Llengua i ús
Revista Tècnica de Política Lingüística
Número 48 (2010)

Recerca

La tecnologia de la parla en català. Avenços i reptes

Autor

José Adrián Rodríguez Fonollosa

Universitat Politècnica de Catalunya

L'article brinda un panorama sumari però exhaustiu de la situació de les tecnologies de la parla en català. Una primera part introductòria dóna pas a una relació d'aplicacions i recursos. Finalment, s'aborda el projecte Tecnoparla, amb els mòduls que l'integren, i s'apunten unes perspectives de futur.

Parlant amb les màquines

Dia a dia, la tecnologia ens envaeix. D'una banda, ens ofereix millores i noves possibilitats innovadores fins i tot en els objectes més tradicionals i quotidians. D'altra banda, però, són tantes i tan sofisticades les opcions disponible en els nous electrodomèstics, l'automòbil o el telèfon mòbil, que la seva utilització pot resultar complexa, si no frustrant.

La tecnologia de la parla estudia com utilitzar la veu, la forma més habitual de comunicar-nos entre persones, per facilitar l'ús i la interacció amb les màquines, ja sigui per si sola o en combinació amb altres dispositius com el teclat o les pantalles tàctils.

Les tecnologies fonamentals que permeten aquesta comunicació bidireccional amb les màquines són el reconeixement de la parla i la conversió de text a veu. Convertint en realitat el màgic "Sèsam, obre't" del famós conte oriental *Alí Babà i els quaranta lladres*, el reconeixement de la parla fa possible que les màquines escoltin i executin les nostres ordres i ens permet oblidar-nos del teclat d'ordinadors i mòbils en transformar directament la veu en text. Els sistemes de reconeixement de la parla són aplicacions informàtiques que analitzen el so recollit pel micròfon per extreure'n el contingut textual. Aquesta selecció de les paraules que corresponen a una determinada gravació de veu es fa amb criteris estadístics. En primer lloc, es tenen en compte els models acústics, que modelen els diferents sons que fem servir en parlar, i en segon lloc, el model lingüístic que proporciona informació sobre les possibles paraules que cal reconèixer i la seva pronunciació (vocabulari del sistema de reconeixement), així com una estimació sobre la probabilitat d'una determinada seqüència de paraules o frase (model de llenguatge).

D'altra banda, la conversió de text a veu o síntesi de veu permet la comunicació en sentit contrari en donar a les màquines la possibilitat de llegir-nos qualsevol text i donar-nos així informació de forma oral amb total flexibilitat. Els sistemes de conversió de text a veu realitzen aquesta transformació en dues fases. En primer lloc, la frase és analitzada per extreure la transcripció fonètica corresponent a cada paraula, escollir una entonació global adequada i definir altres característiques sobre el so que cal generar, com la durada de cada fonema i de les pauses. En segon lloc, i una vegada tenim totes aquestes pautes ja fixades, passem a generar el so corresponent.

És important no confondre els missatges de veu gravats prèviament per un locutor amb els missatges orals generats automàticament a partir de qualsevol text amb un convertidor de text a veu. El convertidor de text a veu ens permet escoltar qualsevol text, com ara les notícies d'un diari en línia o un llibre electrònic. També dóna més flexibilitat a dispositius com el navegador GPS, que amb aquesta tecnologia ens pot indicar ara "A la rotonda, pren la sortida cap a Arbúcies", en comptes de la locució fixa "A la rotonda, pren la segona sortida" (que a vegades no és correcta perquè no té en compte sortides noves o auxiliars).

Altres tecnologies complementàries de la parla van més enllà de la informació textual i ens permeten detectar automàticament qui parla (reconeixement del locutor), quina llengua parla (reconeixement de la llengua o el dialecte), el seu estat emocional (reconeixement d'emocions) i si allò que escoltem és veu, música o un altre tipus de so (classificació d'àudio o detecció d'esdeveniments acústics), per posar-ne uns exemples.

Com a mitjà de comunicació amb les màquines, la veu no arriba a ser tan precisa com un teclat, ja que, malgrat els esforços d'investigació fets les últimes dècades, encara no hem aconseguit que les màquines siguin capaces de reconèixer i generar veu amb la precisió i la qualitat de les persones. Això no ha impedit, però, que la tecnologia de la parla sigui ja de gran utilitat per a moltes persones i empreses. En el següent apartat es comenten, sense ànim de ser exhaustius, algunes de les aplicacions d'aquesta tecnologia i la situació del català.

