RECONOCIMIENTO DE LOS NUMEROS CATALANES MEDIANTE SEMISILABAS Y CON LA AYUDA DE UNA GRAMATICA

Màrius Flaquer, Climent Nadeu, José B. Mariño
Depto. de Teoría del Senyal i Comunicacions, U.P.C.

Como herramienta básica se ha utilizado un sistema de reconocimiento para unidades del habla conectadas dependiente del locutor y basado en el ajuste de patrones, sirviéndonos para este ajuste del algoritmo de programación dinámica de un solo paso con restricciones gramaticales [1]. Las unidades elegidas son las semisílabas (ss), entendiendo por ellas cada una de las mitades que surgen de partir una sílaba por su núcleo. Aparecen por tanto ss iniciales *(/ssi/-)* y finales *(/ssf/-)*. Así, el número 4 *(quatre)* fonetizado y silabificado sería *kwA'-trə*, y semisilabificado aparecería como */kwA/--/A'//trə/--/ə/*.

Las ss utilizadas son las que aparecen en los números catalanes del 0 al 1000. Cada patrón de referencia está formado por la parametrización LPC de una señal correspondiente a la realización de una ss. Mientras que para las ssi sólo contemplamos un patrón de referencia, algunas ssf presentan dos e incluso cuatro. Así, se distinguió entre contexto tónico y átono p.ej., *-/iG/ en síG* (5) o en *sıGkwA'ntə* (50)-, y en el caso de las ssf sin coda (acabadas en vocal) y las acabadas en s, también entre contexto interno y final de palabra p.ej., en *sətA'ntə* (70), la primera *-/ə/*, sería interna y la última final-.

Los patrones de referencia se extraen de palabras artificiales o logotomas que incluyen la semisílaba en un contexto que pretende ser máximamente neutro. Esta exigencia de neutralidad contextual determina la forma del logotoma, que en general definiremos bisílabo y de la siguiente manera:

ssi tónica → sílaba tónica y sin coda + *pə* (p.e. *kwA'pə* para */kwA'/-*)
ssi átona → sílaba átona y sin coda + *pA'* (p.e. *trəpA'* para */trə/-*)
ssf tónica interior → *p* + sílaba tónica + *tə* (p.e. *píntə* para *-/ín/*)
ssf átona interior → *p* + sílaba átona + *pA'* (p.e. *pintA'* para *-/in/*)
ssf tónica final → *tə* + sílaba tónica (p.e. *təpíl* para *-/il/*)
ssf átona final → *tA'* + sílaba átona (p.e. *tA'pu* para *-/u/*)

A veces, y para obtener correctamente ciertas referencias, será necesario violar las anteriores reglas (*-énz/* obtenido de *pénzBə* y no de *pénztə*, ...), llegando incluso a logotomas trisílabos para *B* y *D (pəBúpə* → */Bu/-)*.

Diseño de la gramática

Una gramática es un autómata que conoce todas las cadenas de ss de un determinado lenguaje. Cada estado del autómata corresponde a una unidad sintáctica o gramatical distinta. Una unidad fonética determinada (una ss en nuestro caso) puede dar lugar a varias unidades sintácticas diferentes; ello ocurrirá cuando haya que definir conjuntos distintos de unidades predecesoras o posdecesoras. Además, una gramática bien pensada -esto es importante- permite evitar una proliferación del juego de patrones de referencia, soslayando por tanto un aumento de memoria y cálculo.

Para la generación automática de una gramática necesitaremos inicialmente un fichero con el lenguaje fonetizado y semisilabificado de acuerdo con unas normas generales (en nuestro caso el *{/zE/--/É//ru/--/u/,..,/mi/--/íl/}*). Por enumeración del vocabulario se obtiene una gramática inicial, muy redundante, a la que se aplica un algoritmo que consigue una sintaxis equivalente aunque mucho menor. Con todo, este método general no se ha seguido para los números a causa de elevada estructuración de éstos, ya que podemos llegar a la gramática de los números a partir de combinar subgramáticas menores (la de los dígitos, la de las decenas, ..), resultando así mucho más simple el proceso de creación y reducción de una sintaxis.

Con todo, la sintaxis así obtenida -por uno u otro método- aún no es definitiva. Será necesario mejorarla manualmente, mediante la substitución de cadenas que no se darán en la práctica por otras y la adición de transcripciones alternativas de 4 tipos, a veces imbricados:

1) Considerar diferentes pronunciaciones habituales de una misma palabra.

