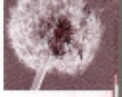

Els sistemes de reconeixement de veu i traducció automàtica en català: present i futur

Mireia Farrús
mfarrus@gps.tsc.upc.es
Jan Anguita
Xavi Anguera
Josep M. Crego
Adrià de Gispert
Javier Hernando
Climent Nadeu
A. Abat
Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (UPC)
Tel. 93 401 10 62

El reconeixement de la parla és el procés que analitza un senyal acústic, capturat per un micròfon o un telèfon, i determina el conjunt de paraules pronunciades per un locutor. Les aplicacions dels sistemes de reconeixement de veu són cada vegada més nombroses i més habituals en la vida quotidiana, com la consulta per telèfon del correu electrònic, la banca electrònica o els productes comercials de dictat de textos, per esmentar alguns exemples.

El reconeixement de la parla és una de les principals àrees d'interès del Grup de Tractament de la Parla (GTP), que, juntament amb el Grup de Tractament de Llenguatge Natural, formen el TALP (Centre de Tecnologies i Aplicacions del Llenguatge i la Parla). El GTP participa activament en la creació de bases de dades orals per al reconeixement de la parla en català, com l'SpeechDat, i una base de dades per a aplicacions de dictat automàtic.

És de vital importància que el català s'adapti als nous àmbits d'ús i, per tant, a les noves tecnologies de la parla. En aquest sentit, cal continuar desenvolupant noves eines basades en sistemes automàtics de reconeixement de veu, veure quines són les prioritats i les mancances actuals en aquest àmbit i definir, d'aquesta manera, les línies futures de desenvolupament.

Recentment, les tecnologies de la traducció automàtica de la parla també són objecte d'un interès creixent en la comunitat científica. Això ve donat tant pel seu atractiu en una societat globalitzada com l'actual, com pel repte que exigeix la integració de tecnologies sovint massa allunyades entre si (com són el reconeixement i el processament de llenguatge natural). En aquesta comunicació també es presenten les línies actuals i futures de recerca que es desenvolupen al TALP en aquest àmbit.

ELS SISTEMES DE RECONEIXEMENT DE VEU I TRADUCCIÓ AUTOMÀTICA EN CATALÀ : PRESENT I FUTUR

M. Farrús, J. Anguita, X. Anguera, J.M. Crego, A. de Gispert, J. Hernando, C. Nadeu
Centre TALP – Departament de Teoria del Senyal i Comunicacions
Universitat Politècnica de Catalunya
{mfarrus, jan, xanguera, jmcrego, agispert, javier, climent}@gps.tsc.upc.es

INTRODUCCIÓ

La visió actual de la societat de la informació gira fonamentalment al voltant de la llengua escrita. No obstant, és evident que la forma més natural i espontània de comunicació entre els éssers humans és la parla, i no precisament l'escriptura. Per aquest motiu, la recerca sobre les tecnologies de la parla ha despertat un gran interès en l'actual societat de la informació.

En aquesta comunicació es fa referència a algunes de les tecnologies de la parla amb més ressò actualment: el reconeixement automàtic de la veu i la traducció oral. La conversió text-parla es tracta en una altra comunicació presentada per membres del nostre Centre [1]. Així doncs, aquí es presenten només les característiques principals del reconeixement de la parla i de la traducció oral. La traducció oral pot fer-se directament a partir de la parla, però aquí considerarem que es tradueix el text que dóna un sistema de reconeixement i a continuació el text traduït es converteix en parla. La traducció de text es pot tractar mitjançant dues aproximacions bàsiques, l'una basada en el coneixement lingüístic (és a dir, en regles), i l'altra en l'estadística. La primera aproximació es presenta en una altra comunicació signada per membres del nostre Centre [2]. En aquesta comunicació ens centrarem només en la segona aproximació, fent referència a l'enfocament estadístic que tenen en comú el reconeixement i la traducció.

