



Att lära sig språk med en virtuell lärare

Språkbehärskning delas ofta in i fyra kunskapsområden: Tala, lyssna, läsa, och skriva. Dessa fyra kan i sin tur delas in i två undergrupper: produktion - perception, och muntligt - skriftligt. Alla är naturligtvis viktiga, men om man måste prioritera dessa delar av språkbehärskning får nog den muntliga företräde framför den skriftliga. Självfallet beror det på vad man vill använda språket till, men en person klarar sig mycket bättre som analfabet med en väl utvecklad muntlig kommunikationsförmåga än tvärtom. Muntlig kommunikation - öga mot öga - är trots allt vår primära kommunikationskanal. För en person som inte behärskar detta medium tillräckligt väl blir någon form för utanförskap ofta resultatet. Detta gäller för svenskar med någon form för språkligt handikapp, men också för invandrare.

Att börja med att lära sig prata, för att sedan lära sig läsa och skriva är dessutom det mest naturliga sättet att lära sig ett språk. Detta gäller både modersmålet och andra språk. En person som ska lära sig ett nytt språk bör därför i de flesta fall tidigt lära sig att få ett bra uttal med rätt betoning, rytm och satsmelodi.

Om man däremot ser på hur mycket tid som läggs ned på olika färdigheter i språkundervisningen, samt var ansträngningarna läggs på forskning i "Computer assisted language learning" (CALL), finner man att en övervägande del av resurserna går till den skriftliga delen. Detta är förmodligen inte på grund av att man ser det muntliga som oviktigt, men för att det i en klassrumssituation är svårt och tidskrävande att arbeta med tal som kräver individuell instruktion och återkoppling, och att det i CALL sammanhang krävs helt andra utvärderingsmetoder än vad som traditionellt har utvecklats. Att skapa en CALL-applikation med fokus på uttal och muntlig kommunikation känns därför som en både angelägen och berättigad utmaning.

På Centrum för talteknologi (CTT), vid institutionen för Tal, musik och hörsel på KTH i Stockholm, arbetar vi sedan våren 2004 på en ny typ av språkinlärningsprogram som det här kapitlet kommer att handla om (Engwall Wik Beskow Granström 2004). Projektet utgår från en vision om en typ av program som i stor utsträckning använder sig av en personmetafor i stället för den vanliga skrivbordsmetaforen. (dvs. att man ser på gränssnittet till datamaskinen som analogt till en person istället för ett skrivbord med mappar, filer, en papperskorg etc.) Tänk dig ett program som innehåller en virtuell lärare, en datoranimerad tredimensionell person som man kan prata med, och som pratar tillbaka. Den virtuella läraren ger dig möjlighet att träna samtal på det nya språket. Han korrigerar ditt uttal om du gör misstag, och han ger dig individuell återkoppling på det du har sagt eller gjort. Den virtuella läraren lär känna dig bättre ju mer du använder programmet och noterar dina förbättringar, vilket ger honom möjligheten att skraddarsy lektionerna efter

dina personliga behov och eventuella svagheter. Efter några timmars träning kostar språkläraren en bråkdel av vad en privatlärare skulle kosta, och han är alltid tillgänglig på din dator - dag och natt, när du själv har lust att träna.

Hans tålamod är oändligt, och du slipper genomlida genanta situationer som kan uppstå i en klassrumssituation när du säger något felaktigt. Inte nog med det, han kan göra en hel del saker som är omöjligt för en människa att klara av, som till exempel att göra delar av ansiktet transparent för att visa hur tungan ska positioneras när du uttalar ett problematiskt ord. Detta var visionen - idag finns naturligtvis inget dataprogram som inom en överskådlig framtid skulle kunna ersätta en välutbildad och entusiastisk språklärare, men man måste börja någonstans.

Bakgrunden till vårt arbete är alltså dels en vision om hur det skulle kunna bli, men också en strävan efter att nyttiggöra tal- och språkteknologi inom språkinläringen, något som inte är så utbrett inom branschen som man skulle kunna förvänta sig. Det är inte så att man inom talteknologin inte har sysslat med språkinlärningsfrågor tidigare. Tvärtom har man haft ett flertal konferenser på temat, och där faktiskt vi i Sverige, i regi av CTT arrangerade den första av dessa, på Marholmen i Stockholm 1998 - Speech Technology in Language Learning (STiLL). Den har sedan följts upp ungefär vartannat år som en serie workshops.

