguación de su valor fonológico.
- Silabeado de los fonemas de la palabra y asignación de tonicidad a las sílabas.
- Cálculo de los alófonos que corresponden a los fonemas de la palabra, según el entorno fonético de la palabra aislada.
- Resilabeo de las palabras en contacto.

En los apartados siguientes se detallan todos estos puntos.

3.1. El corpus de texto de español contemporáneo

Para hacer un recuento fonético de la secuencia fónica, lo ideal sería contar con un gran corpus de transcripciones reales de muestras orales, principalmente espontáneas, pero los corpus disponibles tienen un tamaño muy limitado.

Una segunda opción sería realizar la transcripción fonética de algún corpus oral con transliteraciones de audios de conversaciones espontáneas. Sin embargo, vuelve a surgir la misma limitación, ya que los corpus orales espontáneos son muy escasos y de muy reducido tamaño: el *Corpus Oral de Referencia de la Lengua Española Contemporánea* (CORLEC; Marcos Marín, 1992), de 1 100 000 palabras, es el más amplio.

Como se pretende hacer un recuento de un corpus bastante mayor, se ha optado por crear un corpus específico para este recuento, utilizando una recopilación de libros (en su inmensa mayoría, novelas, pero también ensayos y artículos) de escritores españoles. Aunque el dialecto del español que hable el escritor no influye en la transcripción fonética realizada, se ha evitado en general incluir autores hispanoamericanos por las diferencias en expresiones, léxico o giros lingüísticos que pudieran darse con respecto al español del centro-norte peninsular.

El uso de este corpus tiene la ventaja de que podrá ser transcrito automáticamente con facilidad, puesto que se ajusta convenientemente a las normas ortográficas del espa-

[1] Python es un software estándar, disponible en https://www.python.org.

ñol. Al tratarse, además, de textos originalmente escritos por autores de lengua española (y no de traducciones), se minimiza la aparición de palabras extranjeras, principalmente topónimos o antropónimos, que presentan problemas para su transcripción fonética y para la asignación de tonicidad.

Por último, puesto que el objetivo del trabajo es el análisis del español contemporáneo de España, se requiere que el texto se haya escrito en las últimas décadas, a ser posible por un autor aún vivo o fallecido en el siglo XXI.

Así pues, se han obtenido las versiones en formato electrónico de 560 obras de 63 autores y se han convertido a formato de texto sin formato (con codificación de caracteres UTF8) utilizando el programa de libre distribución Calibre (Versión 2.53; Goyal, 2016). El corpus así creado ocupa 299 MB de memoria. En la Tabla A3 se listan las 560 obras incluidas en él.

Tras analizarlo, se comprobó que el tamaño del corpus, medido en distintas unidades, es el siguiente:

- 301 685 816 caracteres;
- 236 024 884 segmentos (tras resilabeo 233 326 472);
- 104 439 849 sílabas (tras resilabeo 100 674 739);
- 52 661 222 palabras;
- 3 699 383 oraciones.

El corpus utilizado en este trabajo es el corpus en español de mayor tamaño analizado fonéticamente del que se tenga constancia. El precio que se ha de pagar por esta ventaja es el hecho de que este corpus pueda no ser la mejor representación del registro oral. En el artículo de Moreno *et al.* (2008) se muestra que los porcentajes de aparición de los distintos fonemas y sílabas del español varían según sea el tipo del corpus utilizado (oral o escrito). No obstante, el mencionado estudio no tiene en cuenta el resilabeo (se transcriben automáticamente las palabras aisladas, incluso en el corpus oral), lo que altera en gran medida los valores obtenidos, especialmente por lo que se refiere al recuento de los tipos de sílaba más frecuentes.

3.2. El transcriptor fonético automático

A grandes rasgos, el transcriptor creado para realizar este trabajo tiene un diseño y una estructura interna similares a los del comentado TexAFon (Garrido *et al.*, 2012), pero añade funcionalidad extra. También se ha implementado en Python, aunque no se ha utilizado de ningún modo código de TexAFon.

El transcriptor toma la representación ortográfica del texto y crea una transcripción fonética estrecha siguiendo las directrices que se indican en el manual de Fernández Planas y Carrera Sabaté (2001). En dicho manual se presenta un modelo de castellano estándar (variante centro-norte peninsular) en un registro cuidado de habla sin llegar a adoptar un nivel enfático.

El repertorio fonológico de la variante dialectal escogida (castellana) aparece en la Tabla 2 y la Tabla 3. Consta de 28 fonemas: 19 consonánticos, cinco vocálicos y cuatro paravocales (semiconsonantes y semivocales). Es importante notar que se han diferenciado expresamente las cuatro paravocales como fonemas independientes, resultando en un sistema vocálico con nueve fonemas. Esta postura no es la más extendida entre los fonólogos del español, que suelen postular un sistema con cinco vocales, en las que las paravocales son variantes alofónicas de las vocales cerradas; o a lo sumo, añaden a estas cinco vocales dos fonemas paravocálicos, sin diferenciar su uso como semiconsonante o semivocal. No se pretende con este trabajo abogar por un sistema vocálico de nueve fo-

Tabla 2. Repertorio de fonemas consonánticos utilizados en la transcripción, y sus realizaciones alofónicas en ataque y coda silábica.

Fonema	Alófonos de ataque	Alófonos de coda
/m/	[m]	[m ɱ n̪ n̺ n ⁿ n̻ ⁿ̻ ŋ ɴ][2]
/n/	[n nː]	[m ɱ n̪ n̺ n ⁿ n̻ ⁿ̻ ŋ ɴ]
/ɲ/	[ɲ]	–
/b/	[b β̪ ᵝ̪ β̪ː][3]	[β̪ ᵝ̪]
/p/	[p pː]	[β̪]
/d/	[d̪ ð̪ ᵟ̪ ð̪ː]	[ð̪ ᵟ̪]
/t/	[t̪ t̺ tː]	[ð̞]
/g/	[g ɣ ˠ ɣː]	[ɣ ˠ]
/k/	[k k̟ k̺ kː]	[ɣ]
/ʧ/[4]	[ʧ ʧː]	[ʧ]
/f/	[f fː]	[f f̟]
/θ/	[θ θː]	[θ θ̟]
/s/	[s sː]	[s s̪ s̺ z̪][5]
/x/	[x ˣ χ xː χː]	[ˣ]
/ʝ/	[ʝ̟ j͡ʝ̟][6]	–
/l/	[l lː]	[l l̪ l̺ l̻][7]
/ʎ/	[ʎ]	–
/r/	[r rː]	–
/ɾ/	[ɾ]	[ɾ]

[2] Se ha optado por utilizar la notación [n̪] frente a [nʲ] por simplicidad y para poder utilizar la forma [ⁿ̻].
[3] Los alófonos [ᵝ̪ ᵟ̪ ˠ ˣ] son alófonos de coda que pueden llegar a aparecer en ataque por efecto del resilabeo.
[4] Se prefiere la forma /ʧ/ frente a /t͡ʃ/ por simplicidad.
[5] Se utiliza la notación [z̪] en vez de la forma, quizá más correcta, [s̬] para evitar el uso de dos diacríticos.
[6] Se sigue la propuesta de transcripción de Martínez Celdrán y Fernández Planas (2000) para la africada palatal sonora: [j͡ʝ̟].
[7] Como ocurre con la nasal, se utiliza la notación [l] frente a [lʲ] por simplicidad.

nemas. Sin embargo, se ha optado por él únicamente porque presenta ventajas, en términos de computación, con respecto a las opciones de cinco o siete fonemas: facilita el recuento final, acelera el cálculo de los alófonos (que, en cualquier caso, coinciden en los tres sistemas fonológicos comentados) y simplifica tanto el proceso de asignación de grafemas a fonemas, como la tarea de extracción de la estructura silábica de la palabra y el cálculo del efecto del resilabeo. En cualquier caso, para obtener los datos de los recuentos de fonemas suponiendo un sistema de cinco fonemas vocálicos, o de dos paravocales más cinco vocales, basta con sumar los porcentajes de aparición de los cuatro fonemas paravocálicos que se consideran en este trabajo a los del fonema que corresponda.

Estos 28 fonemas se realizan con un total de 106 alófonos distintos, de los que 62 son consonánticos, 36 vocálicos y 8 paravocálicos. De los 62 alófonos consonánticos, 31 aparecen en coda y 46 en ataque (hay 15 que aparecen en ambas posiciones). De los 46 alófonos consonánticos de ataque, 31 son simples y 15 son geminados.

En cuanto a los 36 alófonos vocálicos, se cuenta con 20 alófonos orales (10 de duración normal y otros 10 largos) y 16 alófonos nasalizados (ocho de cada duración). De los ocho alófonos de paravocales considerados, cuatro son orales y otros cuatro están nasalizados. Por último, se consideran como fonemas las pausas: largas en el caso del punto, los puntos suspensivos y los signos de exclamación o interrogación; y cortas en el caso de los demás signos de puntuación.

Por otra parte, también se transcribe la tonicidad de la sílaba. Se marcan los acentos primarios de las palabras tónicas, además de los acentos secundarios de los adverbios acabados en <-mente>, pero no los de las palabras compuestas. En el caso de las palabras átonas, no se marca ninguna sílaba como tónica.

Tabla 3. Repertorio de fonemas vocálicos (y paravocálicos), y de sus alófonos en entornos orales y nasales.

Fonema	Alófonos orales	Alófonos nasales
/j/	[j][8]	[j̃]
/i/	[i i̯ iː i̯ː]	[ĩ ĩ̯ ĩː ĩ̯ː]
/i̯/	[i̯]	[ĩ̯]
/e/	[e ɛ eː ɛː]	[ẽ ẽː]
/a/	[a ɑ aː ɑː]	[ã ãː]
/o/	[o ɔ oː ɔː]	[õ ɔ̃ õː ɔ̃ː]
/w/	[w][8]	[w̃]
/u/	[u u̯ uː u̯ː]	[ũ ũ̯ ũː ũ̯ː]
/u̯/	[u̯]	[ũ̯]

Como se comentó, la arquitectura del sistema está compuesta por niveles jerárquicos colocados en capas, que se implementan como clases independientes en Python. Así, cada clase de nivel superior incluye uno o más elementos de las clases inferiores. Se han implementado siete clases, de las que las siguientes cuatro representan la estructura del texto:

- *Texto:* segmenta el texto de entrada en oraciones, y genera las estadísticas de aparición al acabar la transcripción.
- *Oración:* divide las oraciones en palabras y las expande en su forma ortográfica si contienen caracteres no alfabéticos. También realiza el resilabeo final.
- *Palabra:* silabea las palabras y calcula fonemas y alófonos.
- *Grafema:* representa los caracteres de entrada y sus posibles correspondencias fonológicas.

Las otras tres clases se encargan de la transcripción propiamente dicha:

- *Sílaba:* es una estructura auxiliar que representa una sílaba, su tonicidad y sus fonemas/alófonos.
- *Fonema:* representa a un fonema, con sus características y reglas de realización según su entorno fonético.
- *Alófono:* representa a los alófonos de forma similar a la clase *Fonema*.

Este diseño hace más fácil la división en subtareas y la corrección de errores, que usualmente son atribuibles a un nivel en concreto según el tipo de fallo.

El transcriptor transcribe de media unos 14 000 segmentos por segundo. Aunque no se han encontrado datos de velocidad equivalentes para otros sistemas de transcripción, se considera que se ha conseguido un diseño eficiente teniendo en cuenta las limitaciones del hardware utilizado.[10] Esta velocidad permite que el corpus se analice por completo en menos de cinco horas.

En las siguientes secciones se presentan las clases utilizadas y se explica el funcionamiento del transcriptor.

3.2.1. La clase Texto

Esta clase, que inicia el proceso, toma como entrada el texto total que se ha de transcribir y utiliza expresiones regulares para segmentarlo en oraciones, siguiendo un algoritmo basado en el método enunciado por Grefenstette y Tapanainen (1994).

[8] Las realizaciones [j j͡j̯] se consideran alófonos del fonema consonántico /j/.
[9] La realización de /w/ como [u̯w] o [gw] se considera como una inserción del fonema /g/.
[10] Un ordenador portátil Lenovo IdeaPad Z510, con un procesador Intel Core i7-4702MQ a 2,2 GHz y una memoria RAM de 16 Gb DDR3L a 1333 MHz.

Para ello, se identifican los caracteres segmentadores de oración, tanto ambiguos[11] como no ambiguos[12]. Estos caracteres se incluyen al inicio o al final de la oración y representan pausas.

3.2.2. La clase Oración

Los objetos de la clase *Oración* reciben el texto de una oración como entrada y lo segmentan en palabras. De nuevo, se sigue el algoritmo de Grefenstette y Tapanainen (1994) (v. unas líneas más arriba) y se usan expresiones regulares para segmentar por caracteres blancos (espacios, tabulaciones…), comas (siempre que no sean parte de un número), barras (si no son parte de una fecha) y algunos caracteres especiales como <°ºª&>.

Posteriormente, se identifica cuáles de estas palabras son números (cardinales y ordinales), porcentajes, grados, horas, fechas o números romanos, en cuyo caso la cadena de caracteres se expande en su representación ortográfica. Así, un texto como <27,3º a las 15:30 con un 78 % de humedad en el sector XII> se transforma en <veintisiete coma tres grados a las tres y media con un setenta y ocho por ciento de humedad en el sector doce>.

También se detectan acrónimos (que se deletrean si no pueden silabearse), combinaciones de caracteres alfabéticos y numéricos (que se deletrean carácter a carácter, salvo los números, que se transforman en su representación ortográfica agrupando dígitos consecutivos), además de otros caracteres no alfanuméricos (que también se deletrean) y algunas otras combinaciones de caracteres especiales comunes. Cuando se deletrea una cadena de caracteres, los nombres de las vocales y de algunas consonantes como <d> y <t> se transcriben con tilde (<á>, <dé>, <té>) aunque sean monosílabos, para que más adelante, durante el procesado, no se confundan con palabras átonas. Además, se considera que las letras se pronuncian como en una enumeración, separadas por pausas cortas.

Como ejemplo, el texto de entrada <Mª Victoria, la mexicana del 5º, es la socia nº 27 401 de CC.OO., y trabaja en la O.N.U. Su NIF es 56.765-396DW, y cobra 5.000$. Pues tiene 38,7ºC de fiebre> se convierte en <María Victoria, la mejicana del quinto, es la socia número veintisiete mil cuatrocientos uno de ce, ce, ó, ó, y trabaja en la onu. Su nif es cincuenta y seis mil setecientos sesenta y cinco, guion, trescientos noventa y seis, dé, uve doble y cobra cinco mil dólares. Pues tiene treinta y ocho coma siete grados centígrados de fiebre>.

Tras conseguir una transcripción ortográfica segmentada por palabras, y que únicamente contiene caracteres alfabéticos —y signos de puntuación en su inicio o final—, se crea un objeto de la clase *Palabra* con cada una de ellas, que realiza el siguiente paso del proceso.

3.2.3. La clase Palabra

Al crear un objeto de la clase *Palabra,* se apartan los posibles signos de puntuación que haya en el texto, y se verifica si se trata de una palabra tónica. Para ello el vocablo en cuestión se compara con una pequeña base de datos en la que se incluyen las palabras átonas del español tal y como especifica la Real Academia Española y la Asociación de Academias de la Lengua Española (2005) o en artículos como el de Pérez Tobarra (2005). Si la palabra no es átona, se asignará tonicidad a una de sus sílabas tras haberla silabeado.

Tras asignar la tonicidad, el objeto *Palabra* divide el texto de entrada en grafemas y crea un objeto de la clase *Grafema* para cada uno.

3.2.4. Las clases Grafema y Fonema

Para identificar los grafemas, se procesa la palabra de izquierda a derecha, intentando primero encontrar dígrafos, que pueden ser tanto los compuestos tradicionales <ch>, <ll> y <rr>[13], como las agrupaciones puntuales <gu> y <qu> (seguidos de vocal anterior). Además, se han añadido los dígrafos propios de la lengua vasca <tx> y <tz> puesto que en el corpus aparecen en ocasiones en préstamos (principalmente antropónimos) con el mismo valor fonológico que <ch>, como ocurre con <Etxenike> o <ertzaina>.

Si no se encuentra un dígrafo, se compara el carácter con los monógrafos que el sistema es capaz de reconocer. Además de los grafemas propios del español y de los dígrafos vascos a los que se acaba de aludir, también se ha incluido <ç> como grafema aceptable con el mismo valor fonológico que <s>. Esto se debe a que su uso está bastante extendido en algunas palabras catalanas que se integran en textos en español, como pueden ser <Barça> o <calçot>.

Si el carácter individual es ajeno a la lengua española, se descarta (se considera 'mudo'). Si el carácter sí pertenece al alfabeto español pero contiene algún diacrítico que no es propio del español o no es combinable con dicho carácter (como ocurre en <València>, <Citroën> o <João>), se eliminan los diacríticos del carácter (dándonos en este caso las *españolizaciones* <Citroen>, <Valencia> y <Joao>).

Por último, en caso de que la palabra comience por un ataque compuesto, con una <s> al inicio, se añade la próte-

[11] Caracteres ambiguos son el punto, que puede no segmentar oraciones si es parte de un acrónimo, de una abreviatura o de un número; y los dos puntos, que pueden ser parte de una hora.

[12] Esto es, los saltos de línea, además de los caracteres <¿?¡!()[]{}…;>.