Aplicacions

Sistemes d'informació i gestió telefònica

Cada dia milions d'usuaris són atesos telefònicament de forma automatitzada. Si volem saber el saldo del nostre compte corrent, no hi ha res tan senzill com trucar a un sistema telefònic automatitzat i pronunciar la paraula *saldo*. En les operacions més senzilles i habituals no ens cal que una persona ens faci d'intermediària per aconseguir informació per telèfon. Igual que els caixers automàtics, els sistemes telefònics automatitzats suposen un estalvi de temps i diners, ja que permeten que les persones facin altres tasques més creatives i assessorin millor en gestions més complexes. Les principals empreses de tecnologia de la parla disposen dels productes comercials necessaris per desenvolupar aquest tipus d'aplicacions en català. Però són les empreses que ofereixen el servei les que decideixen finalment en quins idiomes s'ofereix, i en aquest sentit la presència del català encara hauria de millorar bastant.

Sistemes de dictat i control de l'ordinador

La tecnologia de reconeixement de la parla es pot utilitzar també per escriure documents i controlar l'ordinador amb la veu. Encara que és cert que el seu ús no s'ha generalitzat perquè no sempre resulta la forma més còmoda de treballar, hi ha certes tasques i certs col·lectius, com ara els advocats, en què comença a ser habitual la seva utilització per redactar llargs informes amb rapidesa. També els metges, per exemple, poden anar dictant informes o tractaments a una petita gravadora que després es connecta a l'ordinador per generar-ne la transcripció automàtica amb un sistema de reconeixement de la parla. Les últimes versions dels sistemes operatius Windows de Microsoft incorporen ja aquesta tecnologia de reconeixement de la parla en alguns idiomes sense

cost addicional, i també inclouen un convertidor de text a veu. Malauradament, ni Microsoft ni cap altra empresa ofereixen actualment un sistema de dictat en català.

Transcripció, cerca i traducció de documents audiovisuals

Dia a dia, els documents audiovisuals van guanyant pes als documents escrits en àmbits com l'oci, l'educació i la cultura. Les noves tecnologies n'han multiplicat la producció diària i en faciliten la difusió, però per a l'espectador o usuari cada vegada és més necessari poder seleccionar i filtrar amb criteris propis. La tecnologia de la parla és la clau per poder subtitular i classificar aquests documents de forma automàtica o semiautomàtica. Combinada amb la traducció automàtica, aquests subtítols els podem generar també en qualsevol altre idioma per facilitar-ne la difusió internacional. En l'apartat dedicat al projecte Tecnoparla comentarem el desenvolupament d'una aplicació d'aquest tipus en català.

A l'automòbil

Quan tenim les mans ocupades perquè conduïm o fem algun tipus de treball, la veu pot ser la forma òptima d'interactuar amb les màquines. S'ha comprovat que conduint és molt més segur controlar el telèfon mòbil, el GPS o el climatitzador amb la veu que no pas amb les mans. El català disposa en aquest àmbit dels recursos lingüístics adequats i també de la tecnologia bàsica, però la presència al mercat és encara escassa. S'ha de tenir en compte, no obstant això, que parlem de productes recents desenvolupats inicialment en només tres o quatre idiomes.

Als mòbils i les videoconsoles

Empreses com Nintendo, Sony i Microsoft incorporen ja reconeixement de veu en alguns jocs i també comencen a sorgir aplicacions més ambicioses com la traducció de veu (d'un idioma a un altre) en temps real. Malgrat tot, la presència de la tecnologia en català en el món dels videojocs comercials és nul·la, si bé és cert que ni tan sols es doblen o se subtitulen al català.

Altres aplicacions recents de la tecnologia de la parla en els mòbils és el dictat de missatges SMS amb la veu o la conversió a text dels missatges rebuts a la bústia de veu, o bé al contrari, poder escoltar amb el telèfon i un convertidor de text a veu els missatges de text o el correu electrònic. Algunes companyies com Vodafone i Microsoft han anunciat que oferiran aquests serveis també en català.

En l'aprenentatge d'idiomes

Mitjançant la tecnologia de la parla adequada, les màquines poden esdevenir també un professor de llengua i ensenyar-nos la pronunciació adequada de les paraules o l'entonació correcta d'una frase. Qualsevol sintetitzador de veu de qualitat en català pot servir per escoltar un text i tenir-ne així una referència bastant bona, i, encara que és més complicat, també es pot intentar detectar errades fonètiques de pronunciació amb un sistema de reconeixement.