Talteknologi har potential att bli en mycket värdefull bidragsgivare i CALL-sammanhang, eftersom det är med kunskaper från detta forskningsområde man har möjlighet att utvärdera och ge återkoppling på en students uttal. Det är inte en lätt uppgift, och delar av det här kapitlet kommer att diskutera problemställningar man ställs inför när det är muntligt kommunikation i stället för skriftspråk som ska behandlas.

VILLE - Den virtuella läraren

I ett tidigt skede av projektet bestämde vi oss för att försöka göra en så universell språklärare som möjligt, och inte bara ett program för en typ av användare eller för ett språk. Vi separerar därför de språkspecifika delarna av programmet från de språkoberoende delarna, och skiljer de användarspecifika delarna från de mer universella. Samtidigt som detta har legat till grund för hur arkitekturen av den virtuella språkläraren ska utformas, har vi sett det som nödvändigt och viktigt att börja i någon ände, med en specifik typ av användare och ett specifikt målspråk. Den första versionen av den virtuella språkläraren undervisar vuxna invandrare i svenska och har fått namnet Ville.

Vem har svårt för vad? – Lingvistiska kunskaper

Som svensk kan man ofta höra vilket land eller språkområde en person kommer från trots att de pratar svenska. Klassiska brytningar som till exempel svenska med engelsk accent, eller svenska med fransk accent, kan de flesta svenskar urskilja, och med ett tränat öra kan man öka på listan. En persons modersmål påverkar med andra ord hur deras uttal och grammatik utvecklas i ett andraspråk. Det finns naturligtvis individuella skillnader, men man kan tydligt se generella trender såsom att italienare har svårt med vissa aspekter av svenska, och japaner har svårt med andra.

Språk varierar med avseende på vilka ljudkontraster som är betydelsebärande i språket. I till exempel det indiska språket Hindi finns det en betydelsebärande skillnad mellan

aspirerat och oaspirerat k som vi saknar i svenskan. Det vill säga, /kal/ och /khal/ är två olika ord. Ett annat klassiskt exempel är /l/ och /r/. I svenskan är det en betydelsebärande skillnad mellan /l/ och /r/. Vi kan därmed på svenska göra minimala ordpar som bara skiljer sig med avseende på denna kontrast såsom /låda–råda/ eller /lapa–rapa/. Så är inte fallet för japanska och flera andra asiatiska språk, där det bara finns ett ”l/r”-ljud. Som en generell regel kan man säga att det är svårt för en person att lära sig ljudkontraster i ett nytt språk som saknar motsvarighet i personens modersmål.

I svenskan finns det flera ljudkontraster som många tycker är exotiska och svåra. Till exempel har svenskan ett kvantitetssystem med s k komplementär distribution, i folkmun ofta kallat lång och kort vokal. Skillnaden mellan /ful – full/ eller /sil-sill/ är exempel på detta, en lång betonad vokal följs av en kort konsonant och tvärt om. Personer med t ex kinesiska eller spanska som modersmål har ofta mycket svårt att både höra och producera denna skillnad korrekt. Skillnaden mellan de svenska tonerna - accent I (akut accent) och accent II (grav accent), som till exempel *tanken* (behållare) och *tanken* (tänka) är ett annat exempel på en kontrast som finns i svenskan och som många som ska lära sig svenska har svårt för, men brukar vara lätt för kineser som har ett mer komplicerat tonspråk som modersmål.

Ett sätt att ta reda på vilka aspekter av ett målspråk som en student med en viss språklig bakgrund sannolikt kommer få svårigheter med, är att göra en *kontrastiv fonologisk analys*. Det innebär att man jämför vilka betydelsebärande ljudkontraster som finns i en persons modersmål och i målspråket. Man kan på så sätt ofta predicera vilka svårigheter en person kommer att få när han/hon ska lära sig ett nytt språk.

Detta är exempel på lingvistisk kunskap som vi försöker att dra nytta av i designen av den virtuella läraren. Med hjälp av sådana förkunskaper ska den virtuella språkläraren kunna skapa en kursplan som är anpassad till individuella svårigheter, och skapa en undervisningssituation som är skraddarsydd till varje individs behov.