[13] Aunque los ejemplos se dan siempre en minúsculas, en realidad, cuando se trata con caracteres (incluso dígrafos) es indiferente si están en mayúsculas o minúsculas o combinados de alguna manera.

sis <e> antes de realizar la transcripción. Este tipo de combinación consonántica, que no es posible según la fonotáctica del español, es bastante habitual en términos extranjeros (sobre todo nombres propios), y los hablantes suelen pronunciarla, de hecho, con la mencionada prótesis.

Así, se crea un objeto de la clase *Grafema* por cada grafema que contenga la palabra. Cada uno, a su vez, contiene una lista de los posibles fonemas que puede representar, junto con restricciones acerca del entorno fonético necesario para que el grafema los represente.

Tales fonemas pueden ser de seis tipos: consonántico, semiconsonántico, vocálico, semivocálico, mudo o de pausa. La mayor parte de los grafemas consonánticos tiene fácil desambiguación, pues representan siempre un único fonema consonántico, a excepción de:

- <g>, <c>, que son fricativos ante vocal anterior, u oclusivos en los demás casos.
- <r>, que es vibrante múltiple en inicio de palabra y simple o percusiva en los demás casos.
- <x>, que puede representar el fonema /s/, el grupo /ks/ o el fonema /k/ según su entorno fonético ([eʂ.ʈe.ˈɾjɔɾ] <exterior>, [ɛɣ̟.ˈsa.mẽn] <examen>, [ˈɛɣ̟.reɪ̯] <exrey>[14]).

Los grafemas vocálicos (con o sin tilde) también representan habitualmente un único fonema vocálico a excepción de las vocales <i> y <u>, que pueden tener valor vocálico (<pinta> [ˈpi̯n.ta], <punta> [ˈpu̯n.ta]), semivocálico (<faina> [ˈfai̯.na], <fauna> [ˈfau̯.na]), semiconsonántico (<pues> [pwes], <pies> [ˈpjes]) o consonántico (<hiela> [ˈʝ̟e.la], <huela> [ˈɡwe.la]).

Por otra parte, la <y> puede tener valor vocálico (<muy> [mwi̯]), semivocálico (<hay> [ˈai̯]) o consonántico (<yaya> [ˈʝ̟a.ʝ̟a]). La <ü> es siempre semiconsonante en español (por ejemplo, en <pingüe>), aunque el transcriptor acepta también su interpretación como elemento plenamente vocálico para poder transcribir algunas pocas palabras extranjeras que aparecen en el corpus (como <führer> [fu.ˈɾɛɾ]).

Hay que hacer notar que el transcriptor decide que un carácter vocálico es vocal, paravocal (semivocal o semiconsonante) o consonante utilizando únicamente el contexto fonético. Esto origina que se consideren erróneamente como diptongos algunas combinaciones de caracteres vocálicos que ortográficamente sí que son diptongos, pero que en la lengua oral se realizan como hiatos, al menos en el dialecto castellano. Así, <odiando> y <vaciando> se transcriben con diptongo ([o.ˈðjan̪.d̪o] y [ba.ˈθjan̪.d̪o]), cuando en realidad la transcripción de <vaciando> debería realizarse como hiato ([ba.θi.ˈan̪.d̪o]).

Los grafemas mudos no representan ningún fonema, tal y como ocurre con la <h> siempre que no forme parte del dígrafo <ch>. También se consideran grafemas mudos algunos signos de puntuación como las comillas, los guiones y similares. Estos grafemas se eliminan de la cadena y no se consideran parte del entorno fonético.

Por último, los signos de puntuación se consideran como grafemas de pausa, algo que afecta al tipo de alófono que puede aparecer en la palabra previa o en la posterior.

Una vez identificados todos los grafemas, se tienen que purgar las variantes fonéticas que no sean adecuadas, escogiendo el fonema que corresponda a su entorno fonético. Así, la lista de objetos de la clase *Grafema* contenidos en un objeto de la clase *Palabra* se convierte en una lista de objetos de tipo *Fonema,* que representan los fonemas contenidos en la palabra.

3.2.5. La clase Sílaba

Tras crear la lista de objetos de tipo *Fonema,* el objeto *Palabra* los silabea, creando tantos objetos de la clase *Sílaba* como se necesiten. El objeto *Sílaba* no es otra cosa que un contenedor para un máximo de siete fonemas: dos fonemas consonánticos en ataque, semiconsonante, vocal, semivocal y dos consonantes de coda. Además, la *Sílaba* puede contener una pausa al inicio o al final, y puede ser átona o tónica.

Para agrupar la lista de fonemas en sílabas se utiliza un algoritmo sencillo. En primer lugar, se calcula el contenido de los núcleos de las sílabas:

- Se toman los fonemas que se han identificado como vocálicos y se crea una sílaba por cada uno de ellos, asignándolos como núcleos.
- Seguidamente se incluyen en dichas sílabas los fonemas correspondientes a las semiconsonantes y semivocales adyacentes al núcleo (saltando, si fuera necesario, los grafemas mudos).

En segundo lugar, se procesan los fonemas consonánticos, procediendo de la siguiente forma:

- Se añaden como ataque de la primera sílaba todas las consonantes previas a la primera vocal, y, como coda de la última sílaba, todas las consonantes posteriores a la última vocal.
- Todas aquellas consonantes seguidas de un fonema vocálico o semiconsonántico se añaden como ataque de la sílaba a la que pertenezca el vocoide.
- Las consonantes seguidas de consonante se añaden como ataque complejo si su combinación está permitida en español. En cualquier otro caso la consonante en cuestión se añade a la coda de la sílaba cuyo núcleo es la vocal más próxima a dicha consonante por la izquierda.

Por último, se procesan los posibles signos de puntuación y se asignan las pausas al inicio o al final de la palabra.

[14] El manual de transcripción de Fernández Planas y Carrera Sabaté (2001) considera que el fonema /s/ en coda se elide si va seguido de /ɾ/ en ataque.

Así, se coloca cada fonema en su posición correcta dentro de la sílaba a la que pertenezca. Si se tienen más de dos consonantes en coda o en ataque absoluto solo se mantienen la primera y la última. En cualquier caso, estas palabras serán de origen extranjero y su pronunciación no será la estándar.

Tras ello, si la palabra no está integrada en el grupo de palabras átonas, se asigna una de las sílabas como tónica:

- Si la palabra es un adverbio en <-mente>, se asigna un acento primario en la penúltima sílaba, y se sigue procesando el resto como si la palabra careciera de las dos últimas sílabas (pero se asignará un acento secundario en vez de primario).
- Si la palabra contiene un grafema vocálico con tilde, se asigna como tónica la sílaba cuyo núcleo es dicha vocal. Si la palabra contiene más de una tilde, solo se considera la última.
- Si no hay vocales acentuadas, se siguen las normas de acentuación estándar.

Llegados a este punto se obtiene la transcripción fonológica de cada palabra aislada.

3.2.6. La clase Alófono

Esta clase permite calcular el alófono con el que se realiza en el habla cada fonema. Los objetos de esta clase contienen información sobre cuál es el fonema del que derivan, además de restricciones que especifican el entorno fonético en el que pueden aparecer. Estas restricciones están condicionadas por los tres fonemas previos y los tres fonemas posteriores en la secuencia. La necesidad de un entorno fonético tan amplio viene dada por el fenómeno de nasalización de las semivocales o de las semiconsonantes, que depende de que al núcleo al que pertenecen le preceda o le suceda una consonante nasal: por ejemplo, la combinación <ni aun> se transcribe como [nj̃ãn̪]).

Gracias a estas restricciones de aparición se calcula la variante alofónica con la que se realizará cada uno de los fonemas, de forma similar a la que se utilizaba para desambiguar el valor fonológico de cada grafema. En este punto, se obtiene una transcripción fonética estrecha de cada palabra aislada.

3.2.7. Resilabeo de las oraciones

Una vez que el objeto de clase *Oración* ha creado todas las estructuras de clase *Palabra,* silabeadas y transcritas fonéticamente según lo indicado hasta ahora, se procede al resilabeo.

El resilabeo es un proceso inconsciente y natural, que se produce al encadenar palabras en el habla. La existencia de este fenómeno puede ocasionar los siguientes efectos (Hualde, 1989):

- Una consonante pasa de coda de una sílaba a ataque de la siguiente: <el avión hindú> = [ɛl] + [a.ˈβ̪jɔn] + [iṇ.ˈn̪u] → [e.la.ˈβ̪jo.nĩn̪.ˈn̪u]. Esto no ocurre si la sílaba siguiente comienza por consonante, incluso aunque la consonante de la coda y la del ataque siguiente se puedan combinar como ataque complejo (si bien en estos casos puede reforzarse una pronunciación debilitada): <club latino> = [ˈkluᵝ] + [la.ˈt̪i.no] → [ˈkluβ̪.la.ˈt̪i.no].
- Una vocal se consonantiza y pasa a formar parte de la sílaba siguiente: <Eva y Ana> = [ˈe.β̪a] + [i] + [ˈa.na] → [ˈe.β̪a.ˈj̪a.na].
- Una vocal se convierte en paravocal (bien en una semivocal de la sílaba previa, o bien en una semiconsonante de la sílaba siguiente): <uno y dos> = [ˈu.no] + [i] + [ˈd̪ɔs] → [ˈu.nɔi̯.ˈð̞ɔs]; <él y otro> = [ɛl] + [i] + [ˈo.t̪ro] → [ˈe.ˈljo.t̪ro].
- Una consonante o una vocal se alarga: <dame el libro> = [ˈd̪a.me] + [ɛl] + [ˈli.β̪ro] → [ˈd̪a.meː.ˈlːi.β̪ro].

Como se ve en los ejemplos presentados, al realizarse cambios en el ataque y en la coda, además de en la semiconsonante y en la semivocal, se producen modificaciones adicionales en la abertura de la vocal, así como en su nasalidad, debido a la creación de un nuevo entorno fonético. Se pueden incluso perder sílabas si todos sus segmentos migran a las sílabas adyacentes.

Por otra parte, aun cuando no se produzcan traslaciones de segmentos de una sílaba a otra, el mero contacto de dos palabras puede producir cambios alofónicos en el núcleo y en la coda final de la palabra previa o en el ataque absoluto de la palabra siguiente: <la jueza> = [la] + [ˈχwe.θa] → [lɑ.ˈχwe.θa], <con vino> = [ˈk̟ɔn] + [ˈbi.no] → [k̟ɔm.ˈbi.no], <es ron > = [ˈes] + [ˈrɔn] → [ˈe.ˈrɔn], <veo bien> = [ˈbe.o] + [ˈbjen] → [ˈbe.o.ˈβ̪jen].

El proceso del resilabeo depende de la tonicidad de las sílabas. Así, <muy alta> se transcribe como [mu.ˈj̪ɑl̪.t̪a], pero <muy altanera> se transcribe como [mwi.ɑl̪.t̪a.ˈne.ɾa], debido al cambio de tonicidad.

Tras realizar el resilabeo, se unen todas las oraciones incluyendo pausas en sus junturas. Con este último paso se consigue la transcripción fonética estrecha del texto completo, pero además se crea una estructura organizada en fonemas/alófonos integrados en sílabas, con lo que será fácil hacer el tipo de recuento y de estadísticas que se necesitan en este trabajo.

3.3. Recuento de apariciones de segmentos y tipos de sílaba

Llegados a este punto y gracias a las estructuras de datos que se han creado, las estadísticas que se pueden extraer son muchas. Se fijará la atención tan solo en dos aspectos:

- La frecuencia de aparición de cada fonema o alófono según su posición en la sílaba (y según la tonicidad de esta).
- La frecuencia de aparición de los distintos tipos de sílabas, categorizadas en función del número de

segmentos en el ataque, el núcleo y la coda. El hecho de que sea posible tener de cero a dos consonantes en el ataque o en la coda, y la posibilidad de que aparezca una semiconsonante[15] o una semivocal, dan como resultado 30 posibles tipos de sílaba (que, a su vez, pueden ser átonas o tónicas). Estos tipos abarcan desde la sílaba más simple (tipo V) hasta la más compleja (de tipo CCS_cVS_vCC)[16].

4. RESULTADOS

El primer resultado que se puede analizar es el del desempeño del transcriptor. Tal y como se ha podido comprobar en los tests exhaustivos a los que se ha sometido como parte de su desarrollo, la transcripción automática obtenida es de gran calidad y prácticamente libre de errores. Como ejemplo del texto de salida del transcriptor, en la Tabla A4 se muestra una serie de once textos cortos (1618 palabras, 7542 fonemas) que aparecen en el mencionado manual de transcripción de Fernández Planas y Carrera Sabaté (2001). En esta tabla, aparece a la izquierda la transcripción ortográfica del texto, y a la derecha una comparación entre su transcripción fonética automática y la transcripción realizada por las propias autoras del manual de transcripción. Se utilizan estos textos porque han sido estudiados a fondo y se han tomado como modelo para la salida que debe generar el transcriptor automático.

Las transcripciones fonéticas de la tabla incluyen espacios para facilitar la identificación de las palabras con sus símbolos fonéticos. Además, puesto que se trata de una comparación entre dos transcripciones fonéticas distintas, se sigue la convención de que todos los alófonos que coinciden en ambas transcripciones se muestran en fondo blanco sin ningún formato especial, mientras que las divergencias entre la versión manual y automática se marcan de la siguiente forma:

- Los alófonos que aparecen en la transcripción automática pero no en la manual, se muestran con fondo gris, recuadro negro y en negrita, como por ejemplo [ɑ].
- Aquellos alófonos que, por el contrario, aparecen únicamente en la transcripción manual, se muestran con fondo negro, fuente blanca y en negrita, como por ejemplo [ɪ].

Siguiendo esta notación, cuando un alófono no se inserta ni se elide, sino que se modifica, se indica la situación como un alófono que sólo aparece en la transcripción automática, seguido de un alófono que sólo aparece en la transcripción manual. De esta manera, si por ejemplo la transcripción automática incluye el alófono nasalizado [ẽ] mientras que la transcripción manual utiliza [e], la transcripción de la Tabla A4 es [ẽe].

Las dos transcripciones difieren en 175 casos, que involucran el cambio de uno o más alófonos. Se puede comprobar que, de todos estos casos, solo en uno la divergencia puede atribuirse sin dudas a un (predecible) error en la transcripción automática (en concreto con respecto a la palabra [ˌe.ljo.θen̪.ˈt̪ris̪.t̪a], en la que se omite la marca de acento secundario). Este error se produce por la imposibilidad de identificar las palabras compuestas y de asignarles la tonicidad convenientemente, como ya se hizo notar, si bien este tipo de vocablos es poco habitual.

Sin embargo, existen 174 divergencias entre las dos transcripciones que son principalmente atribuibles a diferencias de criterio y a algunos pocos (e inevitables) fallos humanos. Estas 174 divergencias pueden categorizarse en los siguientes tipos:

- El transcriptor automático añade una marca de tonicidad que no aparece en la transcripción manual: 57 casos. De ellos, 52 son debidos a que la transcripción manual considera, por lo general, que los artículos indeterminados y algunas otras palabras gramaticales son átonos, mientras que el transcriptor automático sigue la indicación de la Real Academia Española y la Asociación de Academias de la Lengua Española (2005), que indica lo contrario.
- El transcriptor automático elimina una marca de tonicidad que sí aparece en la transcripción manual: 37 casos. En la transcripción manual se considera que muchas preposiciones y conjunciones son tónicas, en oposición al criterio de la RAE, que es el que sigue el transcriptor automático.
- El transcriptor transcribe un alófono vocálico distinto: 33 casos. Hay 21 casos en los que, siguiendo el criterio del manual de Fernández Planas y Carrera Sabaté (2001), el transcriptor utiliza una variante abierta cuando la transcripción manual lo transcribe como cerrado. En otros 12 casos la transcripción manual no utiliza el diacrítico de nasalización que sí se incluye en la transcripción automática debido a que se dan las condiciones para que el fenómeno ocurra.
- Divergencias en las pausas: 25 casos. La transcripción automática añade 19 pausas y elide otras dos. En otros cuatro casos la pausa escogida en ambas transcripciones es de distinta duración. Estas diferencias se deben a que el transcriptor automático se ciñe a lo que indican los signos de puntuación, mientras que la transcripción manual es más libre en este aspecto (y quizá más realista).

[15] No se permiten sílabas sin ataque y con semiconsonante. En esos casos se produce una consonantización de la semiconsonante: <hielo> [ˈʝe.lo], <cacahuete> [ka.ka.ˈɣwe.t̪e].

[16] Para identificar los elementos de los distintos tipos de sílaba, se utiliza la siguiente convención: C para consonante, S_c para semiconsonante, V para vocal y S_v para semivocal.

- El transcriptor automático combina dos fonemas iguales consecutivos, que en la transcripción manual aparecen separados, y los agrupa en un alófono largo: 12 casos, tanto con vocales como con consonantes. De nuevo, la transcripción manual es más liberal en cuanto a dónde incluir un alófono largo y dónde utilizar dos de duración normal, mientras que el transcriptor automático sigue siempre las reglas de combinación fonética convenidas.
- Divergencia en alófonos consonánticos: 8 casos. Usualmente, se omite un diacrítico en la transcripción manual que, posiblemente, se haya olvidado.
- El transcriptor no incluye un alófono que sí aparece en la transcripción manual: 1 caso. Se trata de la palabra <éxtasis>, transcrita manualmente como [ˈeɣ̟s̪t̪asi̯s̪], con una [ɣ̟] que se omite en la transcripción automática. En este caso la transcripción manual se aleja de la norma expuesta en el manual de transcripción, pero es una pronunciación que también es válida.
- Uso en la transcripción manual de un carácter que no es un símbolo fonético: 1 caso. La transcripción manual utiliza [v] en lugar de [β̞].