En aquesta comunicació es descriu breument l'estat actual d'aquestes tecnologies, concretant-ho en l'àmbit de la llengua catalana, així com les línies futures de recerca que caldria seguir per continuar desenvolupant noves eines en català o millorar les existents. Farem referència a algunes eines o recursos desenvolupats al Centre TALP, (Centre de Tecnologies i Aplicacions del Llenguatge i la Parla), un centre de recerca interdepartamental de la Universitat Politècnica de Catalunya, l'àmbit tecnològic del qual és el tractament automàtic del llenguatge natural, tant en la modalitat oral com en l'escrita.

EL RECONEIXEMENT DE LA PARLA

El reconeixement automàtic de la parla és la determinació del missatge contingut dins un senyal de veu sense la intervenció d'un operador humà. Actualment té un creixent grau d'interès degut als seus avantatges: és la forma més natural de comunicació entre les persones, permet tenir les mans lliures per realitzar altres tasques, és més ràpid que utilitzar un teclat, permet accedir a un ordinador mitjançant la xarxa telefònica, etc. [3]

Els models matemàtics més utilitzats actualment per als sistemes de reconeixement de la parla són els models ocults de Markov (HMM), que *modelitzen* estadísticament tant la pronunciació de les paraules del vocabulari que es poden reconèixer (ho fan connectant els models de les unitats fonètiques elementals: models acústico-fonètics) com la connexió de les paraules per formar frases (models gramaticals o de llenguatge) [4]. Els relativament bons resultats que s'aconsegueixen actualment en aquesta tasca són deguts a la utilització d'eines d'aprenentatge automàtic potents que permeten l'entrenament dels paràmetres dels models estadístics a partir de grans bases de dades, o corpus. Amb bases de dades de veu es formen els models d'unitats fonètiques elementals i amb bases de dades textuals (que poden provenir de material oral) es construeix un model ajustat al domini sintacticosemàntic en què té lloc la comunicació. La figura 1 mostra l'esquema general del funcionament d'un sistema de reconeixement de la parla.

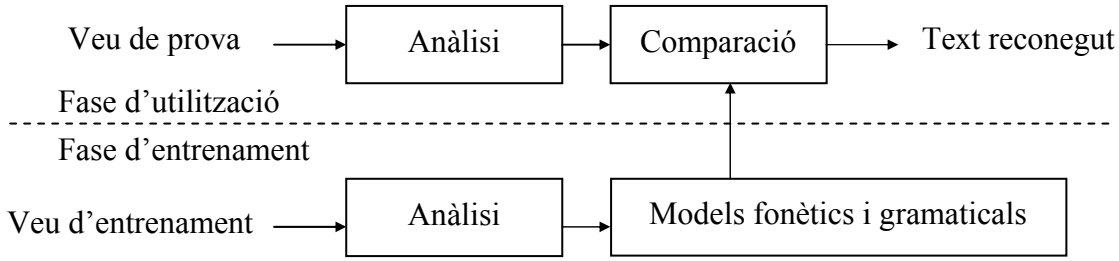

Figura 1. Esquema bàsic d'un sistema de reconeixement de la parla

Un dels principals problemes de la veu és la seva variabilitat. Els senyals de veu corresponents a dues pronunciacions d'un mateix missatge realitzades per la mateixa o per diferents persones poden ser molt diferents. L'etapa d'anàlisi consisteix a extreure un conjunt reduït de característiques de la veu que presentin la mínima variabilitat possible dins del mateix so [5]. En el reconeixement es comparen les característiques de la veu d'entrada amb els models fonètics i gramaticals obtinguts a la fase d'entrenament, i s'escull la seqüència de paraules que presenta la màxima probabilitat.