Accessibilitat

En les persones amb alguna discapacitat auditiva, visual o de la parla aquesta tecnologia pot ser vital no tan sols per interactuar amb les màquines, sinó també amb altres persones. El reconeixement de

la parla permet generar subtítols automàticament per a les persones sordes, mentre que la conversió de text a veu permet a les persones amb deficiències visuals transformar el text d'una pantalla o un missatge en veu, o donar veu a persones amb dificultats greus per parlar amb claredat.

La innovació i l'automatització són constants en la societat moderna i no hi ha cap dubte que la tecnologia de la parla continuarà ajudant a desenvolupar noves aplicacions d'utilitat social i a millorar les existents. Cal seguir investigant per millorar la qualitat, però també hem de ser conscients que moltes vegades és tan important disposar d'una bona tecnologia com fer-la servir de forma adequada. No fos cas que ens passés com a Alí Babà i ens quedéssim tancats dins la cova perquè no recordem les paraules màgiques.

Recursos lingüístics

Comunicar-nos amb la veu ho fem des de molt petits, i potser per això podem arribar a pensar que no hauria de ser tan complicat que els potents ordinadors actuals fossin capaços de reconèixer les nostres paraules. Però el cert és que les característiques del so recollit pel micròfon en pronunciar una paraula determinada varien enormement en funció de la persona i de l'entorn en què ens trobem. No existeix el reconeixedor de veu universal, i cada tipus d'aplicació requereix el desenvolupament d'un producte específic per obtenir el rendiment òptim.

Per a cada entorn (aplicacions telefòniques, dictat a l'ordinador, dictat de SMS, videojocs, documents audiovisuals) i per a cada llengua calen també recursos lingüístics específics. En el cas del reconeixement de la parla aquests recursos lingüístics i aquests desenvolupaments específics requereixen una forta inversió, ja que s'han de recollir i transcriure centenars d'hores de gravacions orals de milers de persones. Aquestes bases de dades o corpus orals han de cobrir adequadament no tan sols la variabilitat de dialectes, accents, edats i sexes, sinó també la variabilitat d'entorns (telefonia fixa, telefonia mòbil, oficina, automòbil, ràdio i televisió, etc.). A més, també poden ser necessaris grans corpus textuals, de milers de milions de paraules, per obtenir models estadístics adequats del llenguatge.

Afortunadament, en el cas de la conversió de text a veu no hi ha tants problemes. Una mateixa tecnologia de síntesi i una mateixa veu pot servir per a gairebé tots els entorns i aplicacions, tot i que també cal adaptar-se a les capacitats del dispositiu que hagi de funcionar.

Tant en reconeixement com en síntesi de la parla, el català es troba actualment en una situació relativament bona pel que fa a recursos lingüístics i textuals. Els últims quatre anys, el centre TALP de la Universitat Politècnica de Catalunya, amb el suport econòmic de la Generalitat de Catalunya, ha completat el conjunt de recursos lingüístics i textuals necessaris per desenvolupar tecnologia de la parla en català. D'una banda, en reconeixement de la parla s'han recollit bases de dades orals en els entorns més habituals: telefonia, oficina, automòbil i televisió. I d'altra banda, en síntesi de veu s'han gravat noves veus de gran qualitat.

A més, aquests recursos lingüístics s'han fet servir per millorar els sistemes de reconeixement de la parla i de síntesi de veu en català, i per desenvolupar aplicacions que mostrin la utilitat d'aquesta tecnologia, com ara el sistema de subtitulació i traducció automàtica de veu a veu que es descriu al següent apartat.

És important destacar finalment que tots aquests recursos lingüístics i una gran part d'aquesta tecnologia està disponible de forma gratuïta per a qualsevol empresa, centre de recerca o persona interessada.

El projecte Tecnoparla

Fa tres anys va néixer el projecte Tecnoparla amb l'impuls i el finançament de la Secretaria de Política Lingüística de la Generalitat de Catalunya i amb la idea de potenciar el desenvolupament de recursos lingüístics i de tecnologia de la parla en català de nivell internacional, i facilitar-ne així la utilització en tots els àmbits. En aquest sentit, en el projecte es marquen dos objectius. En primer lloc, es persegueix un objectiu general: dur la tecnologia de la parla en català al nivell d'altres llengües europees en tots els entorns d'interès. En segon lloc, es tria desenvolupar una aplicació concreta que mostri la utilitat de la tecnologia desenvolupada: la subtitulació i la traducció automàtica de continguts audiovisuals.

Com ja hem comentat, l'interès d'aquesta aplicació és clar si es pensa en la gran quantitat de produccions audiovisuals disponibles en múltiples idiomes, bé creades de forma professional per a cinema, televisió o ràdio, bé creades per particulars i posades a l'abast de qualsevol persona a través d'Internet. La subtitulació i la traducció automàtiques faciliten l'accessibilitat i la difusió internacional d'aquests continguts audiovisuals.