Como se ve, la mayor parte de las divergencias no pueden estrictamente considerarse como fallos en ninguna de las dos transcripciones —en concreto las debidas a diferencias de criterio sobre la tonicidad de ciertas palabras gramaticales, la posición o duración de las pausas, y el alargamiento de segmentos—. Las divergencias originadas por cambios en la apertura o nasalización de vocales también son discutibles, puesto que son fenómenos que no siempre ocurren de forma sistemática.

Sin embargo, aún quedarían varias divergencias que podrían ser consideradas como fallos en la transcripción manual. Su existencia prueba que la tarea de la transcripción no es una labor sencilla, sino que requiere gran concentración y puede resultar tediosa, lo que facilita la aparición de errores humanos. Estos errores son difíciles de corregir, ya que a menudo pasan inadvertidos en las revisiones posteriores (especialmente cuando se trata de la ausencia de algún diacrítico en una transcripción larga). Este es un problema que no afecta a un transcriptor automático, el cual, aun con sus limitaciones, ofrece transcripciones que, especialmente sin son largas, pueden llegar a ser incluso más exactas que las realizadas manualmente por profesionales.

En cuanto al resultado del recuento, como se ha comentado ya, el corpus contiene 100 674 739 sílabas, de las cuales el 68,33 % son átonas y el 31,67 % son tónicas. Por lo que se refiere a los fonemas, suman 233 326 472, de los cuales el 66,22 % se encuentra en sílabas átonas y el 33,78 %, en tónicas. Se puede decir, por tanto, que en el español hablado aproximadamente uno de cada tres fonemas o una de cada tres sílabas son tónicos.

De los fonemas analizados, un 46,87 % son vocoides (el 43,15 % vocales, y el 3,72 % deslizantes), y el 53,13 % restante está dividido entre un 41,04 % de fonemas consonánticos en ataque y un 12,09 % en coda. En la Tabla 4 se puede observar la distribución de los distintos fonemas según su porcentaje de aparición, discriminando entre posiciones átonas y tónicas. También se añade una columna con la llamada *tonicidad relativa,* que relaciona el porcentaje de apariciones del fonema en sílabas tónicas con el valor medio mencionado del 33,78 %.

Se puede comprobar que las vocales fuertes /a e o/ son los fonemas más frecuentes. Juntos representan más del 36 % de todos los fonemas de la cadena hablada. Si se ciñe la atención a las vocales, se observa que la siguiente en aparecer, /i/, ocupa la 9.ª posición y la /u/ queda en la 15.ª. En cuanto a las paravocales, la más común es la /j/,

Tabla 4. Porcentajes de aparición de los fonemas del español, desglosados por tonicidad, y su tonicidad relativa (el porcentaje de apariciones tónicas dividido entre la tonicidad media, 33,78 %).

Fonema	% Ap. total	% Ap. átona	% Ap. tónica	Ton. rel.
/a/	13,30	9,27	4,03	0,90
/e/	13,10	9,31	3,79	0,86
/o/	9,85	7,54	2,31	0,70
/s/	7,47	5,72	1,74	0,69
/n/	7,07	4,30	2,77	1,16
/ɾ/	5,88	3,43	2,45	1,23
/d/	4,95	3,79	1,16	0,69
/l/	4,95	3,74	1,21	0,72
/i/	4,67	2,37	2,30	1,46
/t/	4,29	2,70	1,59	1,10
/k/	3,93	2,92	1,01	0,76
/m/	3,04	1,78	1,26	1,23
/b/	2,81	1,58	1,23	1,30
/p/	2,58	1,85	0,73	0,83
/u/	2,23	0,99	1,24	1,65
/j/	2,09	0,71	1,38	1,95
/θ/	1,67	0,79	0,88	1,56
/ɡ/	1,12	0,55	0,57	1,50
/w/	0,87	0,28	0,58	1,99
/r/	0,73	0,50	0,23	0,92
/x/	0,73	0,40	0,33	1,33
/f/	0,64	0,32	0,31	1,46
/i̯/	0,61	0,45	0,16	0,77
/ʎ/	0,42	0,31	0,11	0,77
/j̰/	0,36	0,16	0,20	1,66
/ʧ/	0,28	0,19	0,09	0,93
/ɲ/	0,21	0,13	0,08	1,12
/u̯/	0,15	0,12	0,03	0,61
Total	**100**	**66,22**	**33,78**	**100**

que está en la 16.ª posición. Sigue en esta lista la semiconsonante posterior /w/, en 19.ª posición, y, por último, las semivocales /i̯ u̯/ en 23.ª y 28.ª (y última) posición.

En cuanto a los datos de tonicidad relativa, se puede apreciar que las vocales abiertas son, en proporción, menos tónicas que las cerradas. Por otra parte, las semiconsonantes son los fonemas que más tienden a aparecer en posición tónica, mientras que las semivocales se sitúan en el extremo opuesto.

En cuanto a las consonantes, se observa que, con una diferencia muy apreciable, las consonantes más frecuentes son las alveolares o dentoalveolares /s n ɾ d l t/, que se reparten entre la 4.ª y la 10.ª posición (la /i/ está en la 9.ª, como se comentó). La /r/ sin embargo, es un fonema poco común, que no aparece hasta la 20.ª posición. En un dialecto con seseo, la /s/ sería, con diferencia, el fonema consonántico más habitual del español, con un 9,13 % del total (aproximadamente, una de cada seis consonantes), con una frecuencia de aparición muy cercana a la de la /o/.

Tras las alveolares, el grupo de las bilabiales /m b p/ es el siguiente tipo de consonante más frecuente: entre la 12.ª y la 14.ª posición. Se podría asimismo concluir que el siguiente grupo de consonantes estaría formado por las interdentales, las velares (excepto la /k/, que es más frecuente que las bilabiales, en 11.ª posición) y las labiodentales, es decir con el orden siguiente: /θ g x f/.

Por último, las consonantes menos frecuentes, con diferencia, son las palatales /ʎ ɟ̯ ʧ ɲ/, ubicadas en cuatro de las últimas cinco posiciones, si bien hay que tener en cuenta que, en una pronunciación que no diferencie la /ʎ/ de la /ɟ̯/, el fonema /ɟ̯/ aparecería en la 20.ª posición del listado.

Es más difícil establecer algún tipo de relación entre el segmento y la tonicidad en el caso de las consonantes. Sí se puede apreciar que los fonemas que aparecen más habitualmente en sílaba tónica son /ɟ̯ θ g f x b/, que se incluyen entre las consonantes menos frecuentes. Lo contrario también se cumple en gran medida, ya que las consonantes más frecuentes tienden más a aparecer en sílaba átona que la media de fonemas.

También se ha realizado un análisis en el nivel de alófono, y la Tabla A5 muestra los resultados para los fonemas situados en el núcleo. En ella se muestra el fonema en cuestión, su porcentaje de aparición entre los fonemas del núcleo[17], los alófonos con los que se presenta, el porcentaje de aparición de dicho alófono para ese fonema[18], y los porcentajes de aparición del alófono en sílaba átona o tónica, pero esta vez en relación con el total de fonemas analizados y no solo con los fonemas del núcleo[19].

Puede apreciarse que las variantes alofónicas más frecuentes de los vocoides son sus versiones orales, principalmente las variantes cerradas [j i i̯ e a o w u u̯]. Menos frecuentes que ellas son las variantes orales abiertas [i̯ ɛ ɑ ɔ u̯], y por último, las variantes nasalizadas [j̃ ĩ ĩ̯ ẽ ã õ w̃ ũ ũ̯] son las más escasas. Este mismo patrón se repite para las vocales largas: las orales cerradas son más habituales que las abiertas, y estas, a su vez, son más frecuentes que las nasalizadas.

También se han analizado los alófonos de las consonantes, distinguiendo entre posición de ataque y de coda. En la Tabla A6 y en la Tabla A7 (con las mismas columnas que la Tabla A5) aparecen los resultados para las posiciones de ataque y coda, respectivamente. Como puede observarse, las dos tablas ofrecen resultados bastante diferentes.

Al considerar la distribución de los fonemas en posición de ataque, lo primero que se observa es que las dos consonantes más frecuentes en el recuento total (/s n/) bajan hasta la 5.ª y la 7.ª posición cuando solo se considera el ataque. Estas dos consonantes son eminentemente consonantes de coda, lo que se debe a la propia morfología del español, ya que son terminaciones habituales de muchos morfemas de flexión y derivación nominal y verbal.

En cambio, las dentoalveolares son las consonantes más frecuentes en ataque: /d t/ aparecen en 1.ª y 2.ª posición, de forma que su agrupación constituye el 22,13 % de los fonemas en ataque. Por lo demás, el orden de frecuencias es similar al recuento consonántico global ya presentado. No debe olvidarse que los fonemas consonánticos en ataque suman el 77,24 % del total de fonemas consonánticos, con lo que el recuento global ha de ser forzosamente parecido al recuento de ataque.

Los resultados para las consonantes en coda también son sustancialmente diferentes de los del recuento general de consonantes. Los dos fonemas consonánticos más frecuentes, /s n/, también lo son en coda, pero en posición inversa (algo más frecuente la /n/ que la /s/). Además, en este caso estos dos fonemas por sí mismos representan el 64,21 % del total. Si se les suman las otras dos consonantes alveolares del español /ɾ l/ (la /r/ no puede aparecer en coda salvo en pronunciación enfática), en 3.ª y 4.ª posición, se obtiene el 91,70 % del total. Como puede comprobarse, la distribución de consonantes en coda es mucho más irregular, con unas pocas consonantes que aparecen en la mayoría de los casos.

Puede verse en la Tabla A7 que, de estos cuatro fonemas, la /ɾ/ se encuentra en sílaba tónica mucho más frecuentemente que el resto. Otros fonemas más frecuentes que la media en codas de sílabas tónicas son /x d θ ʧ f/. El resto tiende a ubicarse más en posición átona que la media. Los fonemas más habituales en sílabas átonas que la media son /s b l g p/.

En la Tabla A8 se puede ver un listado de los 106 alófonos utilizados, ordenados por frecuencia de aparición y distribuidos según su aparición en sílabas tónicas y átonas.

En cuanto al análisis del corpus en el nivel de sílaba, la Tabla A9 muestra los 30 distintos tipos silábicos posi-

[17] La suma de porcentajes es, pues, 100 %, y no 46,87 %.
[18] La suma de todas las filas para un fonema dado es por lo tanto el 100 %.
[19] La suma total de la tercera y cuarta columnas es el porcentaje de aparición de vocoides, un 46,87 %.

bles en español, con sus porcentajes de aparición (discriminados por tonicidad). En las últimas columnas se recoge la sílaba concreta con la que se encuentra representado más frecuentemente cada tipo, junto con el dato de la tonicidad relativa, ya utilizado anteriormente.

Según los datos de la Tabla A9 la sílaba por excelencia en español (en la cadena hablada) es la del tipo CV, con el 57,45 % del total, seguida muy de lejos por las que responden a los tipos CVC y V. Solo con la suma de estos tres tipos de sílaba (de los 30 posibles) se aglutina el 83 % de las sílabas. Si se añaden los siguientes cuatro tipos de sílaba más frecuentes (CSCV, CCV, CC y CSCVC) se alcanza el 96,93 %. También puede apreciarse en la Tabla A9 que, por lo general, las sílabas con semivocal tienden a ser átonas y las sílabas con semiconsonante tienden a ser tónicas en mayor proporción que el resto.

Por último, la Tabla A10 recoge las 40 sílabas más frecuentes en la cadena hablada, cuya frecuencia de aparición en conjunto supera ligeramente el 50 %. Son sílabas del tipo CV casi en su totalidad, salvo por dos sílabas de tipo V, otras dos de tipo VC y otra del tipo CVC.

En total, el corpus analizado contiene 13 424 sílabas distintas a nivel de fonema. Dichas sílabas tienen 33 777 realizaciones distintas a nivel alofónico.

5. CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

Comparar los resultados de este trabajo con los de estudios previos (que aparecen en la Tabla A1 y en la Tabla A2) no es un asunto trivial, debido a que en ellos la metodología es diferente: se mezclan distintos dialectos del español y además se agrupan los fonemas de formas diversas. Las paravocales suelen agruparse con las vocales cerradas correspondientes y algunos trabajos utilizan archifonemas en la coda o los desglosan en (algunos) alófonos. En cualquier caso, puede comprobarse que los resultados de este trabajo van en la misma línea que los estudios previos y que las variaciones pueden ser atribuibles al distinto corpus utilizado en el recuento.

Los resultados obtenidos describen bastante bien la realidad fonética del castellano hablado. Sin embargo, tienen claras limitaciones. Los principales fallos se producen debido a los problemas expuestos por Ríos (1993), principalmente con relación al silabeo y a la asignación de los acentos secundarios. Además, no se puede evitar que en el corpus aparezca alguna palabra extranjera que no siempre se transcribirá como es debido, ya que el transcriptor solo está preparado para trabajar con la ortografía española.

Quizá el problema más importante sea el hecho de que existen casos ambiguos en los que la norma indica que existe un diptongo ortográfico aunque oralmente pueda realizarse con diptongo o hiato dependiendo de la palabra y del hablante. Por ejemplo, las secuencias <ia>/<ua> forman habitualmente un diptongo (como en <odiar> o <aguar>), pero existen verbos (y algunas otras pocas palabras) en los que dicha combinación constituye un hiato (<vaciar>, <actuar>, <truhan>, <guion>), algo que no se refleja en la ortografía. Debido a ello, no es posible crear un procedimiento de silabeo que funcione de forma correcta sin tener un lexicón con formas con hiato y formas con diptongo. Estas formas son poco numerosas y se podría crear una pequeña base de datos que incluyera tales excepciones.

En este sentido, la existencia de prefijos que no se silabean junto con el lexema principal también causa problemas. Como se indica en Hualde (1989), estos prefijos se silabean en un nivel más bajo, como si el prefijo y el lexema fueran dos palabras consecutivas. Así, el transcriptor producirá la forma [su.β̞lu.ˈnaɾ] para la palabra <sublunar>, cuando una transcripción más correcta sería [ˌsuβ.lu.ˈnaɾ]. Para corregir este defecto, se podría disponer de una lista de aquellos prefijos que experimentan este fenómeno, de manera que se identificaran y se listaran como palabras independientes, con la particularidad de que llevarían acento secundario. La identificación de los prefijos facilitaría que se produjera la forma correcta también en aquellos casos en que la sucesión de dos vocales produce en realidad un hiato y no un diptongo (<antihéroe> debería transcribirse como [ˌan̪.t̪i.ˈe.ɾo.e] y no como [an̪.ˈtje.ɾo.e]).

Surgen asimismo problemas en cuanto a los acentos secundarios en las palabras compuestas. El transcriptor solo es capaz de detectar los adverbios acabados en <-mente>, pero, según se ha visto, existen prefijos que adquieren un acento secundario, como también ocurre en palabras compuestas del tipo de [ˌsal̪.t̪a.ˈmɔ̃n̪.t̪es], y que se transcriben incorrectamente. Sucede igualmente el fenómeno contrario, es decir, que se transcribe incorrectamente una marca de tonicidad que no debería aparecer, como en los compuestos de dos o más palabras donde solo la última es tónica, como en <centro derecha>, <José Luis> o <nueve mil>, algo de lo que se habla con más detalle en Gómez Torrego (2011).

Aparte del mencionado problema de la tonicidad, la expansión de los numerales no siempre es correcta, ya que se emplea por defecto su forma masculina no apocopada. Así, la transcripción de <21 semanas> será [bɛi̯n̪.ˈt̪ju.no.se.ˈmã.nas]. Esto se podría evitar realizando un análisis sintáctico previo, lo que añadiría bastante complejidad al sistema.

Por otra parte, el tratamiento de siglas, acrónimos y abreviaturas es incompleto. Aunque los acrónimos se deletrean correctamente, existen casos en los que la transcripción producida no es la apropiada. También puede ocurrir que se confunda una palabra o un acrónimo con un número romano. Así, <talla XL> se transcribe erróneamente como [ˈt̪a.ʎa.k̟wa.ˈren̪.t̪a].

El transcriptor creado es muy potente, y es capaz de calcular la transcripción fonética correcta (en la práctica totalidad de los casos) como una sucesión de los 106 alófonos posibles, incluyendo el (re)silabeo y la asignación de tonicidad. Por ello, podría ser un recurso muy importante para un sistema de conversión de texto a habla, en el que la transcripción se tomara como *input* para un dispositivo que

la convirtiera en habla sintética. En este caso, sería deseable analizar, además, las características entonativas del texto, como se hace en TexAFon (Garrido *et al*., 2012).

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a Juan María Garrido Almiñana por atender mis consultas y por sus sugerencias durante la realización de este trabajo.

REFERENCIAS

Alameda, J. R., y Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Alarcos Llorach, E. (1965). *Fonología española* (4.ª ed.). Madrid: Gredos.