Com més ben estimats estiguin els paràmetres dels models millor funcionarà el sistema; per aquest motiu, la fase d'entrenament és molt important de cara al bon funcionament d'un sistema de reconeixement de la parla. Per al bon entrenament d'un sistema independent del locutor fan falta bases de dades orals de molts locutors, una per cada llengua de treball.. Aquestes bases de dades són costoses d'obtenir, fins i tot per als sistemes dependents del locutor (és a dir, aquells que només pot utilitzar un locutor determinat), en què una sola persona ha d'entrenar tot un sistema. A la pràctica es poden obtenir sistemes dependents del locutor a partir dels models independents mitjançant tècniques d'adaptació que requereixen poques mostres de la veu del locutor objectiu.

El GTP (Grup de Tractament de la Parla) del Centre TALP participa activament en la creació de bases de dades orals i textuals per al reconeixement de la parla en català, com l'*SpeechDat,* una base de dades de 1005 locutors dels dos dialectes principals del Principat sobre telefonia fixa, amb un corpus dissenyat per a la creació de teleserveis per veu, que conté dígits, nombres naturals, dates, noms i cognoms, noms de ciutats i d'empreses, etc., a més d'un diccionari de pronunciacions. També ha recollit la base de dades oral per entrenar la versió catalana del programa *FreeSpeech* [6], desenvolupat per Philips per a aplicacions de dictat automàtic.

El Centre TALP ha desenvolupat programari per a entrenament i reconeixement, especialment l'anomenat *Ramsès* [7], pensat per a la recerca, creat inicialment en castellà i adaptat posteriorment al català, i també ha generat la tecnologia bàsica del reconeixedor IberVox que comercialitza una empresa amb versions per a cada una de les dues llengües. A part dels sistemes de reconeixement, també s'han desenvolupat eines per a la síntesi de la parla i sistemes de diàleg que engloben tant el reconeixement com la síntesi. Recentment s'ha implementat, per exemple, l'*aTTemps*, un servei en català d'accés a informació meteorològica a través del telèfon [8].

LA TRADUCCIÓ AUTOMÀTICA DE LA PARLA

La traducció automàtica de la parla és tot el procés automàtic capaç de convertir una frase parlada en una llengua de partida en una altra frase amb el mateix missatge, però en una llengua d'arribada. Inclou processos complexos - i encara no resolts - com el reconeixement de la parla, la traducció del text reconegut i la síntesi de la frase generada (veure figura 2), essent una de les tecnologies de la parla més desafiants actualment.

En l'àmbit de la traducció automàtica hi ha dues línies de recerca principals: la traducció guiada per regles i la traducció guiada per dades (o estadística). Al Centre TALP es treballa en el desenvolupament d'un sistema de traducció enterament estadístic, basat en l'anàlisi automàtica de grans quantitats de textos bilingües paral·lels, aprenent els paràmetres del model matemàtic que regeix el procés de traducció [9].

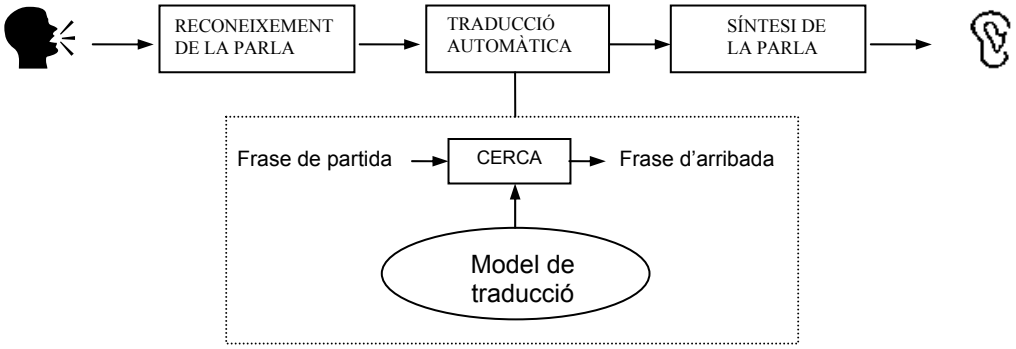

Figura 2. Esquema bàsic del traductor de la parla del Centre TALP

Les claus de la traducció estadística són els algorismes d'estimació del model de traducció i de cerca de la frase idònia, i l'ús de corpus de gran mida a partir dels quals s'estima aquest model. La qualitat i la mida d'aquest corpus o base de dades incideix directament en la qualitat de les traduccions finals.