Akustisk analys

Den grundläggande akustiska analysen i Vile görs med hjälp av ett fritt tillgängligt ljudanalysprogram, Snack, som utvecklats vid CTT (Sjölander 2004). Det finns i Snack en mängd metoder för att analysera den akustiska signalen. Med Snack kan man få fram akustiska parametrar såsom till exempel stämbandston, intensitet (ljudstyrka) och spektrumanalys för utvalda delar av en inspelning. Vile har alltså en mängd akustiska parametrar att tillgå, men dessa måste tolkas på något sätt för att vara av värde i ett språkundervisningssammanhang.

Ett annat viktigt verktyg i Vile kallas *Styrd segmentering*. Det är en teknik som gör det möjligt att hitta tidsgränserna för varje del av ett yttrande. Detta används på flera olika sätt i Vile. Genom att till exempel mäta längden av den betonade vokalen och den efterföljande konsonanten i ett yttrande, och sedan jämföra dessa med någon form av referens, kan Vile ge återkoppling på hur studenten behärskar lång och kort vokal. Det metoden gör är att på bästa möjliga sätt tidsmässigt anpassa ett yttrande till en transkription över vad som sagts. En begränsning med denna teknik är att man måste veta vad som blev sagt, och det innebär att det finns många situationer där metoden inte passar. I en vokabulärövning till exempel, där man vill testa om en student kommer ihåg ett ord,

kan man inte använda sig av segmenteringsmetoden. Om studenten säger något helt annat än det han förväntades säga, kommer metoden fortfarande att försöka passa in de segment som den förväntas hitta, trots att de kanske inte finns där.

Uttalsanalys:

En av de viktigaste målsättningarna i vårt arbete med att skapa en virtuell språklärare har varit att utveckla metoder som klarar av att detektera typiska uttalsfel som språkstudenter gör. Vilka detektorer som behövs kan till viss mån härledas från den kontrastiva analys av ljudkontraster som beskrevs i förra avsnittet. Man kan tänka sig en lång lista av detektorer som klarar av olika typer av kontraster i ett språk, och även satsmelodiska och rytmiska aspekter av tal.

I valet av vilka detektorer som ska utvecklas först, måste en prioritering göras av vilka uttalsfel som kan anses som viktigast ur en kommunikativ synvinkel. (Vissa uttalsfel upplevs av infödda talare som mer störande än andra, och dessa är därför viktigare att kunna detektera och rätta till.) Alla typer av uttalsfel är inte lika vanliga, och man bör därför rimligtvis lägga ner mer möda på att utveckla en detektor som kan hjälpa många studenter jämfört med en som bara används av några få. Sedan finns det en del uttalsfel som är lättare än andra att rätta till för eleven. Slutligen finns det även en pragmatisk avvägning eftersom alla typer av detektorer inte är lika lätta att göra.

Några detektorer vi redan experimenterat med är: *duration*, *ordbetoning*, *vokalkvalitet*, *vokalinskott* och *vokalreduktion*. Durationsdetektorn som vi nämnt tidigare och som används för att lära ut t ex skillnaden mellan /sil-sill/) använder styrd segmentering och mäter längden på den betonade vokalen och den efterföljande konsonanten i studentens uttal. Dessa tider jämförs sedan med en referens. Om tiderna avviker mer än ett visst tröskelvärde får studenten en återkoppling som informerar om detta.

Ordbetoningsdetektorn används för att mäta vilken stavelse i ett ord som har betonats (till exempel me-lo-**di**). Den är också baserad på styrd segmentering, samt extraktion av stämbandston och röststyrka. Här jämförs studentens yttrande med ett facit med information om vilken stavelse i ett ord som ska vara den mest prominenta.

Vokalkvalitet är viktigt att kunna ge återkoppling på eftersom svenskan har en relativt stor uppsättning av vokaler jämfört med de flesta språk. Detta innebär att flera av de svenska vokalerna, t ex U, Y och Ö skapar problem för många. Vokalkvalitet är dock svårare att utvärdera än många andra aspekter av tal, eftersom den mätning av vokalspektrum som ligger till grund för vokalkvaliteten måste normaliseras, på grund av att människor är olika stora.