Álvarez, C. J., Carreiras, M., y de Vega, M. (1992). Estudio estadístico de la ortografia castellana: (1) La frecuencia silábica. *Cognitiva, 4*(1), 75-105.

Bonaventura, P., Giuliani, F., Garrido, J. M., y Ortín, I. (1998). Grapheme-to-phoneme transcription rules for Spanish, with application to automatic speech recognition and synthesis. *Proceedings of the Workshop on Partially Automated Techniques for Transcribing Naturally Occurring Continuous Speech, 16th August 1998, Université de Montréal.* Montréal: COLING-ACL, 33-39. https://doi.org/10.3115/1628291.1628295

Castro, M. J., España, S., Salvador, I., y Marzalz, A. (2001). Transcriptor ortográfico-fonético para el castellano. *Procesamiento del lenguaje natural, 27,* 241-246.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French*. Heidelberg: Groos.

Fernández Planas, A. M., y Carrera Sabaté, J. (2001). *Prácticas de transcripción fonética en castellano*. Barcelona: Salvatella.

Garrido, J.M., Laplaza, Y., Marquina, M., Schoenfelder, C., y Rustullet, S. (2012). TexAFon: A multilingual text processing tool for text-to-speech applications. *Proceedings of IberSPEECH 2012. Séptimas Jornadas en Tecnología del Habla and 3rd Iberian SLTech Workshop,* 281-289. http://iberspeech2012.ii.uam.es/IberSPEECH2012_OnlineProceedings.pdf

Gómez Torrego, L. (2011). Algunos compuestos sintagmáticos con el primer componente átono y algunas formas prefijadas con la preposición «sin». En M. V. Escandell Vidal, M. Leonetti y C. Sánchez López, *60 problemas de gramática* (pp. 366-379). Madrid: AKAL.

González Rátiva, M. C., y Mejía Escobar, J. A. (2011). Frecuencia fonemática del español de Colombia. *Forma y Función, 24*(2), 69-102. http://www.bdigital.unal.edu.co/37067/

Goyal, K. (2016). Calibre (Versión 2.53) [software]. Obtenido de https://calibre-ebook.com/

Grefenstette, G., y Tapanainen, P. (1994). What is a word, What is a sentence? Problems of Tokenization. *Third International Conference on Computational Lexicography, Budapest, 1994,* 79–87.

Guerra, R. (1983). Estudio estadístico de la sílaba en español. En M. Esgueva y M. Cantarero, *Estudios de fonética 1* (pp. 9-112). Madrid: Consejo Superior de Investigaciones Científicas e Instituto «Miguel de Cervantes».

Guirao, M., y Borzone de Manrique, A. M. (1972). Fonemas, sílabas y palabras en el español de Buenos Aires. *Filología, 16,* 135-165.

Guirao, M., y García, M. A. (1993). *Estudio estadístico del español*. Buenos Aires: Consejo Nacional de Investigaciones Científicas y Técnicas.

Hualde, J. I. (1989). Silabeo y estructura morfémica en español. *Hispania*, *72*(4), 821-831. https://oadoi.org/10.2307/343560

Llisterri, J., y Mariño, J. B. (1993). *Spanish adaptation of SAMPA and automatic phonetic transcription* (informe SAM-A/UPC/001/V1). ESPRIT Project 6819.

Lloyd, P. M., y Schnitzer, R. D. (1967). A statistical study of the structure of the Spanish syllable. *Linguistics, 5*(37), 58-72. https://oadoi.org/10.1515/ling.1967.5.37.58

Marcos Marín, F. (Dir.) (1992) *Corpus Oral de Referencia de la Lengua Española CORLEC* [corpus]. http://www.lllf.uam.es/ESP/Corlec.html

Martínez Celdrán, E., y Fernández Planas, A. M. (2000). Características fonéticas de la africada palatal sonora del español. *Actas del IV Congreso de Lingüística General* (Vol. 4). Universidad de Cádiz, pp. 1751-1761.

Moreno Sandoval, A., Torre Toledano, D., Curto, N., y de la Torre, R. (2006). Inventario de frecuencias fonémicas y silábicas del castellano espontáneo y escrito. http://www.lllf.uam.es/ESP/Publicaciones/LLI-UAM-4JTH.pdf

Moreno Sandoval, A., Toledano, D. T., de la Torre, R., Garrote, M., y Guirao, J. M. (2008). Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories. *Proceedings of the 6th Conference on Language Resources and Evaluation (LREC)*, pp. 1097-1100.

Mosterín, J. (1981). *La ortografía fonémica del español*. Madrid: Alianza Universidad.

Navarro Tomás, T. (1966). Escala de frecuencia de fonemas españoles. *Estudios de fonología española* (2.ª ed.), 15-30. New York: Las Américas.

Pérez, H. E. (2003). Frecuencia de fonemas. *Revista Electrónica en Tecnologías del Habla*, *1*.

Pérez Tobarra, L. (2005). El acento en español. *redELE: Revista Electrónica de Didáctica ELE, 4,* 14-33.

Pineda, L. A., Villaseñor, L., Cuétara, J., Castellanos, H., y López, I. (2004). DIMEx100: A new phonetic and speech corpus for Mexican Spanish. *Advances in Artificial Intelligence-IBERAMIA 2004,* 974-983. Berlin: Springer. https://oadoi.org/10.1007/978-3-540-30498-2_97

Quilis, A., y Esgueva, M. (1980). Frecuencia de fonemas en el español hablado. *Lingüística Española Actual, 2*(1), 1-25.

REAL ACADEMIA ESPAÑOLA, y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2005). *Diccionario panhispánico de dudas*. Madrid: Santillana.

Ríos Mestre, A. (1993). La información lingüística en la transcripción fonética automática del español. *VIII Congreso de la SEPLN, Granada,* 381-387. https://rua.ua.es/dspace/bitstream/10045/4619/1/PLN_13_28.pdf

Rojo Sánchez, G. (1991). Frecuencia de fonemas en español actual. En M. Brea y F. Fernández Rei (Coords.), *Homenaxe ó Profesor Constantino García, vol. 1* (pp. 451-467). http://hdl.handle.net/10347/12469

Sebastián-Gallés, N., Martí, M. A., Carreiras, M. y Cuetos, F. (2000). *LEXESP: Una base de datos informatizada del español* [CD-ROM y libro]. Barcelona: Universitat de Barcelona.

Zipf, G. K., y Rogers, F. M. (1939). Phonemes and variphones in four present-day Romance languages and Classical Latin from the viewpoint of dynamic philology. *Archives néerlandaises de phonétique expérimentale, 15,* 111-147.

ANEXO

A1. Resultados de los trabajos previos (1939-1981) que incluyen recuentos de frecuencias de aparición de fonemas del español (cifras expresadas como porcentaje del total).

Fonema	Zipf y Rogers (1939)	Navarro (1966)	Alarcos (1965)	Delattre (1965)	Lloyd y Schnitzer (1967)	Guirao y Borzone (1972)	Quilis y Esgueva (1980)	Mosterín (1981)
/i/	4,20	6,81	8,60	4,63	8,53	7,27	7,38	7,97
/e/	12,20	12,92	12,60	14,06	9,73	14,51	14,67	13,89
/a/	14,06	13,55	13,70	13,06	15,21	12,45	12,19	12,13
/o/	9,32	9,11	10,30	9,21	10,27	9,85	9,98	9,23
/u/	1,76	2,82	2,10	1,87	2,77	3,08	3,33	3,27
/w/				1,25				
Diptongos	2,02							
Vocoides	**43,56**	**45,21**	**47,30**	**44,08**	**46,51**	**47,16**	**47,55**	**46,49**
/p/	2,92	2,97	2,10	2,29	2,36	2,76	2,77	2,40
/t/	4,46	4,67	4,60	4,74	5,32	4,92	4,53	4,40
/k/	3,84	4,10	3,80	4,64	3,76	4,37	3,98	3,95
/b/	3,26	2,46	2,50	2,85	2,92	2,45	2,37	2,10
/B/			0,10		0,13		0,03	
/d/	5,06	4,85	4,00	5,20	3,81	4,16	4,24	4,59
/D/			0,25		0,23		0,31	
/g/	1,02	1,01	1,00	0,71	1,46	0,94	0,94	0,84
/G/			0,25		0,31		0,28	
/f/	0,72	0,70	1,00	0,52	1,46	0,67	0,55	0,88
/θ/	1,74	2,16	1,70	1,42	2,49		1,45	1,95
/s/	8,12	8,24	8,00	8,35	4,26	9,72	8,32	9,01
/x/	0,58	0,49	0,70	0,37	1,02	0,65	0,57	0,73
/ʧ/	0,30	0,29	0,40	0,32	0,57	0,33	0,37	0,15
/j̯/	2,40	0,39	0,40	2,94	0,15		0,41	0,09
/m/	2,98	3,00	2,50	3,73	2,48	3,04	3,06	2,74
/n/	5,94	6,73	2,70	7,01	2,34	7,67	2,78	7,99
/ɲ/	0,36	0,35	0,20	0,30	0,28	0,28	0,25	0,00
/N/			3,70		4,44		4,86	
/ɾ/	5,90	5,73	2,50		8,24	5,58	3,26	5,59
/R/			4,50	6,23			1,93	
/r/	1,04	0,78	0,60		1,17	0,50	0,43	0,81
/l/	5,20	5,29	4,70	3,84	2,96	4,25	4,23	4,98
/ʎ/	0,60	0,58	0,50	0,47	0,60		0,38	0,31
/ʒ/						0,55		
Consonantes	**56,44**	**54,79**	**52,70**	**55,93**	**52,76**	**52,84**	**52,30**	**53,51**

A2. Resultados de trabajos previos (1991-2011) que incluyen recuentos de frecuencias de aparición de fonemas del español (cifras expresadas como porcentaje del total). La última columna corresponde a este trabajo.

Fonema	Rojo (1991)	Guirao y García (1993)	Llisterri y Mariño (1993)	Pérez (2003)	Pineda *et al.* (2004)	Moreno *et al.* (2006)	González y Mejía (2011)	Arias (2016)
/i̯/								0,61
/i/	7,50	6,59	4,29	7,46	6,60	7,59	7,82	4,67
/j/			2,60		1,90			2,09
/e/	13,51	14,99	13,72	14,13	10,94	12,74	12,11	13,10
/ɛ/					2,59			
/a/	13,40	13,27	13,43	12,31	11,45	12,89	14,38	13,30
/ɑ/					0,62			
/o/	9,57	10,75	10,37	9,28	5,05	9,32	11,20	9,85
/ɔ/					4,26			
/u/	3,16	2,79	1,98	3,05	1,87	3,04	2,95	2,23
/w/			1,35		1,14			0,87
/u̯/								0,15
Vocoides	**47,14**	**48,39**	**47,74**	**46,23**	**46,42**	**45,58**	**48,46**	**46,87**
/p/	2,66	2,68	2,60	2,58	2,57	2,73	2,76	2,58
/t/	4,48	4,48	4,63	4,92	4,66	4,31	5,26	4,29
/k/	3,98	4,31	4,04	3,94	3,33	3,80	3,55	3,93
/k̟/					0,57			
/b/	2,66	3,08	0,45	1,92	0,32	2,55	3,60	2,81
/β̞/			2,47		1,79			
/d/	4,79	4,00	0,76	4,84	0,98	5,42	4,41	4,95
/ð̞/			3,20		4,59			
/g/	0,95	1,11	0,11	0,94	0,09	1,04	1,36	1,12
/ɣ̞/			0,79		0,73			
/f/	0,68	0,53	0,51	0,75	0,80	0,92	0,71	0,64
/θ/	1,68		1,53			2,00		1,67
/s/	7,58	9,39	6,95	9,61	7,86	7,33	5,84	7,47
/s̪/			1,33		1,27			
/s̟/					1,11			
/x/	0,73	0,70	0,63	0,74	0,70	0,77	0,91	0,73
/tʃ/	0,28	0,40	0,40	0,32	0,15	0,18	0,39	0,28
/tɕ/							0,10	
/j̟/	0,22	0,72	0,19	0,69	0,31		0,85	0,36
/t̪j̟/					0,01			
/m/	3,09	3,17	3,63	2,62	2,77	2,76	3,63	3,04
/n/	6,99	7,14	7,02	7,78	4,85	7,09	7,16	7,07
/n̪/					1,86			
/ŋ/			0,46		0,38			
/ɲ/	0,19	0,24	0,27	0,24	0,13	0,31	0,28	0,21
/ɾ/	5,67	5,40	4,25	6,19	5,70	6,19	6,47	5,88
/r/	0,79	0,39	0,40	0,64	0,62	0,99	0,98	0,73
/l/	5,08	3,88	4,25	5,05	5,43	5,46	3,28	4,95
/ʎ/	0,38		0,54			0,53		0,42
Consonantes	**52,88**	**51,62**	**51,41**	**53,77**	**53,58**	**54,38**	**51,54**	**53,13**

A3. Listado de las obras que componen el corpus y su año de publicación, agrupados por autor.

Autor/Obra	Año
Adón, Pilar	
El mes más cruel	2010
Aguilar, Carlos	
Nueve colores sangra la luna	2005
Árbol, Víctor del	
La tristeza del samurái	2011
Respirar por la herida	2013
Un millón de gotas	2014
Asensi, Matilde	
El salón de ámbar	1999
Iacobus	2000
El último catón	2001
El origen perdido	2003
Peregrinatio	2004
Todo bajo el cielo	2006
Tierra Firme	2007
Venganza en Sevilla	2010
La conjura de Cortés	2012
Benítez Reyes, Felipe	
Mercado de espejismos	2007
Caballero Bonald, José Manuel	
Dos días de septiembre	1962
Campo de Agramante	1992
Carrillo, Mónica	
La luz de Candela	2014
Caso, Ángeles	
El peso de las sombras	1994
Un largo silencio	2000
Las olvidadas, una historia de mujeres creadoras	2005
Contra el viento	2009
Donde se alzan los tronos	2012
Cela Conde, Camilo José	
Cela, mi padre. La vida íntima y literaria de Camilo José Cela contada por su hijo	1989
Cela, Camilo José	
La familia de Pascual Duarte	1942
Viaje a La Alcarria	1948
La colmena	1951
Mrs Caldwell habla con su hijo	1953
Primer viaje andaluz. Notas de un vagabundaje por Jaén, Córdoba, Sevilla, Huelva y sus tierras	1959
Garito de hospicianos o guirigay de imposturas o bombollas	1963
San Camilo 1936	1969
La insólita y gloriosa hazaña del cipote de Archidona	1977
Mazurca para dos muertos	1983
Nuevo viaje a la Alcarria	1986
Cachondeos, escarceos y otros meneos	1993
La cruz de San Andrés	1994
Madera de boj	1999
Cercas, Javier	
El móvil	1987
El vientre de la ballena	1997
Soldados de Salamina	2001
La velocidad de la luz	2005
Anatomía de un instante	2009
Las leyes de la frontera	2012
El impostor	2014
Corral, José Luis	
El salón dorado	1996
El amuleto de bronce	1998
El invierno de la corona	1999
El Cid	2000
Trafalgar	2001
La torre y el caballero. El ocaso de los feudales	2002
Numancia	2003
El número de Dios	2004
¡Independencia!	2005
Breve historia de la Orden del Temple	2006
El caballero del templo	2006
Fulcanelli, el dueño del secreto	2008
El rey felón	2009
Fátima, el enigma de las apariciones	2009
¿Qué fue la Corona de Aragón?	2010
La prisionera de Roma	2011
El Códice del Peregrino	2012
El médico hereje	2013
El trono maldito	2014
Delibes, Miguel	
La sombra del ciprés es alargada	1947
El camino	1950
La partida	1954
Diario de un cazador	1955
Diario de un emigrante	1958

La hoja roja	1959
Las ratas	1962
Viejas historias de Castilla la Vieja	1964
Cinco horas con Mario	1966
La mortaja	1970
El príncipe destronado	1973
Mis amigas las truchas	1977
El disputado voto del señor Cayo	1978
Un mundo que agoniza	1979
Los santos inocentes	1981
Dos viajes en automóvil: Suecia y Países Bajos	1982
Cartas de amor de un sexagenario voluptuoso	1983
Tres pájaros de cuenta	1987
Mi querida bicicleta	1988
Señora de rojo sobre fondo gris	1991
Diario de un jubilado	1995
El hereje	1998
España 1939-1950: Muerte y resurrección de la novela	2004
Dueñas, María	
El tiempo entre costuras	2009
Misión Olvido	2012
Eslava Galán, Juan	
En busca del unicornio	1987
Roma de los césares	1988
Yo, Aníbal	1988
Tartessos y otros enigmas de la historia	1991
El viaje de Tobías	1992
Historia secreta del sexo en España	1992
Historias de la Inquisición	1992
Los templarios y otros enigmas medievales	1992
Historia de España contada para escépticos	1995
Julio César, el hombre que pudo reinar	1995
El fraude de la Sábana Santa y las reliquias de Cristo	1997
Tumbaollas y hambrientos. Los españoles comiendo y ayunando a través de la historia	1997
Señorita	1998
Escuela y prisiones de Vicentito González	1999
Los dientes del Dragón	2001
El paraíso disputado. Ruta de los castillos y las batallas	2003
La mula	2003
Los íberos. Los españoles como fuimos	2004
Una historia de la guerra civil que no va a gustar a nadie	2005
El mercenario de Granada	2006
Califas, guerreros, esclavas y eunucos. Los moros en España	2008
Los años del miedo	2008
El catolicismo explicado a las ovejas	2009
De la alpargata al seiscientos	2010
Rey Lobo	2010
Homo Erectus	2011
La década que nos dejó sin aliento	2011
La primera guerra mundial contada para escépticos	2014
Etxebarría, Lucía	
Amor, curiosidad, prozac y duda	1997
Beatriz y los cuerpos celestes	1998
Nosotras que no somos como las demás	1999
Un milagro en equilibrio	2004
Lo verdadero es un momento de lo falso	2010
El contenido Del Silencio	2011
Tu corazón no está bien de la cabeza. Cómo salí de una relación tóxica	2013
Falcones, Ildefonso	
La catedral del mar	2006
La mano de Fátima	2009
La reina descalza	2013
Freire, Espido	
Melocotones helados	1999
Cuando comer es un infierno. Confesiones de una bulímica	2002
Cuentos malvados	2003
Mileuristas: cuerpo, alma y mente de la generación de los 1000 euros	2006
La flor del Norte	2011
Los malos del cuento. Cómo sobrevivir entre personas tóxicas	2013
García Montero, Luis	
Habitaciones separadas	1994
Mañana no será lo que Dios quiera	2009
No me cuentes tu vida	2012
Una forma de resistencia	2012
Alguien dice tu nombre	2013
Giner, Gonzalo	
La cuarta alianza	2005
El secreto de la logia	2007
El sanador de caballos	2008
El jinete del silencio	2011
Pacto de lealtad	2015
Gómez Jurado, Juan	