Al Centre TALP es treballa en la traducció entre català, anglès i castellà. Les bases de dades disponibles inclouen el corpus trilingüe generat al projecte europeu LC-STAR [10], d'àmbit turístic i d'aproximadament 500.000 paraules per idioma, el qual té una part destacada que prové de gravacions de diàlegs orals de caràcter força espontani realitzades al nostre Centre, així com el corpus d'*El Periódico de Catalunya*, que inclou les edicions bilingües d'aquest rotatiu dels darrers anys. Altres bases de dades menors produïdes en el projecte europeu FAME [11] inclouen 10 diàlegs orals sobre reserva d'hotel (transcrits i traduïts al castellà i a

l'anglès), i textos traduïts provinents de bases de dades externes a la UPC, produïdes en el projecte C-STAR [12] en els dominis turístic i mèdic, entre d'altres.

CONCLUSIONS, NECESSITATS I REPTES

El reconeixement automàtic de la parla és el primer pas i probablement el més important cap a la interacció natural home-màquina. La traducció automàtica de la parla se situaria en un pas encara més endavant, en permetre la comunicació entre parlants de llengües diferents a través d'un sistema capaç d'interactuar amb l'home de manera automàtica i natural.

La parla presenta una gran variabilitat. Cada fonema admet moltes realitzacions acústiques, segons els fonemes adjacents, l'accentuació, la identitat, l'estat d'ànim del parlant, etc. D'altra banda, el canal acústic pot introduir pertorbacions al senyal, com soroll ambiental o reverberació de la sala, i el canal elèctric aporta una distorsió addicional (per exemple, la línia telefònica). Però a més a més, la parla té en general un component d'espontaneïtat molt més accentuat que l'escriptura, fet que provoca disfuncions de diferent caire: paraules inacabades o suprimides, frases mal estructurades, recomençaments, sons sense contingut semàntic, etc. [13].

Els sistemes actuals de reconeixement de la veu obtenen uns resultats força satisfactoris, amb taxes d'error en paraules que poden ser inferiors al 5% si es tracta de veu llegida en un entorn acústic favorable, encara que el vocabulari sigui d'unes quantes desenes de milers de paraules, i sempre que es disposi d'un model estadístic de llenguatge ben adaptat al domini semàntic del discurs.Els reptes principals es troben en la parla conversacional, que introdueix un alt grau d'espontaneïtat, i en el reconeixement a partir de veu recollida en micròfons allunyats del parlant. D'altra banda, els models de llenguatge només tenen en compte el context més immediat de cada paraula, de manera que es necessiten representacions més potents, que englobin les frases senceres [13][14].

El problema actual de la traducció de la parla és anàleg al del reconeixement: els sistemes funcionen bé per a situacions controlades i dominis semàntics molt restringits, però a mesura que s'amplia el domini d'ús s'incrementa la dificultat del sistema degut a les múltiples accepcions, ambigüitats, falta d'equivalència directa entre les llengües de partida i d'arribada, etc. La millora dels sistemes de traducció s'ha de centrar en l'ampliació del domini semàntic utilitzat, amb l'objectiu d'aconseguir que els sistemes siguin útils per a un domini obert de la parla.

Pel que fa als sistemes en llengua catalana, els esforços s'han de centrar, principalment, en la creació de bases de dades més completes, tant en grandària com en diversitat d'estil i de domini. En primer lloc, caldria completar l'*SpeechDat,* que disposa de l'ordre de 1000 locutors quan la majoria de bases de dades europees ronden els 5000 locutors. En segon lloc, cal ampliar a altres entorns acústics, com s'ha fet en castellà i altres llengües en els projectes europeus SpeechDatCar [15] (entorn de cotxe) i SpeeCon [16] (cases, llocs públics, etc). No obstant, els corpus anteriors són bàsicament de parla llegida, i cada vegada faran més falta, també en català, els corpus de parla espontània..