Anàlisi, subtitulació i traducció de programes de televisió

El sistema desenvolupat s'ha dissenyat específicament per a programes televisius de notícies o de debat, i inclou moltes tecnologies per extreure tota la informació possible de l'àudio:

- En primer lloc, es fa una classificació d'àudio per detectar si parla un locutor o tenim un altre tipus de so (música, silenci, veu telefònica, etc.). En els següents mòduls es faran servir només el segments d'àudio on es detecti la veu d'un locutor.

- En segon lloc, la tecnologia de reconeixement de locutors ens permet detectar automàticament quants locutors diferents tenim en un programa i quin d'ells parla cada vegada.

- En tercer lloc, detectem en quina llengua es parla (castellà, català o anglès).

- En quart lloc, tenim el sistema de reconeixement de la parla, corresponent a la llengua detectada, que ens proporciona la subtitulació en l'idioma original i també informació addicional d'utilitat per al traductor i el sintetitzador de veu.

- A continuació podem generar els subtítols en altres idiomes amb el traductor automàtic.

- I, finalment, podem fer servir la conversió de text a veu per escoltar els subtítols traduïts. Aquest últim pas inclou tecnologia específica de sincronització i de conversió de veu amb l'objectiu que la veu sintètica traduïda soni sincronitzada amb la imatge i amb un to de veu semblant al del parlant original.

A continuació es descriuen amb una mica més de detall tots aquests mòduls i altres contribucions del projecte Tecnoparla.

Mòdul de classificació de l'àudio

Per aconseguir una transcripció completa i automàtica de documents audiovisuals és imprescindible realitzar en primer lloc una classificació de l'àudio per detectar en quin moment es parla.

Per evitar transcripcions absurdes amb grans errors, el reconeixedor de la parla ha d'actuar només quan es parla i no hi ha música de fons o és a penes perceptible. Per a programes televisius de debat o documentals s'ha desenvolupat un classificador de segments sonors capaç de detectar música, sorolls, silenci, veu de bona qualitat o veu telefònica, així com la presència de música de fons sobre la veu. Aquesta classificació és útil també per a persones sordes.

Mòdul de detecció de locutors

Com a informació addicional a la transcripció textual dels segments de veu, el sistema desenvolupat identifica, a més, quin locutor parla cada vegada. Per obtenir aquesta informació, el mòdul de detecció de locutors segmenta automàticament les intervencions de cada convidat al programa, detecta les veus diferents que hi intervenen i agrupa els segments corresponents al mateix locutor.

Mòdul d'identificació d'idioma

En els programes televisius analitzats en el projecte (programa de debat *Àgora* de TVC), la major part d'intervencions són en català, però no és estrany que hi hagi també convidats que s'expressin en castellà durant el debat. Per això cal detectar l'idioma, per escollir el sistema de reconeixement corresponent. Aquesta informació permet obtenir també estadístiques sobre l'ús de cada llengua de forma automatitzada.

Sistema de reconeixement de la parla

El centre TALP de la Universitat Politècnica de Catalunya, juntament amb la Universitat d'Aquisgrà, hem desenvolupat un sistema de reconeixement de notícies radiofòniques i debats televisius en català amb una qualitat semblant a l'aconseguida pels millors sistemes de reconeixement de la parla en altres llengües europees. El sistema incorpora els mètodes més avançats d'anàlisi i modelatge de la parla, i s'ha dissenyat per obtenir la màxima qualitat possible amb documents audiovisuals prèviament gravats i treballant amb vocabularis d'unes 100.000 paraules. El sistema és de codi obert i de lliure ús per a activitats de recerca, està totalment documentat i permet desenvolupar amb facilitat models acústics i de llenguatge adaptats a altres tipus d'aplicacions o entorns.

Traductor estadístic català-castellà (-anglès)

Fins fa pocs anys els sistemes de traducció comercials es basaven fonamentalment en unes regles lingüístiques de transformació que eren el resultat d'una anàlisi morfològica i sintàctica clàssica.

Avui en dia, tanmateix, són els denominats *sistemes estadístics de traducció* els que proporcionen més qualitat en moltes aplicacions, com ara el traductor de Google i el desenvolupat pel centre TALP en el marc del projecte Tecnoparla. Aquests sistemes estadístics són capaços d'aprendre a traduir a partir d'exemples de traduccions, els anomenats *corpus paral·lels*. Un avantatge addicional en el context de l'aplicació de traducció de veu desenvolupada és que aquests sistemes estadístics treballen millor que els basats en regles amb parla espontània o amb textos amb errors deguts al mateix locutor o al sistema de reconeixement de la parla.