La leyenda del ladrón	2012
Goytisolo, Juan	
Juegos de manos	1954
Para vivir aquí	1960
Señas de identidad	1966
Paisajes después de la batalla	1982
Coto vedado	1985
En los reinos de taifa	1986
Las virtudes del pájaro solitario	1988
El sitio de los sitios	1995
Carajicomedia	2000
Grandes, Almudena	
Las edades de Lulú	1989
Te llamare Viernes	1991
Malena es un nombre de tango	1994
Modelos de mujer	1996
Atlas de geografía humana	1998
Los aires difíciles	2002
Castillos de cartón	2004
Estaciones de paso	2005
El corazón helado	2007
Inés y la alegría	2010
El lector de Julio Verne	2012
Las tres bodas de Manolita	2014
Los besos en el pan	2015
Lindo, Elvira	
Manolito Gafotas	1994
Pobre Manolito	1995
¡Cómo molo!	1996
Los trapos sucios	1997
El otro barrio	1998
Manolito on the road	1998
Yo y el imbécil	1999
Tinto de verano	2001
Tinto de verano 2. El mundo es un pañuelo	2002
Manolito tiene un secreto	2002
Tinto de verano 3. Otro verano contigo	2003
Una palabra tuya	2008
Lo que me queda por vivir	2010
Lugares que no quiero compartir con nadie	2011
Mejor Manolo	2012
Llamazares, Julio	
Luna de lobos	1985
La lluvia amarilla	1988
El río del olvido	1990
Trás-os-Montes	1998
El cielo de Madrid	2005
Las lágrimas de San Lorenzo	2013
Lloréns, Chufo	
La otra lepra	1993
Catalina, la fugitiva de San Benito	2001
La saga de los malditos	2003
Te daré la tierra	2008
Mar de fuego	2011
Mallorquí del Corral, César	
El último trabajo del Señor Luna	1997
La cruz de El Dorado	1999
La catedral	2000
Las lágrimas de Shiva	2002
El viajero perdido	2005
El juego de Caín	2008
El juego de los herejes	2010
La isla de Bowen	2012
Marías, Javier	
Los dominios del lobo	1971
Travesía del horizonte	1973
El hombre sentimental	1986
Mientras ellas duermen	1990
Corazón tan blanco	1992
Vidas escritas	1992
Mañana en la batalla piensa en mí	1994
Cuando fui mortal	1996
Mala índole	1998
Fiebre y lanza	2002
Baile y sueño	2004
Veneno y sombra y adiós	2007
Los enamoramientos	2011
Así empieza lo malo	2014
Marsé, Juan	
Encerrados con un solo juguete	1960
Últimas tardes con Teresa	1966
La oscura historia de la prima Montse	1970
Si te dicen que caí	1973
La muchacha de las bragas de oro	1978
Un día volveré	1982
Ronda del Guinardó	1984

El amante bilingüe	1990
El embrujo de Shanghai	1993
Rabos de lagartija	2000
Caligrafía de los sueños	2011
Matute, Ana María	
Los soldados lloran de noche	1963
La trampa	1969
Luciérnagas	1993
Demonios familiares	2014
Mejide, Risto	
#Annoyomics: El arte de molestar para ganar dinero	2000
No busques trabajo	2000
El pensamiento negativo	2008
Que la muerte te acompañe	2011
Mendoza, Eduardo	
La verdad sobre el caso Savolta	1975
El misterio de la cripta embrujada	1978
El laberinto de las aceitunas	1982
La ciudad de los prodigios	1986
Nueva York	1986
La isla inaudita	1989
Sin noticias de Gurb	1991
El año del diluvio	1992
La aventura del tocador de señoras	2001
Baroja, la contradicción	2001
El último trayecto de Horacio Dos	2002
Mauricio o las elecciones primarias	2006
El asombroso viaje de Pomponio Flato	2008
Tres vidas de santos	2009
Rina de gatos. Madrid 1936	2010
El enredo de la bolsa y la vida	2012
Millás, Juan José	
Visión del ahogado	1977
La soledad era esto	1990
Tonto, muerto, bastardo e invisible	1995
El orden alfabético	1998
No mires debajo de la cama	1999
Dos mujeres en Praga	2002
El mundo	2007
Lo que sé de los hombrecillos	2010
Articuentos completos	2011
La mujer loca	2014
Moix, Terenci	
No digas que fue un sueño	1986
El cine de los sábados	1990
Garras de astracán	1991
La herida de la esfinge	1991
Memorias. El peso de la paja. El beso de Peter Pan	1993
Mujercísimas	1995
El amargo don de la belleza	1996
Memorias. El peso de la paja. Extraño en el paraíso	1998
Chulas y famosas	1999
El arpista ciego	2002
Montero, Carla	
La piel dorada	2014
Montero, Rosa	
Te trataré como a una reina	1983
Amado amo	1988
Temblor	1990
Bella y oscura	1993
La hija del caníbal	1997
Amantes y enemigos. Cuentos de parejas	1998
El corazón del tártaro	2001
La loca de la casa	2003
Historia del Rey Transparente	2005
Lágrimas en la lluvia	2011
Dictadoras: Las mujeres de los hombres más despiadados de la historia	2013
La ridícula idea de no volver a verte	2013
Muñoz Molina, Antonio	
Beatus ille	1986
El invierno en Lisboa	1987
Beltenebros	1989
Córdoba de los Omeyas	1991
El jinete polaco	1991
Los misterios de Madrid	1992
Ardor guerrero	1995
Plenilunio	1997
Carlota Fainberg	1999
En ausencia de Blanca	2001
Ventanas de Manhattan	2004
El viento de la luna	2006
La noche de los tiempos	2009
Navarro, Julia	
La hermandad de la sábana santa	2004
La Biblia de barro	2005
La sangre de los inocentes	2007

Dime quién soy	2010
Dispara, yo ya estoy muerto	2013
Historia de un canalla	2016
Palomas, Alejandro	
El alma del mundo	2011
El tiempo que nos une	2016
Pérez Reverte, Arturo	
El maestro de esgrima	1988
La tabla de Flandes	1990
El club Dumas o La sombra de Richelieu	1993
La sombra del águila	1993
Territorio comanche	1994
La piel del tambor	1995
Un asunto de honor	1995
El capitán Alatriste	1996
Limpieza de sangre	1997
El sol de Breda	1998
Patente de corso	1998
El oro del rey	2000
La carta esférica	2000
La Reina del Sur	2002
El caballero del jubón amarillo	2003
Cabo Trafalgar	2004
El húsar	2004
Corsarios de Levante	2006
El pintor de batallas	2006
Ojos azules	2009
El asedio	2010
El puente de los asesinos	2011
El tango de la guardia vieja	2012
El francotirador paciente	2013
Pombo, Álvaro	
Relatos sobre la falta de sustancia	1977
Aparición del eterno femenino contada por S. M. el Rey	1993
Donde las mujeres	1996
Contra natura	2005
La Fortuna de Matilda Turpin	2006
El temblor del héroe	2012
Quédate con nosotros, Señor, porque atardece	2013
Posadas, Carmen	
Pequeñas infamias	1998
La cinta roja	2008
Invitación a un asesinato	2010
El testigo invisible	2013
Posteguillo Gómez, Santiago	
Africanus: el hijo del cónsul	2006
La traición de Roma	2009
Los asesinos del emperador	2011
Circo Máximo	2013
La sangre de los libros	2014
Trajano y Decebalo en la Rumanía del siglo XXI	2014
Las legiones malditas	2008
Prada, Juan Manuel de	
Las máscaras del héroe	1996
La tempestad	1997
Rivas, Manuel	
¿Qué me quieres, amor?	1997
El lápiz del carpintero	2002
Los libros arden mal	2006
Todo es silencio	2010
Las voces bajas	2012
Ruescas, Javier	
Cuentos de Bereth I: Encantamiento de luna	2009
Cuentos de Bereth II: La maldición de las Musas	2010
Tempus Fugit, ladrones de almas	2010
Cuentos de Bereth III: Los versos del destino	2011
Salas, Antonio	
El año que trafiqué con mujeres	2004
Diario de un skin	2010
El palestino	2011
Operación princesa	2013
Sampedro, José Luis	
El río que nos lleva	1961
Octubre, Octubre	1981
La sonrisa etrusca	1985
La vieja sirena	1990
Real Sitio	1993
El amante lesbiano	2000
Escribir es vivir	2005
Sánchez Adalid, Jesús	
La luz de Oriente	2000
El mozárabe	2001
Félix de Lusitania	2002
La tierra sin mal	2003
El cautivo	2004
La sublime puerta	2005

En compañía del sol	2006
El caballero de Alcántara	2008
Los milagros del vino	2010
Galeón	2011
Alcazaba	2012
El camino Mozárabe	2013
Treinta doblones de oro	2014
Y de repente, Teresa	2014
Sánchez Dragó, Fernando	
La prueba del Laberinto	1992
Sánchez, Mamen	
Gafas de sol para días de lluvia	2007
Agua del limonero	2010
Juego de damas	2011
La felicidad es un té contigo	2013
Se prohíbe mantener afectos desmedidos en la puerta de la pensión	2014
Sánchez-Garnica, Paloma	
La brisa de oriente	2009
El alma de las piedras	2011
Las tres heridas	2012
La sonata del silencio	2014
Santandreu, Rafael	
El arte de no amargarse la vida	2014
Las gafas de la felicidad	2014
Savater, Fernando	
Ética para Amador	1991
Política para Amador	1992
El jardín de las dudas	1993
La hermandad de la buena suerte	1993
El valor de educar	1997
Las preguntas de la vida	1999
Mira por dónde. Autobiografía razonada	2003
Los diez mandamientos en el siglo XXI	2004
El gran laberinto	2005
La vida eterna	2007
Ética de urgencia	2012
Los invitados de la princesa	2012
Sierra, Javier	
Roswell, secreto de Estado	1995
La dama azul	1998
En busca de la edad de oro	2000
Las puertas templarias	2000
La cena secreta	2004
La ruta prohibida y otros enigmas de la Historia	2007
El ángel perdido	2011
El quinto mundo	2012
El maestro del Prado	2013
La guía secreta del Prado	2013
La pirámide inmortal	2014
Silva, Lorenzo	
Noviembre sin violetas	1995
La sustancia interior	1996
La flaqueza del bolchevique	1997
El lejano país de los estanques	1998
El ángel oculto	1999
El urinario	1999
El alquimista impaciente	2000
La lluvia de París	2000
Del Rif al Yebala	2001
El nombre de los nuestros	2001
La isla del fin de la suerte	2001
Niños feroces	2001
La niebla y la doncella	2002
El déspota adolescente	2003
Carta blanca	2004
Nadie vale más que otro	2004
La reina sin espejo	2005
El blog del inquisidor	2008
El Derecho en la obra de Kafka	2008
La estrategia del agua	2010
Sereno en el peligro	2010
Tres mil metros en la noche	2011
Zona desdinerizada	2011
La marca del meridiano	2012
Siete ciudades en África: Historia del Marruecos español	2013
Los cuerpos extraños	2014
Torres, Maruja	
Un calor tan cercano	1998
Mientras vivimos	2000
Esperadme en el cielo	2009
Fácil de matar	2011
Sin entrañas	2012
Diez veces siete	2014
Umbral, Francisco	
Valle-Inclán, los botines blancos de piqué	1968
Las ninfas	1975

Mortal y rosa	1975
Ramón y las vanguardias	1978
Nada en el domingo	1988
El fulgor de África	1989
La leyenda del césar visionario	1992
Memorias borbónicas	1992
Madrid 650	1995
La forja de un ladrón	1997
La bestia rosa	1998
El socialista sentimental	1999
Un ser de lejanías	2001
Cela: un cadáver exquisito	2002
Días felices en Argüelles	2005
Carta a mi mujer	2008
Ussía, Alfonso	
Memorias del marqués de Sotoancho. La albariza de los juncos	1998
El secuestro de Mamá y otros relatos del marqués de Sotoancho	1999
Lo que Dios ha unido que no lo separe Mamá	2000
Pachucha tirando a mal	2001
Un talibán en La Jaralera	2002
Las dos bodas. El Príncipe y Sotoancho se casan	2004
Las canicas, las cuquis y el novio tontito de Mamá	2005
Mamá se quiere morir y no hay manera	2006
¡Milagro! Se ha muerto Mamá	2007
El diario de Mamá	2009
Vargas Llosa, Mario	
Los jefes y otros cuentos	1959
La ciudad y los perros	1963
La casa verde	1966
Conversación en La Catedral	1969
García Márquez: historia de un deicidio	1971
Pantaleón y las visitadoras	1973
La orgía perpetua. Flaubert y Madame Bovary	1975
La tía Julia y el escribidor	1977
La guerra del fin del mundo	1981
Historia de Mayta	1984
¿Quién mató a Palomino Molero?	1986
El hablador	1987
Elogio de la madrastra	1988
La verdad de las mentiras	1990
El loco de los balcones	1993
El pez en el agua	1993
Lituma en Los Andes	1993
Cartas a un joven novelista	1997
Los cuadernos de don Rigoberto	1997
La fiesta del chivo	2000
Artículos y ensayos	2002
El Paraíso en la otra esquina	2003
Travesuras de niña mala	2006
El viaje a la ficción	2008
El sueño del celta	2010
La civilización del espectáculo	2012
El héroe discreto	2013
Vázquez Figueroa, Alberto	
Bajo siete mares. Largo viaje al paraíso	1968
Tierra virgen. La destrucción del Amazonas	1973
¿Quién mató al embajador?	1974
Ébano	1975
Manaos	1975
¡Panamá, Panamá!	1977
Marea negra	1977
Nuevos dioses	1980
Tuareg	1980
Matar a Gadafi	1981
La iguana	1982
Océano	1984
Yaiza	1984
Maradentro	1985
Cienfuegos	1987
Palmira	1987
Viracocha	1987
El perro	1989
Caribes	1990
Azabache	1991
Sicario	1991
Montenegro	1992
Brazofuerte	1993
África llora	1994
Negreros	1996
Piratas	1996
Leon Bocanegra	1998
Los ojos del tuareg	2000
Tiempo de conquistadores	2000
Bora Bora	2001
Delfines	2001

La puerta del Pacífico	2004
Alí en el país de las maravillas	2005
El rey leproso	2005
Centauros	2007
Pederastas	2007
Por mil millones de dólares	2007
Vivos y muertos	2007
Coltán	2008
Saud el leopardo	2009
Garoé	2010
El mar en llamas	2011
Irina Dogonovic	2011
Codicia	2012
Bimini	2013
Vázquez Montalbán, Manuel	
Yo maté a Kennedy. Impresiones, observaciones y memorias de un guardaespaldas	1972
Tatuaje	1974
La soledad del mánager	1977
La rosa de Alejandría	1984
El balneario	1986
Historias de padres e hijos	1987
Escritos subnormales	1989
Galíndez	1990
El hombre de mi vida	2000
La aznaridad	2003
Vila Matas, Enrique	
Bartleby y compañía	2001
París no se acaba nunca	2003
Doctor Pasavento	2005
El viento ligero en Parma	2008
Aire de Dylan	2012

A4. Comparación entre transcripción manual y automática. Con fondo gris los símbolos que aparecen en la transcripción automática pero no en la manual, y con fondo negro los que solo aparecen en la transcripción manual.