Tanmateix, no sembla gaire raonable l'augment continuat –que té lloc actualment– de les bases de dades a fi de capturar tota la varietat de parlants, entorns acústics, estils de parla, etc.

Com a alternativa, s'haurà d'aconseguir a mitjà o llarg termini una adaptació més completa dels models durant l'ús del sistema. Tampoc és convenient haver de tornar a dissenyar el sistema quasi des de la base quan es canvia de llengua o domini semàntic. Els investigadors en aquesta àrea, especialment els que treballen amb llengües minoritzades, són conscients de la necessitat de crear tècniques que siguin fàcilment transportables d'una llengua a una altra i d'una aplicació a una altra.

En resum, tot i que encara som lluny d'aconseguir conversar amb una màquina de la mateixa manera que ho fem les persones, o de parlar amb una persona d'una altra llengua sense moure'ns de la nostra, els primers passos en recerca sobre aquestes tecnologies ja s'han consolidat. Ara el repte és millorar els sistemes actuals de manera que puguin ser factibles i aplicables en qualsevol condició, i no només en entorns i dominis restringits com fins ara. Respecte a la llengua catalana, tot i que en els darrers anys s'han generat els recursos mínims per possibilitar el desenvolupament de sistemes d'accés a informació per telèfon, encara falta produir les bases de dades orals que situïn la nostra llengua al nivell de les llengües europees més parlades pel que fa a aquestes tecnologies.

BIBLIOGRAFIA

[1] J. Adell, P.D. Agüero, A. Bonafonte, H. Duxans, I. Esquerra, A. Moreno, J. Pérez, D. Sündermann. "Els talps també parlen. Línies de recerca en síntesi de la parla al centre TALP", *II Congrés d'Enginyeria en Llengua Catalana,* Andorra, Nov. 2004.
[2] V. Arranz, E. Comelles, D. Farwell, C. Nadeu, J. Padrell. "Sistema de traducció oral per al català, castellà i anglès", *II Congrés d'Enginyeria en Llengua Catalana,* Andorra, Nov. 2004.
[3] D. O'Shaughnessy. "Interacting with Computers by Voice: Automatic Speech Recognition and Synthesis". *Proceedings of the IEEE,* vol. 91, núm. 8, pàg. 1272-1305, Set. 2003.
[4] L. Rabiner, B.H. Juang. *Fundamentals of Speech Recognition.* Prentice Hall, 1993.
[5] C. Nadeu. "Representación de la voz en reconocimiento del habla", *Quark*, núm. 21, pàg. 63-71, 2001.
[6] http://www.upc.es/catala/noticies/acrecerca/1999/freeSpeech.htm
[7] A. Bonafonte, J.B. Mariño, A. Nogueiras, J.A. Rodríguez. "RAMSES: el sistema de reconocimiento del habla continua y gran vocabulario desarrollado por la UPC", *VIII Jornadas de I+D en Telecomunicaciones*, pàg. 399-408, Madrid, Oct. 1998.
[8] http://www.meteocat.com/home/diptic.pdf
[9] A. de Gispert, J.B. Mariño. "TALP: Xgram-based Spoken Language Translation System". Pendent de publicació a: *Proceedings of the International Workshop on Spoken Language Translation*, Kyoto (Japó), Set.-Oct. 2004.
[10] http://www.lc-star.com
[11] http://isl.ira.uka.de/fame
[12] http://www.c-star.org/
[13] J.B. Mariño, C. Nadeu, "La tecnologia digital", *Disseny del futur*, Editorial Proa, pàg. 111-144, 2002.
[14] B.H. Juang, S. Furui. "Automatic recognition and understanding of spoken language - a first step toward natural human-machine communication". *Proceedings of the IEEE,* vol.88(8), pàg.1142-1165, Agost 2000.
[15] http://www.speechdat.org/SP-CAR
[16] http://www.speecon.com/