El traductor desenvolupat està disponible en línia a http://www.n-ii.org i ofereix una gran qualitat especialment en la traducció de notícies entre català i castellà, ja que ha estat entrenat amb les edicions bilingües d'*El Periódico de Catalunya*.

Conversió de text a veu

En aquesta tecnologia s'ha treballat amb dos objectius. El fonamental ha estat aconseguir un sistema de conversió de text a veu en codi lliure de gran qualitat en català i que, a més, estigués fàcilment disponible per a les principals distribucions Linux (FestCat). Un altre objectiu ha estat la millora de les tècniques d'adaptació o conversió de veu que faciliten i abarateixen el desenvolupament de veus personalitzades.

FestCat

Els últims anys, s'ha continuat desenvolupant el convertidor de text a veu en català FestCat incorporant-hi noves veus de gran qualitat amb bases de dades més grans, de deu hores, i resolent els problemes computacionals que comporten. També s'hi ha incorporat una veu infantil. A més, s'ha desenvolupat una versió del sistema de conversió de text a veu per a dispositius amb poca memòria com ara mòbils i navegadors GPS.

Finalment, s'ha integrat el sistema desenvolupat en el sistema públic Festival i s'ha creat un repositori per facilitar la instal·lació de les veus en diverses distribucions de Linux.

Reptes de futur

Els últims anys, l'avanç tant en recursos lingüístics orals en català com en tecnologia bàsica ha estat important. Ara tenim el repte de facilitar-ne l'ús i també d'animar i ajudar les empreses a portar aquesta tecnologia a la societat. I no em refereixo només a les empreses tecnològiques, sinó també a les empreses de serveis. I és que a vegades una determinada tecnologia sí que està disponible comercialment, però no acaba d'arribar a la societat perquè no és utilitzada de manera habitual per desenvolupar aplicacions o serveis en català.

Per analitzar aquests reptes, novament hem de distingir entre conversió de text a veu i reconeixement de la parla.

En conversió de text a veu la situació actual del català és bona tant pel que fa a la disponibilitat comercial com al programari lliure per a l'entorn Linux (FestCat). A més, al centre TALP treballem per oferir-ne una versió gratuïta també per al sistema operatiu Windows i altres entorns.

Pel que fa a productes i serveis, la situació no és tan bona, però ja podem trobar uns quants serveis telefònics d'informació automatitzats amb tecnologia de la parla en català, i també algun exemple al carrer, com els missatges de veu generats automàticament amb un convertidor de text a veu en català que fan servir alguns autobusos de Barcelona. A més, també el podem fer servir sense restriccions a l'ordinador personal, per exemple amb les aplicacions informàtiques per llegir la pantalla de l'ordinador que utilitzen les persones amb deficiències visuals.

Pel que fa al reconeixement de la parla, la situació del català no és tan bona i encara hi trobem bastants llacunes no cobertes pels sistemes comercials. El català està disponible comercialment per a aplicacions telefòniques, que són les que més impacte econòmic han tingut tradicionalment, però en altres entorns la presència del reconeixement del català és pràcticament nul·la.

També és cert que molt poques llengües disposen, per exemple, de sistemes de dictat i control de l'ordinador de bona qualitat. El sistema operatiu Windows només ofereix el sistema de reco-

neixement de la parla en sis idiomes: anglès, francès, espanyol, alemany, italià i xinès; mentre que l'altre producte comercial de qualitat i d'ús general disponible comercialment, el Dragon Naturally Speaking de l'empresa Nuance, només està disponible en holandès, anglès, francès, alemany, italià i espanyol. Tampoc existeix, en cap llengua, un sistema equivalent en codi obert o de lliure distribució, simplement perquè no es tracta d'una tasca senzilla ni barata de desenvolupar i mantenir.

Amb la finalitat de contribuir a cobrir aquests forats de la tecnologia en català, des del centre TALP col·laborem ja en projectes de recerca a mitjà i llarg termini amb diversos grups d'investigació i empreses nacionals i internacionals. El sistema de reconeixement de la parla en català i el sistema de traducció que ja hem desenvolupat en el marc del projecte Tecnoparla és un bon punt de partida, però caldrà incrementar notablement la inversió privada i també mantenir la pública per poder adaptar el sistema a cada tasca concreta i arribar a la societat amb productes de qualitat.

Esperem que les empreses i les diferents administracions s'adonin de la rendibilitat econòmica i social de potenciar el desenvolupament i l'ús de tecnologia de la parla en català.