Transcripción ortográfica	Transcripción fonética estrecha
Los metros son históricos, mientras que el ritmo se confunde con el lenguaje mismo. No es difícil distinguir en cada metro los elementos intelectuales y abstractos y los más puramente rítmicos. En las lenguas modernas los metros están compuestos por un determinado número de sílabas, duración cortada por acentos tónicos y pausas. Los acentos y las pausas constituyen la porción más antigua y puramente rítmica del metro; están cerca aún del golpe de tambor, de la ceremonia ritual y del talón danzante que hiere la tierra. El acento es danza y rito. Gracias al acento, el metro se pone en pie y es unidad danzante. La medida silábica implica un principio de abstracción, una retórica y una reflexión sobre el lenguaje. Duración puramente lineal, tiende a convertirse en mecánica pura. Los acentos, las pausas, las aliteraciones, los choques o reuniones inesperadas de un sonido con otro, constituyen la porción concreta y permanente del metro.	[lɔs ˈmetro ˈs ːon isˈtorikɔs‖ ˈmĩẽntras ke ːl ˈriθmo se kɔŋˈfunde kon el ːeŋˈgwɑxe ˈmismo‖ no ˈez ðiˈfiθil diˈstiŋˈgir eŋ ˈkaða ˈmetro los eleˈmẽntos inteleˈtwales j asˈtraytos i lɔs ˈmas ˌpuraˈmẽnte ˈriθmikɔs‖ en las ˈleŋgwas moˈðɛrnasˈs‖ lɔs ˈmetros esˈtaŋ kɔmˈpwestɔs poˈr un deterˈmĩnaðo ˈnumero ðe ˈsilaβas| duraˈθjɔŋ kɔrˈtaða por aˈθentɔs ˈtonikos i ˈpausas‖ los aˈθentos i las ˈpausas kɔnˈstiˈtujen la pɔrˈθjɔn ˈmas anˈtiywa i ˌpuraˈmẽnte ˈriθmika ðel ˈmetro| esˈtan ˈθɛrka ːun del ˈʝolpe ðel tamˈbɔr| de la θereˈmõnja riˈtwal i ðel taˈlɔn danˈθante ke ˈjere la ˈtjera‖ el aˈθento ˈez ˈðanθa i ˈrito‖ ˈgraθjas al aˈθento| ɛl ˈmetro se ˈponẽe ːm ˈpje ˈj es uniˈðað ːanˈθante‖ la meˈðiða siˈlaβika imˈplikɑ ˈum priˈθipjo ðe astraʏˈθjɔn| ˈuna reˈtorika j una rɛfleʏˈsjɔn soβre ːl ːeŋˈgwɑxe‖ duraˈθjɔm ˌpuraˈmẽnte lineˈɑl| ˈtjende a kɔmbɛrˈtirse ːⁿ meˈkanika ˈpura‖ los aˈθentɔs| las ˈpausas| las alitɛraˈθjones| lɔs ˈʧokes ɔ reuˈnjõnes inespeˈraðaz ðe ˈuun soˈniðo koˈn otro| kɔnˈstiˈtujen la pɔrˈθjɔn kɔŋˈkreta j pɛrmãˈnẽnte ðɛl ˈmetro]
Es posible que si Galileo hubiese concretado su modelo heliocentrista tras un éxtasis místico, la Iglesia hubiese acabado por aceptarlo sin usar la Inquisición como medio disuasorio para obligarle a reconocer su error. Pero lo hizo con una máquina diabólica: el telescopio; y con un don derivado de un pacto con el diablo: la observación, ampliada, eso sí, por el efecto de las lentes y la experimentación, merced a la cual el mundo pierde la escala humana y se acerca a la divinidad.	[ˈes poˈsiβle ke si ʝaliˈleo uˈβjese kɔŋkreˈtaðo su moˈðelo eljoθenˈtriˈsta ˈtraˈs uˈn eʏˈstasis ˈmistiko| la jˈʏlesjɑa uˈβjese akaˈβaðo por aθeβˈtarlo sin uˈsar la iŋkisiˈθjɔŋ komõ ˈmeðjo ðiswaˈsorjo parɑa oβliˈʝarle a rɛkonoˈθɛr sw ɛˈrɔr‖ pero lo ˈiθo koˈn ũnã ˈmakina ðjaˈβolika| ɛl teles kopjo| i koˈn ũn ˈdɔŋ deriˈβaðo ðe ˈuum ˈpaʏto kon ɛl ˈdjaβlo| lɑa ɔβserβaˈθjɔn| aam ˈpljaða| ˈeso ˈsi| por el eˈfɛʏto ðe las ˈlentes i la esperimẽnta ˈθjɔn| mɛrˈθeð a la ˈkwal ɛl ˈmũndo ˈpjɛrðe la esˈkalɑ uˈmãna j se aˈθɛr ka ː la ðiβiniˈðað]

Señorita: esta misma mañana, bajo la dulce llovizna del cielo, cruzó usted, aparición fortuita, por delante de la puerta de la casa donde aún vivo y ya no tengo hogar. Cuando desperté, fui a la puerta de la suya, donde ignoro si tiene usted hogar o no le tiene. Me habían llevado allí sus ojos, sus ojos, que son refulgentes estrellas mellizas en la nebulosa de mi mundo. Perdóneme, Eugenia, y deje que le dé familiarmente este dulce nombre; perdóneme la lírica. Yo vivo en perpetua lírica infinitesimal.	[seɲoˈɾiɾa| ˈesɾa ˈmismã mãˈɲãna| ‖bɑχo la ˈðuɭθe ʌoˈβiθna ðɛɭ ˈθjelo| kɾuˈθo ɯsˈɾeð| apaɾiˈθjɔɱ fɔɾˈɾwiɾa| pɔɾ ðeˈlaɲɾe ðe la ˈpwɛrɾa ðe la ˈkasa ‖ðɔɲde ɑˈum ˈbiβɔ i ˈja no ˈɾeŋgo ːˈɣaɾ| ‖kwaɲdo ðespɛrˈɾe| ˈfwi a la ˈpwɛrɾa ðe la ˈsuja| ɖɔɳdɛ iɣˈnoɾo si ˈɾjene ɯsˈɾeð oˈɣaɾ o no le ˈɾjene‖ me aˈβiaɲ ʌeˈβaðo aˈʌi suˈs‖ ɔχɔs| suˈs‖ ɔχɔs| ke ˈsɔn ɾefuˈxeɲɾes esˈɾɾeʌaʂ meˈʌiθas en la neβuˈlosa ðe mĩ ˈmũɲdo‖ pɛɾˈðonẽme| euˈxenja| i ˈðɛxe ke le ˈðe famiˌljaɾˈmẽɲɾe ːˈesɾe ˈðuɭθe ˈnɔ̃mbɾe| pɛɾˈðonẽme la ˈliɾika‖ t͡ʃo ˈβiβo em pɛɾˈpeɾwa ˈliɾika ɨɱfiniɾesiˈmɑl]
Los amores de la Guindilla y Quino, el Manco, tardaron en conocerse en el pueblo. Además, progresaron con una lentitud crispante. Era un paso definitivo, a la postre. Quino, el Manco, ya había pensado en ella, en la Guindilla, antes del incidente con los mozos. La Guindilla no era joven y él tampoco. Por otro lado, la Guindilla era enjuta y delgada y poseía un negocio en marcha; y un evidente talento comercial. Precisamente de lo que él carecía. Últimamente, Quino estaba asfixiado por las hipotecas. Bien mirado, propiedad de él, lo que se dice de él, no restaba ni un hierbajo del huerto.	[los aˈmoɾeʐ ðe la ɣiɲˈɖiʌa i ˈki no‖ ɛl ˈmãŋko‖ ɾaɾˈðaron ẽŋ konoˈθɛrse ːn ɛl ˈpweβlo‖ aðeˈmas| pɾoɣɾeˈsaɾɔŋ ko̞ˈn ũna leɲtiˈɾuð kɾiˈsˈpaɲɾe‖ ˈeɾɑ ˈɯ̰m ˈpaso ðefiniˈɾiβo| a la ˈpɔsɾɾe‖ ˈkino‖ ɛl ˈmãŋko‖ ‖ɟiɟa ːˈβia penˈsaðo eˈn eʌa| en la ɣiɲˈɖiʌa| ˈanɾeʐ ðel iɲθiˈðeɲɾe kɔn lɔʂ ˈmoθɔs‖ la ɣiɲˈɖiʌa no ˈeɾɑ ˈχo βeˈn ‖ɟiɟɛl ɾamˈpoko‖ poˈɾ oɾo ˈlaðo| la ɣiɲˈɖiʌa ˈeɾa eɴˈχuɾa i ðɛlˈɣaðа i poseˈiɑ ˈɯ̰n ːeˈɣoθjo eⁿ ˈmaɾʧa‖ ‖ɟiɟun eβiˈðeɲɾe ɾaˈleɲɾo komɛrˈθjɑl‖ pɾeˌθisaˈmẽɲɾe ðe lo ‖kɛe ːɛl kaɾeˈθia‖ ˌulɾimãˈmẽɲɾe| ˈkino esˈɾaβa ːsfiɣˈsjaðo pɔɾ las ipoˈɾekas‖ ˈbjeⁿ miˈɾaðo| pɾopjeˈðaˈð ːɛe ːɛl| lo ke se ˈðiθe ˈðɛe ːˈɛl| nɔ rɛʂˈɾaβa ‖nj̃ ũ̃n ‖ɟiɛrˈβɑχo ðɛl ˈɣwɛrɾo]
Julio vivía con unas tías viejas; su padre, empleado en una capital de provincia, era de una posición bastante modesta. Julio se mostraba muy independiente; podía haber buscado la protección de su primo Enrique Aracil, que por entonces acababa de obtener una plaza de médico en el hospital, por oposición, y que podía ayudarle; pero Julio no quería protección alguna, no iba ni a ver a su primo; pretendía debérselo todo a sí mismo. Dada su tendencia práctica, era un poco paradójica esta resistencia suya a ser protegido. Julio, muy hábil, no estudiaba casi nada, pero aprobaba siempre. Buscaba amigos menos inteligentes que él para explotarles; allí donde veía una superioridad cualquiera, fuese en el orden que fuese, se retiraba. Llegó a confesar a Hurtado que le molestaba pasear con gente de más estatura que él. Julio aprendía con gran facilidad todos los juegos. Sus padres, haciendo un sacrificio, podían pagarle los libros, la matrícula y la ropa. La tía de Julio solía darle, para que fuera alguna vez al teatro, un duro todos los meses, y Aracil se las arreglaba jugando a las cartas con sus amigos, de tal manera, que, después de ir al café y al teatro y comprar cigarrillos, al cabo del mes, no solo le quedaba el duro de su tía, sino que tenía dos o tres más.	[ˈχuljo βiˈβia ko̞ˈn ũnaʂ ˈtiaʂ ˈβjexas| su ˈpaðɾe| empleˈaðo eˈn ũna kapiˈtɑl ɖe pɾoˈβiɲθja| ˈeɾa ðe ˈɯ̰na posiˈθjɔm baʂˈtaɲɾe moˈðesta‖ ˈχuljo se mɔʂˈɾaβa mw̃wĩ ːɲdepeɲˈɖjeɲɾe| poˈðia ːˈβɒɛr βɯsˈkaðo la pɾoɾɛɛɣˈθjɔn ɖe su ˈpɾimo enˈɾiɲe aɾaˈθil| ke poɾ eɲˈtɔnθes akaˈβaβa ðe ɔoβˈte̞neˈɾ una ˈplaθa ðe ˈmeðiko en el ɔspiˈtɑl| poɾ oposiˈθjɔon| i ke poˈðia ːju̯ˈðaɾle| ‖peɾɔ ˈχuljo no ke̞ˈɾia pɾoɾɛɣˈθjon ɑlˈɣuna| no ˈi βa nj a ˈβeɾ a su ˈpɾimo| pɾeɾeɲˈɖia ðeˈβɛrselo ˈɾoðo a ‖si ˈmiʂmo‖ ˈɖaða su ɾeɲˈɖenθja ˈpɾaɣɾika| ˈeɾɑ ˈɯ̰m ˈpoko paɾaˈðɔxika ˈesɾa ɾɛsiʂˈɾenθja ˈsuja ː ˈsɛɾ pɾoɾɛɛˈxiðo‖ ˈχuljo|muˈj aβiɭ| no esɾuˈðjaβa ˈkasi ˈnaða| ‖peɾo a pɾoˈβaβa ˈsjem pɾe‖ bɯsˈkaβa ːˈmiɣɔʂ ˈmẽnos iɲteliˈxeɲɾes ‖ke̞ɛ ːɛl para esploˈɾarles| aˈʌi ðɔɲde βɣeˈia ˈɯ̰na superjoɾiˈðað kwɑlˈkjeɾa| ˈfwese ːn eˈl ɔɾðeɲ ke ˈfwese| sɛ ɾeɾiˈɾaβa‖ ʌeˈɣo a kɔɱfeˈsaɾ ɑ uɾˈtaðo ke le moleʂˈtaβa paseˈaɾ kɔŋ ˈxeɲɾe ðe ˈmas esɾaˈɾuɾa ‖ke̞ɛ ːɛl‖ ˈχuljo apɾeɲˈdia kɔŋ ˈgɾaɱ faθiliˈðað ˈɾoðɔʂ lɔs ˈχweɣɔs‖ sus ˈpaðres| aˈθjeɲdo ˈɯ̰n sakɾiˈfiθjo| poˈðiam paˈɣaɾle lɔʂ ˈliβɾɔs| la maˈɾikula i la ˈrɔpa‖ la ˈtia ðɛe ˈχuljo soˈlia ˈðaɾle| ‖paɾa ke ˈfweɾɑ ːlˈɣuna ˈβeθ ɑl ɾeˈaɾo| ‖uɲ ˈduɾo ˈɾoðɔʂ lɔʂ ˈmeses‖ ɟi aɾaˈθil se las aɾɛˈɣlaβɑ χuˈɣaɲdo a las ˈkaɾɾas kɔn sus aˈmiɣɔs ‖ʐ| d‖ðe ˈɾol mãˈneɾa‖ ke| ɖesˈpwez ðe ˈiɾ ɑl kaˈfe i ɑl ɾeˈaɾɾo i kɔmˈpɾaɾ θiɣaˈɾiʌ‖ɔs| ɑl ˈkaβo ðɛl ˈmes| no ˈsolo le ke̞ˈðaβa ɛl ˈduɾo ðe su ˈɾia‖ ‖sino ke ɾeˈnia ˈðos o ‖ɾreʂ ˈmas]