2) Permitir la ausencia de ssf sin coda en posición interior de palabra.

Una útil modificación fué permitir jugar con la duración de las vocales. En general, los patrones de referencia fueron obtenidos de logotomas bisílabos (o sea, con una dicción pausada), mientras que para palabras largas la velocidad aumenta, difiriendo cada vez más las longitudes de las ss de referencia y prueba. Para combatir este fenómeno, al menos en parte, se permitió prescindir de las semisílabas finales sin coda en posición interior de palabra. Así, alternativamente a */sə/--/ə//SA/--/A'n//tə/--/ə//kwA/--/A'//trə/--/ə/* (64) se permitió también */sə/- /SA/--/A'n//tə/- /kwA/- /trə/--/ə/*, además de las posibilidades intermedias entre ambos extremos. Esta táctica se reveló muy útil, apareciendo con profusión para números de 3 o más sílabas, y no dándose jamás para números con menos de 3, como era de esperar.

3) Diferenciación de las ssf sin coda entre interiores y finales.

Para las ssf sin coda, mejor aún que considerar 2 patrones (interior y final) para una misma unidad (fonética), y dado que la sistemática preferencia del algoritmo en los reconocimientos correctos por el patrón adecuado (al contexto) sólo fallaba para algún mal reconocimiento, se decidió contemplar dos unidades distintas, lo que enderezó los anteriores errores.

4) Modelado de nuevos sonidos sin variar el conjunto de patrones de referencia.

Por ejemplo, se observó que para las palabras acabadas en vocal tónica (p.e. 31 → *trEntəú*, incluso *trEntú*), frecuentemente esta vocal era seguida de ella misma pero átona y sensiblemente menos enérgica (es decir, *trEntúu*), dando lugar a una suerte de vocal alargada. Se permitió pues finales del tipo *-/ú//u/--/u/* y *-/ú/ -/u/* junto al hasta ahora sólo contemplado *-/ú/*, apareciendo profusamente en el reconocimiento la segunda alternativa. El mismo tipo de solución permitió utilizar patrones de referencia de diptongos interiores de palabra en el reconocimiento de diptongos finales, evitando así nuevas referencias.

Veamos otro caso. El grupo (vocal + oclusiva + fricativa) -p.e. *sÉtséns*, 700- es esencialmente (vocal interior + silencio corto + fricativa), a causa de la cortísima duración en nuestro caso de la oclusiva *t*. Así, el uso p.e. de *-/Ét/ (təpÉt)*, aún siendo acertado para posición final -reconocimiento de 7 *(sÉt)* y similares- resulta impropio (a causa de la larga *t*) en posición interior (p.e. 700). Se decidió, pues, para la unidad *-/Ét/*, junto al antiguo patrón de referencia *(tapÉt)*, un segundo idéntico a $-/É/_{int}$ *(pÉpə)*; y los mismo para *-/uit/ (búit,buitséns ...)*, conviviendo el viejo patrón *(təpúit)* y el nuevo, idéntico a *-/ui/ (təpuipA')*. Este cambio se demostró totalmente adecuado ya que se reconoció siempre con el nuevo patrón para el caso interior.

Resultados

El sistema tuvo una tasa de reconocimiento del 100% sobre 62 números probados, elegidos para que cada ss aparezca al menos dos veces en un contexto dado. A lo sumo, se detecta alguna semisílaba falsa que no afecta el mensaje del número reconocido por tratarse de una alternativa (p.e. y a causa del ruido final, reconocer *bintiún* cuando se dijo *bintiú*, 21). Porcentualmente, las mejoras más significativas en la eliminación de errores se encuentran en los puntos 1 a 3, mientras que las otras ideas corrigen los errores restantes, pero sobre todo mejoran reconocimientos ya correctos [2].

Referencias bibliográficas

[1] J.B. Mariño, C. Nadeu, A. Moreno, E. Lleida y E. Monte.
"Recognition of numbers and strings of numbers by using of demisyllabes; one speaker experiment". EUROSPEECH'89, París, Septiembre 1989, págs. 215-218.

[2] M. Flaquer i Molina.
"Reconeixement dels nombres catalans del 0 al 1000". Proyecto Fin de Carrera. E.T.S.E.T.B.-U.P.C., Febrero 1990.