Vetusta, la muy noble y leal ciudad, corte en lejano siglo, hacía la digestión del cocido y de la olla podrida, y descansaba oyendo entre sueños el monótono y familiar zumbido de la campana de coro, que retumbaba allá en lo alto de la esbelta torre en la Santa Basílica. La torre de la catedral, poema romántico de piedra, delicado himno, de dulces líneas de belleza muda y perenne, era obra del siglo dieciséis, aunque antes comenzada, de estilo gótico, pero, cabe decir, moderado por un instinto de prudencia y armonía que modificaba las vulgares exageraciones de esta arquitectura. La vista no se fatigaba contemplando horas y horas aquel índice de piedra que señalaba al cielo; no era una de esas torres cuya aguja se quiebra de sutil, más flacas que esbeltas, amaneradas como señoritas cursis que aprietan demasiado el corsé; era maciza sin perder nada de su espiritual grandeza, y hasta sus segundos corredores, elegante balaustrada, subía como fuerte castillo, lanzándose desde allí en pirámide de ángulo gracioso, inimitable en sus medidas y proporciones. Como haz de músculos y nervios, la piedra, enroscándose en la piedra, trepaba a la altura, haciendo equilibrios de acróbata en el aire; y como prodigio de juegos malabares, en una punta de caliza se mantenía, cual imantada, una bola grande de bronce dorado, y encima otra más pequeña, y sobre esta una cruz de hierro que acababa en pararrayos.	[beˈt̪us̪t̪a| la mw̃wn̥| ˈnoβlɛ i̯ leˈɑl̪ θjuˈðaðˌ| ˈkɔrt̪e ːn lɛˈxano ˈsiʏloˌ| aˈθia la ðˌi̯xeʂˈt͡ʃjɔn̪ d̪el̪ koˈθiðɔ i̯ ðe lɑa ˈoʌa poˈðriða| i ðeskanˈsaβɑ oˈi̯en̪do ɐn̪t̪re ˈsweɲos ɛl mõˈnot̪onɔo i̯ famiˈljar θum̥ˈbiðo ðe la kamˈpana ðe ˈkɔrɔ| k̟e rɛɐt̪umˈbaβa ːˈʌa en lo ˈɑl̪t̪o ðe la eʂˈβɛl̪t̪a ˈt̪ɔrɐe ːn la ˈsan̪t̪a βaˈsilika‖ la ˈt̪ɔrɛ ðe la kat̪eˈðˌɾɑl| poˈema rɔˈmãn̪t̪iko ðe ˈpjeðˌra| d̪eliˈkaðo ˈi̯nm̥ːno |d̪ðe ˈðu̯l̪θeʂ ˈlineaz̪ ðe βeˈʌeθa ˈmuða i̯ peˈrenːe| ˈerɑ ˈoβra ðɛl ˈsiʏlo ðjeθiˈseɨs| ɑu̯n̪k̟e ˈan̪t̪es k̟omẽɲˈθaða| d̪e ːs̪ˈt̪ilo ˈʏot̪iko| perɔ| ˈkaβe ðeˈθi̯ɾ| moðeˈraðo poˈɾ un ɪ̃ɲˈs̪t̪iɲt̪o ðe pɾuˈðẹn̪ðja i̯ armõˈnia k̟e moðifiˈkaβa las̪ βu̯lˈʏares ɛʏsɑa̯xeɾaˈθjonez̪ ˈðe ːs̪t̪a ːɾk̟it̪ɛʏˈt̪uɾa‖ la ˈβis̪t̪a no se faʈiˈʏaβa k̟ɔn̪t̪emˈplan̪ˈdo ːo̯raˈs j oras aˈk̟eˈl i̯n̪diθe ðe ˈpjeðɾa k̟e seɲaˈlaβɑ ːl̪ ˈθjelo| no ˈerɑ ˈu̯n̪a ˈðe ːsaʂ ˈt̪ɔres |ku̯a ːˈʏu̯xa se ˈk̟jeβɾa ðe suˈt̪il̪| ˈmas ˈflakas k̟ɐs̪ˈβɛl̪t̪as| amãneˈɾaðas k̟omo seɲoˈɾiɾas ˈkʉrsis k̟e aˈpɾjeɾan̪ d̪emaˈsjaðo ɛl kɔrˈse| ˈera maˈθiθa sim pɛrˈðɛr ˈnaða ðe sw espiɾiˈt̪wɑl ʏraɲˈd̪eθa| ʄ͡ʝ aʂt̪a sus ːeˈʏu̯n̪d̪ɔs kɔɾeˈðores| eleˈʏan̪t̪e βalɑu̯ʂˈt̪raða| suˈβia |komo ˈfwɛrt̪e kaʂˈt̪iʌo| lan̪ˈθan̪d̪ose |ðeʐðe aˈʌi em piˈɾamiðe ðe ˈaŋgulo ʏraˈθjoso| inĩmiˈt̪aβle ːn suʂ meˈðiðas i pɾopɔrˈθjones‖ |komo ˈaθ̟ ðe ˈmuskulos i ˈnɛrβjɔs| la ˈpjeðɾa| enrɔsˈkan̪d̪ose ːn la ˈpjeðra| t̪reˈpaβa ː lɑ ːl̪ˈt̪uɾa| aˈθjen̪do ek̟iˈliβɾjɔz̪ ðe aˈkroβat̪a en eˈl ai̯ ɾe| i |k̟omo pɾoˈðixjo ðɛ ˈχweʏɔʂ malaˈβares| eˈn ũna ˈpu̯n̪t̪a ðe kaˈliθa se mãn̪t̪eˈnia|ˈkwal imãn̪ˈt̪aða| ˈuna ˈβola ˈʏɾan̪d̪e ðe ˈβɾɔn̪θe ðoˈraðo ʄ͡ʝ en ˈθimɑa ˈot̪ɾa |mas peˈk̟eɲa| i |soˈβre ːe̯s̪t̪ɑ ˈuna ˈkɾu̯θθ ðe ˈi̯ɛɾɔ k̟e akaˈβaβa em paɾaˈraɟɔs]
Se me encogió el corazón cuando vi a ese muchacho sentado en el borde del sofá con los paquetes envueltos en papel de regalo, su pelo recién cortado, sin saber dónde meter ese amor que anduvo juntando todos estos años para Irene; buen mozo me pareció, alto y elegante como un príncipe, bien vestido como siempre anda él, tieso como un palo de escoba, un verdadero caballero, pero de poco le vale su pinta de galán, porque la niña no se fija en esas cosas y menos ahora que anda enamorada del fotógrafo; camarón que se duerme se lo lleva la corriente, no debió irse Gustavo dejándola sola por tantos meses. Yo no entiendo a estas parejas modernas, en mis tiempos no había tanta libertad y todo funcionaba como es debido: la mujer callada en su casa. Las novias esperaban bordando sábanas y no andaban encaramadas al anca de las motocicletas de otros hombres; eso debió prevenirlo el Capitán en vez de partir de viaje tan tranquilo, yo lo vi desde el principio y se lo dije: ausencias causan olvidos; pero nadie me hizo caso, me miraron con lástima, como si yo fuera una estúpida, pero no tengo ni un pelo de tonta, más sabe el diablo por viejo que por diablo.	[se mẽe ːŋkk̟oˈxjo ɛl k̟koraˈθɔŋ k̟waɲd̪o ˈβi a ˈese muˈʧaʧo seɲˈt̪aðo en ɛl ˈβɔrðe ðɛl soˈfa k̟ɔn lɔs paˈk̟et̪es emˈbwɛl̪t̪os emn̥ paˈpɛl̪ d̪ɛ rɛˈʏalo|su ˈpelɔ rɛˈθjeŋ k̟ɔrˈt̪aðo| sin saˈβɛr ˈðɔn̪de meˈt̪eˈɾ ese aˈmɔr k̟e aɲˈd̪uβɔ χu̯nˈt̪an̪do ˈt̪oðo's eʂt̪o's aɲɔs para i̯ˈrene| ˈbweⁿ ˈmoθo me pareˈθjo| ˈɑl̪t̪o i̯ eleˈʏan̪t̪e |komo ˈu̯u̯m ˈpɾiɲθi pe| ˈbjem beʂˈt̪iðo |komo ˈsjempɾe ˈan̪d̪a ˈɛl| ˈt̪jeso |komo ˈu̯u̯m ˈpalo ðe ːsˈk̟oβa| ˈum bɛɾðaˈðero kaβaˈʌe ɾo| |pero ðe ˈpok̟o le ˈβale su ˈpiɲt̪a ðe ʏaˈlanm̥| pɔrk̟e la ˈnĩɲã no se ˈfi̯xa eˈn esas ˈk̟osas i ˈmẽnos ɑˈɔra k̟e ˈan̪d̪a enãmoˈɾaða ðɛl foˈt̪oʏɾafo| kamaˈrɔŋ k̟e se ˈðˌwɛrme se lo ˈʌeβa la k̟ɔˈrjen̪t̪e| no ðeˈβjo ˈi̯ɾse ɣu̯sˈt̪aβo ðɛˈxan̪d̪ola ˈsola pɔr ˈt̪an̪t̪oʂ ˈmeses‖ ʄ͡ʝo no eɲˈt̪jen̪d̪o a ˈeʂt̪as paˈrɛxaʂ moˈðɛrnas| eⁿ miʂ ˈt̪jempɔʂ no aˈβia ˈt̪an̪t̪a liβɛrˈt̪að i ˈt̪oðo fu̯nθjoˈnaβa k̟omo ˈez̪ ðeˈβiðo|‖ la muˈxɛr kaˈʌaða en su ˈkasa‖ las̪ ˈnoβjas espeˈɾaβam bɔrˈðan̪d̪o ˈsaβanas i no an̪ˈd̪aβan ẽɕŋkaraˈmaðas aˈl aŋka ðe las̪ moʈoθiˈkl̪eʈaz̪ ðe ˈoʈɾoˈs ɔmbres| ˈeso ðeˈβjo pɾeβeˈniɾlo ɛl kapiˈʈan ẽm ˈbeθ̟ ðe parˈʈiɾ ðe ˈβjɑxe ʈan ʈraŋˈk̟ilo| ʄ͡ʝo lo ˈβi ðezʂðɛ ːl pɾiɲˈθipjɔo i̯ se lo ˈðˌi̯xe| ɑu̯ˈseɲθjas ˈkɑu̯san ɔlˈβiðɔs| |pero ˈnaðje me ˈiθo ˈkaso| mẽ miˈɾaɾɔn̪ k̟ɔn ˈlaʂt̪ima| k̟omo si ˈi̯o ˈfwerɑ ˈu̯u̯na eʂˈt̪upiða| |pero no ˈt̪eɲgo ˈɲʄ ũu̯m ˈpelo ðe ˈt̪ɔn̪t̪a| ˈmaˈs ːaβɛ ːl̪ ˈd̪jaβlo pɔr ˈβjɛχo k̟e pɔr ˈðjaβlo]
Todo calla; todo reposa. Pasa de tarde en tarde, cruzando el ancho ámbito, con esa indolencia privativa de los perros de pueblo, un alto mastín que se detiene un momento, sin saber por qué, y luego se pierde a lo lejos por una empinada calleja; una bandada de gorriones se abate rápida sobre el suelo, picotea, salta, brinca, se levanta veloz y se aleja piando, moviendo voluptuosamente las alas sobre el azul límpido. A lo lejos, como una nota metálica, incisiva, que rasga de pronto la diafanidad del ambiente, vibra el cacareo sostenido de un gallo.	[ˈt̪oðo ˈkaʌa| ˈt̪oðɔ rɛˈposa‖ ˈpasa ðe ˈt̪arðe ːn ˈt̪arðe|kɾuˈθan̪d̪o eˈl an̪ʧo ˈambi̯t̪o| k̟oˈn esa i̯n̪d̪oˈlen̪θja pɾiβaˈt̪iβa ðe lɔs ˈpɛɾɔz̪ ðe ˈpweβlo| ˈuˈn ɑl̪t̪o maʂˈt̪i̯n k̟e se ðeˈt̪jene ˈuⁿ mõˈmẽn̪t̪o| si̯n saˈβɛr pɔr ˈk̟e| i ˈlweʏo se ˈpjɛrðe a lo ˈlɛχɔs poˈɾ una empiˈnaða kaˈʌɛxa| ˈuna βan̪ˈd̪aða ðe ʏˈrjones ːe aˈβat̪e ˈrapiða |soβɾɛ ːl ˈswelo| piko̯ˈt̪ea| ˈsɑl̪t̪a| ˈbɾi̯ŋka|se leˈβan̪t̪a βeˈloθ i se aˈlɛxa ˈpjan̪d̪o| moˈβjen̪d̪o βoluβ̟t̪wosaˈmẽn̪t̪e laˈs alas ːoβɾe ːl aˈθuˈl ːimpiðo‖a lo ˈlɛχɔs||komo ˈu̯u̯nã ˈnoʈa meˈʈalika| i̯n̪θiˈsiβa| k̟ɛ ɾaʂʏa ðe ˈpɾɔn̪t̪o la ðjafaniˈðaðˀ ːðel amˈbjen̪t̪e| ˈbiβɾa ɛl kakaˈɾeo sɔs̪t̪eˈniðo ðe ˈu̯u̯n̥ ˈgaʌo]

La cumbre. Ahí está el ocaso, todo empurpurado, herido por sus propios cristales, que le hacen sangre por doquiera. A su esplendor, el pinar verde se agria, vagamente enrojecido; y las hierbas y las florecillas, encendidas y transparentes, embalsaman el instante sereno de una esencia mojada, penetrante y luminosa. Yo me quedo extasiado en el crepúsculo. Platero, granas de ocaso sus ojos negros, se va, manso, a un charquero de aguas de carmín, de rosa, de violeta; hunde suavemente su boca en los espejos, que parece que se hacen líquidos al tocarlos él; y hay por su enorme garganta como un pasar profundo de umbrías aguas de sangre. El paraje es conocido, pero el momento lo trastorna y lo hace extraño, ruinoso y monumental. Se dijera, a cada instante, que vamos a descubrir un palacio abandonado... La tarde se prolonga más allá de sí misma, y la hora, contagiada de eternidad, es infinita, pacífica, insondeable...	[la ˈkumbɾe‖ aˈi eˈsta el oˈkaso‖ ˈtoðo empurpuˈɾaðo‖ eˈɾiðo pɔr sus ˈpropjɔs kɾiˈstales‖ ke le ˈaθen ˈsaŋgɾe pɔr ðoˈkjera‖ a sw espleŋˈdɔr‖ el piˈnar ˈβɛrðe se ˈaɣɾja‖ ˌbaɣaˈmẽnte ːnrɔxeˈθiðo‖ i laˈs ˈjɛrβas i las floɾeˈθiʎas‖ enθenˈdiðas i traⁿspaˈɾentes‖ embɑlˈsamãn el iⁿsˈtante seˈɾeno ðe ˈuna eˈsenθja mɔˈxaða‖ peneˈtrante i lumiⁿˈnosa‖ ʧo me ˈkeðo esˈtaˈsjaðo en el kɾeˈpuskulo‖ plaˈtero‖ ˈgranaz ðe oˈkaso suˈs ɔχɔˈs ˈneɣɾɔs‖ se ˈβa‖ ˈmãnso‖ ɑ ˈuŋn ʧarˈkero ðe ˈaɣwaz ðe karˈmĩn‖ doˈɛ ˈrɔsa‖ de βjoˈleta‖ ˈunde ˌswaβeˈmẽnte su ˈβoka en los esˈpɛχɔs‖ ke paˈɾeθe ke se ˈaθen ˈlikiðos ɑl toˈkarloˈs ɛl‖ ʧi ai pɔr sw eˈnɔrme ɣarˈɣanta ˈkomo ˈuŋm paˈsar proˈfundo ðe um'bɾiaˈs aɣwaz ðe ˈsaŋgɾe‖ el paˈrɑˈxe ːs konoˈθiðo‖ ˈpero el mõˈmẽnto lo traˈsˈtɔrna i lo ˈaθe ːsˈtraɲo‖ rwiˈnosɔ i mõnũmẽˈntɑl‖ se ðiˈxera‖ a ˈkaða iⁿsˈtante‖ ke ˈβamos a ðesku'βɾiˈr uŋm paˈlaθjo aβanˈdoˈnaðo‖ la ˈtarðe se proˈlɔŋga ˈmas aˈʎa ðe ˈsi ˈmiʂma‖ i lɑ ˈɔra‖ kɔntɑˈxjaða ðe ːtɛrniˈðað‖ ˈes imfiˈnita‖ paˈθifika‖ insɔnˈdeˈaβle]
La lingüística estudia el lenguaje humano. El lenguaje se manifiesta solo en los seres humanos a través de las lenguas que les permiten hablar entre sí y consigo mismos. Hemos dicho que se manifiesta solo en los seres humanos, pues cualquier otra cosa que reciba la denominación de lenguaje lo será solo metafóricamente; es decir, por similitud al lenguaje humano. Los mismos seres humanos hablan entre sí; esto es, se comunican intercambiando mensajes con los demás seres humanos de su entorno; pero también hablan consigo mismos, aunque no se profieran palabras: el lenguaje es también la base del pensamiento humano; no es posible conectar dos ideas, ni tan siquiera estructurar una sola sin la ayuda del lenguaje. De ahí que sea el centro de nuestra vida intelectual y social. Por esto mismo, las lenguas, manifestaciones concretas del lenguaje humano, juegan un papel fundamental en la cultura de cualquier pueblo.	[la liŋˈgwiˈstika esˈtuðja el ːeŋˈgwɑxe uˈmãno‖ el ːeŋˈgwɑxe se mãniˈfjesta ˈsolo en loˈs ːeres uˈmãnos a ˈtraˈβez ðe laˈs ˈleŋgwas ke les pɛrˈmiten aˈβlar enˈtre ˈsi ːi kɔnˈsiyo ˈmiʂmɔs‖ ˈemɔz ˈðitʃo ke se mãniˈfjesta ˈsolo en loˈs ːeres uˈmãnɔs‖ pwes kwɑlˈkjeˈr oˈtra ˈkosa kɛ reˈθiβa la ðenõmĩnaˈθjɔn de leŋˈgwɑxe lo seˈra ˈsolo meta‚foɾikaˈmẽnte‖ ˈez ðeˈθir‖ pɔr similiˈtuˈðˈ al ːeŋˈgwɑxe uˈmãno‖ lɔʂ ˈmiʂmoˈs ːeres uˈmãnoˈs aβlaˈn ẽnˈtre ˈsi‖ ˈesto ˈes‖ se komũˈnikan iⁿnˈtɛrkamˈbjanˈdo mẽnˈsɑxes kɔn lɔz ðeˈmaˈs ːeres uˈmãnɔz ðe ˈsw eŋˈtɔrno‖ pero tamˈbjeˈn aβlaŋ kɔnˈsiyo ˈmiʂmɔs‖ ˈɑuŋke no se proˈfjeram paˈlaβɾas‖ el ːeŋˈgwɑˈxe ːs tamˈbjen la ˈβase ðɛl pensaˈmʃẽnto uˈmãno‖ no ˈes poˈsiβle koneɣˈtar ˈðos iˈðeas‖ ni tan siˈkjera esˈtruɣtuˈɾaˈr una ˈsola sin la ːˈjuða ðel ːeŋˈgwɑxe‖ de aˈi ke ˈsea ɛl ˈθentro ðe ˈnwesˈtra ˈβiða inˈteleɣˈtwal i soˈθjɑl‖ poˈr esto ˈmiʂmo‖ laʂ ˈleŋgwas‖ mãnifesˈtaˈθjones kɔŋˈkreˈtaz ðel ːeŋˈgwɑxe uˈmãno‖ ˈχweɣaˈn ũˈm paˈpɛl fuɲdamẽˈtal en la kulˈtura ðe kwɑlˈkjeɾ ˈpweβlo]
El hombre viene subiendo por el medio del camino, silbando entre dientes una canción y tirando sin demasiadas ganas de la caballería. Trae un viejo tabardo de piel vuelta, descolorido ya por los años y la lluvia, y un sombrero hongo de fieltro hundido hasta los ojos. Quizá por eso no nos ve hasta que está ya prácticamente encima de nosotros. Aún no son las ocho todavía de un día que ha amanecido hinchado de negros nubarrones, amenazando lluvia, y, aquí arriba, en el puerto de Amarza, la humedad y la luz se funden formando una misma sustancia, una niebla pegajosa y fría que empapa mansamente la tierra y el espacio. Cuando no nos ve, parados en medio del camino, al final de una revuelta, el hombre tira del ronzal a la caballería y se detiene. De reojo, bajo el ala del sombrero, mientras Ramiro y yo nos acercamos, observa los hayedos más cercanos buscando otras personas. Recibe con recelo mi saludo. Pero sus ojos, hundidos bajo el sombrero no dejan traslucir la menor sombra de miedo.	[eˈl ɔmbɾe ˈβjene suˈβjeɲdo poɾ el ˈmeðjo ðɛl kaˈmĩno‖ siˈlˈβaɲdo ˈenˈtɾe ˈðjenˈte‚s una kaɲˈθjon i ˈtiˈɾanˈdo siɲ ˈdema‚sjaðaʂ ˈɣanaz ðe la kaβaʎeˈɾia‖ ˈtrae ˈuŋm ˈbjɛχo ˈtaˈβarðo ðe ˈpjɛl ˈβwɛlta‖ ˈdeskoloˈɾiðo ˈja pɔr loˈs aɲos i la ˈʌuβja‖ ʧi un sɔmˈbɾeˈɾɔo ːɔŋgo ðe ˈfjɛltɾɔo unˈdiðo ˈaʂta loˈs ɔχɔs‖ kiˈθa poˈr eso nõo nɔs ˈβe aʂta ke ːsˈta ˈja ˌpɾaɣtikaˈmẽnte ːnˈθima ðe noˈsoˈtɾɔs‖ ɑˈun ːo ˈsɔn la's oʧo toðaˈβia ðe ˈuŋn ˈdia ke ˈa ːmãneˈθiðo inˈʧaðo ðe ˈneɣɾɔʂ nuβaˈrɔnes‖ amẽnaˈθandɔ ˈʌuβja‖ i‖ aˈki aˈriβa‖ en el ˈpwɛrto ðe aˈmarθa‖ lɑ ume'ðað i la ˈluθ se ˈfuɲdeɱ fɔrˈmãɲdo ˈuŋnã ˈmiʂma suʂˈtanθja‖ ˈunã ˈnjeβla peɣɑˈχosa i ˈfria ke ːmˈpapa ˌmãnsaˈmẽnte la ˈtjɛra i el esˈpaθjo‖ ˈkwanˈdo nõo nɔʂ ˈβe‖ paˈraðos eⁿ ˈmeðjo ðɛl kaˈmĩno‖ ɑl fiˈnɑl de ˈuŋna rɛˈβwɛlta‖ eˈl ɔmbre ˈtira ðɛl rɔnˈθal a la kaβaʎeˈɾia i se ðeˈtjene‖ de rɛˈɔχo‖ ˈbɑχo eˈl ala ðɛl sɔmˈbɾero‖ ˈmʃẽnˈtra raˈmiɾɔ i ˈi nos aθɛrˈkamɔs‖ ɔβˈserβa los aˈi̯eðoʂ ˈmas θɛrˈkanoʂ βuʂˈkanˈdo ːtras pɛr ˈso nas‖ rɛˈθiβe kɔn rɛˈθelo mi saˈluðo‖ ˈpero su's ɔχɔs‖ unˈdiðoʂ βɑχo ɛl sɔmˈbɾero‖ no ˈðexaɲ traʂluˈθir la mẽˈnɔr ˈsombɾa ðe ˈmjeðo]

A.5. Porcentajes de aparición de fonemas y alófonos en núcleo: porcentaje de aparición del fonema en núcleo, y porcentaje global de apariciones del alófono según tonicidad.

Fon.	% apa. núcleo	Al.	% del fonema	% áto. total	% tón. total
/a/	28,39	[a]	88,54	8,24	3,54
		[ɑ]	6,72	0,61	0,28
		[aː]	2,16	0,25	0,04
		[ã]	2,13	0,16	0,12
		[ɑː]	0,41	0,05	0,01
		[ãː]	0,04	<0,01	<0,01
/e/	27,96	[e]	76,98	7,40	2,69
		[ɛ]	16,83	1,51	0,69
		[ẽ]	2,96	0,15	0,23
		[eː]	2,55	0,20	0,13
		[ɛː]	0,59	0,06	0,02
		[ẽː]	0,08	0,01	<0,01
/o/	21,03	[o]	67,21	5,22	1,40
		[ɔ]	30,77	2,20	0,83
		[õ]	1,00	0,08	0,02
		[ɔ̃]	0,52	0,02	0,03
		[oː]	0,33	0,02	0,01
		[ɔː]	0,16	0,01	0,01
		[ɔ̃ː]	<0,01	<0,01	<0,01
		[õː]	<0,01	<0,01	<0,01
/i/	9,94	[i]	76,04	1,81	1,73
		[i̯]	20,36	0,45	0,50
		[ĩ]	2,48	0,08	0,04
		[ĩ̯]	0,87	0,04	<0,01
		[iː]	0,12	<0,01	<0,01
		[i̯ː]	0,08	<0,01	<0,01
		[ĩ̯ː]	0,04	<0,01	<0,01
		[ĩː]	0,01	<0,01	<0,01
/u/	4,78	[u]	61,27	0,67	0,70
		[u̯]	31,63	0,31	0,40
		[ũ̯]	3,78	0,01	0,08
		[ũ]	3,24	0,01	0,06
		[uː]	0,06	<0,01	<0,01
		[u̯ː]	0,03	<0,01	<0,01
		[ũ̯ː]	<0,01	<0,01	<0,01
		[ũː]	<0,01	<0,01	<0,01
/j/	4,47	[j]	95,85	0,71	1,29
		[j̃]	4,15	0,01	0,08
/w/	1,84	[w]	99,59	0,28	0,58
		[w̃]	0,41	<0,01	<0,01
/i̯/	1,27	[i̯]	96,02	0,43	0,14
		[ĩ̯]	3,98	0,02	<0,01
/u̯/	0,32	[u̯]	98,35	0,12	0,03
		[ũ̯]	1,65	<0,01	<0,01
Tot.	**100**		**Tot.**	**31,15**	**15,69**

A.6. Porcentajes de aparición de fonemas y alófonos en ataque: porcentaje de aparición del fonema en ataque, y porcentaje global de apariciones del alófono según tonicidad.

Fon.	% apa. ataque	Al.	% del. fonema	% áto. total	% tón. total
/d/	11,75	[ð̞]	77,91	2,95	0,80
		[d̪]	20,67	0,76	0,24
		[ð̞]	0,76	0,03	0,01
		[ð̞ː]	0,66	0,03	<0,01
/t/	10,38	[t̪]	99,83	2,68	1,57
		[tː]	0,12	<0,01	<0,01
		[t̪]	0,05	<0,01	<0,01
/ɾ/	9,50	[ɾ]	100	2,68	1,22
/k/	9,13	[k̟]	41,74	1,29	0,27
		[k̟]	32,57	0,92	0,30
		[k]	25,61	0,58	0,38
		[kː]	0,08	<0,01	<0,01
/s/	9,29	[s]	95,98	2,68	0,98
		[sː]	4,02	0,11	0,04
/l/	8,80	[l]	99,24	2,79	0,79
		[lː]	0,76	0,02	0,01
/n/	7,19	[n]	98,30	2,04	0,86
		[nː]	1,70	0,03	0,02
/b/	6,77	[β̞]	86,06	1,34	1,05
		[b]	13,82	0,22	0,17
		[β̞ː]	0,08	<0,01	<0,01
		[β̞]	0,04	<0,01	<0,01
/m/	6,31	[m]	100	1,51	1,08
/p/	6,21	[p]	99,95	1,83	0,72
		[pː]	0,05	<0,01	<0,01
/θ/	3,81	[θ]	99,94	0,77	0,80
		[θː]	0,06	<0,01	<0,01
/g/	2,66	[ɣ]	80,46	0,47	0,41
		[g]	19,41	0,06	0,15
		[ɣ̃]	0,11	<0,01	<0,01
		[ɣːˈ]	0,02	<0,01	<0,01
/x/	1,77	[x]	64,08	0,23	0,24
		[χ]	35,86	0,17	0,09
		[x]	0,05	<0,01	<0,01
		[xː]	0,01	<0,01	<0,01
		[χː]	<0,01	<0,01	<0,01
/r/	1,78	[r]	98,49	0,49	0,22
		[rː]	1,51	0,01	<0,01
/f/	1,55	[f]	99,83	0,32	0,31
		[fː]	0,17	<0,01	<0,01
/ʎ/	1,02	[ʎ]	100	0,31	0,11
/j̟/	0,89	[j̟]	72,27	0,12	0,14
		[t͡j̟]	27,73	0,04	0,06
/ʧ/	0,68	[ʧ]	99,99	0,19	0,09
		[ʧː]	<0,01	<0,01	<0,01
/ɲ/	0,52	[ɲ]	100	0,13	0,08
Tot.	**100**		**Tot.**	**27,85**	**13,20**

A.7. Porcentajes de aparición de fonemas y alófonos en coda: porcentaje de aparición del fonema en coda, y porcentaje global de apariciones del alófono según tonicidad.

Fon.	% apa. coda	Al.	% del fon.	% áto. total	% tón. total
/n/	34,05	[n̪]	47,76	0,89	1,07
		[n]	20,31	0,52	0,32
		[ŋ]	11,92	0,30	0,19
		[n̥]	6,54	0,15	0,12
		[m]	6,23	0,15	0,11
		[ⁿ]	4,08	0,12	0,05
		[ɱ]	2,14	0,07	0,02
		[ɳ]	0,71	0,01	0,02
		[ɴ]	0,31	0,01	0,01
		[n̈]	<0,01	<0,01	<0,01
/s/	30,16	[s]	47,73	1,53	0,21
		[s̪]	24,00	0,57	0,31
		[s̺]	15,94	0,44	0,14
		[z̪]	12,34	0,39	0,06
/ɾ/	16,45	[ɾ]	100	0,75	1,24
/l/	11,04	[l]	75,55	0,72	0,29
		[l̪]	21,27	0,17	0,12
		[l̺]	2,28	0,02	0,01
		[l̈]	0,91	0,01	<0,01
/m/	3,72	[m]	97,38	0,27	0,17
		[n]	1,62	<0,01	0,01
		[ⁿ]	0,45	<0,01	<0,01
		[n̪]	0,26	<0,01	<0,01
		[ŋ]	0,19	<0,01	<0,01
		[ɱ]	0,06	<0,01	<0,01
		[n̈]	0,02	<0,01	<0,01
		[n̥]	0,02	<0,01	<0,01
		[ɴ]	<0,01	<0,01	<0,01
		[n̻]	<0,01	<0,01	<0,01
/k/	1,54	[ɣ]	100	0,12	0,07
/d/	1,07	[ð̪]	84,28	0,01	0,10
		[ð̺]	15,72	0,01	0,01
/θ/	0,88	[θ]	60,61	0,01	0,05
		[θ̺]	39,39	0,01	0,04
/b/	0,29	[β]	93,69	0,03	<0,01
		[β̪]	6,31	<0,01	<0,01
/p/	0,25	[β̞]	100	0,02	0,01
/g/	0,24	[ɣ]	86,92	0,02	0,01
		[ɣ̞]	13,08	<0,01	<0,01
/t/	0,23	[ð̞]	100	0,01	0,02
/ʧ/	0,03	[ʧ̞]	100	<0,01	<0,01
/f/	0,03	[f]	76,27	<0,01	<0,01
		[f̞]	23,73	<0,01	<0,01
/x/	0,01	[x]	100	<0,01	<0,01
Tot.	**100**		**Tot.**	**7,32**	**4,76**

A.8. Porcentajes de aparición de los distintos alófonos.

Al.	% total	Al.	% total	Al.	% total
[a]	11,776	[ʌ]	0,423	[ɔː]	0,016
[e]	10,083	[ẽ]	0,387	[ɴ]	0,013
[o]	6,610	[b]	0,384	[ɴ]	0,012
[ɾ]	5,885	[eː]	0,334	[rː]	0,011
[s]	5,400	[aː]	0,287	[ẽː]	0,010
[l]	4,592	[l̪]	0,284	[iː]	0,006
[t̪]	4,252	[ã]	0,283	[tː]	0,005
[ð̪]	3,804	[ʧ̞]	0,283	[ɣ̞]	0,005
[n]	3,743	[n̥]	0,269	[ãː]	0,005
[i]	3,549	[j̥]	0,263	[w̃]	0,005
[m]	3,282	[χ]	0,260	[iː]	0,004
[ɔ]	3,044	[ɲ]	0,215	[β̞]	0,003
[p]	2,548	[g]	0,212	[kː]	0,003
[β̞]	2,453	[ⁿ]	0,170	[ĩː]	0,003
[ɛ]	2,205	[sː]	0,153	[ũ̞]	0,002
[j]	2,005	[u̯]	0,148	[β̞ː]	0,002
[n̪]	1,966	[ð̞]	0,145	[t̪]	0,002
[θ]	1,628	[ĩ]	0,115	[uː]	0,001
[k̟]	1,564	[t͡ʃ̞]	0,101	[x]	0,001
[u]	1,364	[õ]	0,098	[pː]	0,001
[k̠]	1,221	[ɱ]	0,089	[fː]	0,001
[ɣ̞]	1,091	[j̃]	0,087	[θː]	0,001
[d̪]	0,996	[ũ]	0,085	[f̞]	0,001
[k]	0,960	[ɛː]	0,078	[u̯ː]	0,001
[i̯]	0,947	[ũ]	0,073	[ĩː]	0,001
[ɑ]	0,894	[ɑː]	0,055	[ɔ̃ː]	0,001
[s̺]	0,875	[ɔ̃]	0,052	[õː]	<0,001
[w]	0,864	[nː]	0,050	[ɣː]	<0,001
[r]	0,718	[θ̞]	0,042	[xː]	<0,001
[u̯]	0,709	[i̯]	0,041	[χː]	<0,001
[f]	0,637	[ð̞ː]	0,032	[n̪]	<0,001
[i̯]	0,591	[oː]	0,032	[ʧ̞ː]	<0,001
[s̺]	0,581	[l̥]	0,030	[ũ̞ː]	<0,001
[ŋ]	0,491	[n̥]	0,029	[ũː]	<0,001
[x]	0,465	[lː]	0,027		
[z̪]	0,450	[i̯]	0,024		

A.9. Porcentajes de aparición de los distintos tipos de sílaba, diferenciando entre sílabas átonas y tónicas, sílaba más frecuente del tipo, y su tonicidad relativa.

Tipo	% apar. total	% apar. átona	% apar. tón.	Sílaba más frecuen.	Ton. rel.
CV	57,45	42,18	15,26	/de/	0,84
CVC	19,82	12,71	7,11	/kon/	1,13
V	6,13	4,35	1,78	/a/	0,92
CSV	3,74	1,59	2,16	/θjo/	1,82
CCV	3,54	2,44	1,10	/tɾa/	0,98
VC	3,34	2,41	0,93	/en/	0,88
CSVC	2,91	0,63	2,28	/θjon/	2,47
CCVC	1,01	0,55	0,46	/tɾas/	1,43
CVS	0,97	0,73	0,24	/taɪ̯/	0,79
CVSC	0,41	0,33	0,08	/beɪ̯n/	0,61
CVCC	0,12	0,08	0,04	/kons/	1,13
VS	0,12	0,06	0,06	/aɪ̯/	1,57
CCSV	0,09	0,03	0,06	/bɾjo/	2,17
VSC	0,08	0,08	<0,01	/aʊ̯n/	0,16
CCVS	0,07	0,05	0,02	/tɾaɪ̯/	0,81
CSVS	0,06	0,05	0,01	/θjaɪ̯/	0,60
CCSVC	0,04	0,01	0,03	/kljen/	2,49
CCVSC	0,03	0,01	0,02	/treɪ̯n/	2,25
CCVCC	0,02	0,02	<0,01	/trans/	0,61
CVSCC	0,02	0,02	<0,01	/seɪ̯ns/	0,44
CSVSC	0,01	0,01	<0,01	/θjaɪ̯n/	1,03
VCC	0,01	<0,01	0,01	/ins/	1,94
CSVCC	<0,01	<0,01	<0,01	/kwaɾt/	1,88
VSCC	<0,01	<0,01	<0,01	/aɪ̯ns/	0,45
CCSVS	<0,01	<0,01	<0,01	/bɾjoɪ̯/	0,79
CCVSCC	<0,01	<0,01	<0,01	/pɾoʊ̯st/	1,56
CSVSCC	<0,01	<0,01	<0,01	/djoɪ̯ns/	1,59
CCSVCC	<0,01	<0,01	<0,01	/stwaɾt/	2,19
CCSVSC	<0,01	<0,01	<0,01	/bɾjoɪ̯n/	0,70
CCSVSCC	<0,01	<0,01	<0,01	/bɾjoɪ̯ns/	1,58

A.10. Porcentajes de aparición de las sílabas más frecuentes del español.

Sílaba	% aparición	Sílaba	% aparición
/de/	3,71	/ma/	1,09
/a/	3,29	/e/	1,02
/la/	2,34	/so/	0,96
/do/	2,17	/me/	0,91
/ke/	2,05	/bi/	0,91
/ɾa/	1,68	/mo/	0,90
/ta/	1,67	/kon/	0,87
/te/	1,59	/pe/	0,81
/se/	1,58	/di/	0,80
/na/	1,58	/ti/	0,79
/no/	1,54	/mi/	0,79
/to/	1,52	/le/	0,76
/ko/	1,47	/po/	0,75
/sa/	1,32	/ɾo/	0,74
/ka/	1,30	/ne/	0,68
/lo/	1,19	/su/	0,67
/ba/	1,17	/re/	0,66
/da/	1,16	/en/	0,65
/pa/	1,15	/es/	0,63
/si/	1,11	/be/	0,63