Vokalinskott och vokalreduktion förekommer ofta bland studenter som lär sig svenska som andraspråk. Det innebär att konsonantkluster (såsom till exempel SPR i språk, eller SKR i skratta) görs om genom att antingen lägga till någon eller några vokaler (från till exempel /skratta/ till /esekratta/) eller genom att ta bort någon eller några av konsonanterna (från /språk/ till /råk/). Var dessa inskott sker är ganska förutsägbart, och en inskottshypotes samt reduktionshypotes kan användas i kombination med styrd segmentering för att se om det är sannolikt att t ex ett vokalinskott har förekommit.

Lektioner i programmet

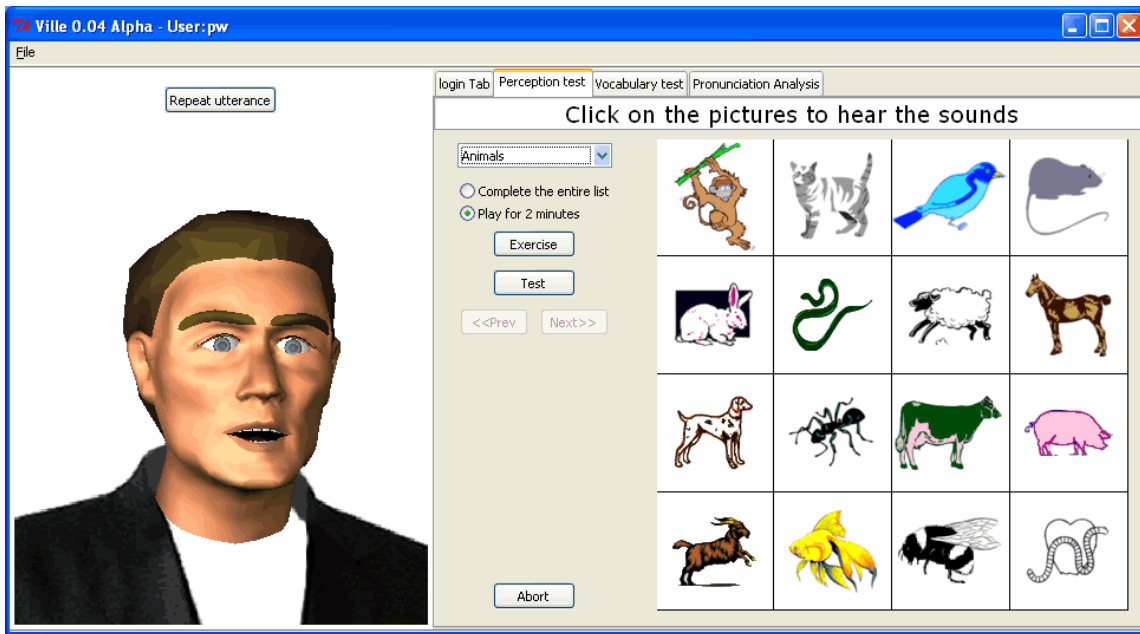
Perceptionsövningar

Det första steget för att lära sig uttala en ny ljudkontrast går normalt genom hörseln. Det vill säga, om man inte kan urskilja ett karakteristiskt drag kan man heller inte producera det. Ville har därför både *perceptionsövningar* och *produktionsövningar*.

Perceptionsövningarna innehåller inte så mycket talteknologiska metoder som produktionsdelen, men är förmodligen ur pedagogisk synvinkel minst lika viktig. Övningarna är skapade för att belysa specifika ljudkontraster som man av erfarenhet vet är svåra för många. En övning går till så att ett minimalt par som till exempel /glas-glass/ blir presenterat på skärmen och Ville säger slumpvis ett av orden. Studentens uppgift är att identifiera vilket av orden som sagts och klicka på det rätta ordet. Ville ger sedan återkoppling på studentens val. Man kan variera svårigheten genom att t ex börja med överdriven kontrast för att sedan gå mot en mer naturlig duration. Eftersom studenten i dessa övningar inte pratar, utan gör sina val genom att trycka på knappar, har man också möjligheten att kartlägga och ge korrekt återkoppling på studentens handlingar med 100% säkerhet, (något som inte är lika lätt i produktionsövningarna).

Vokabulärövningar

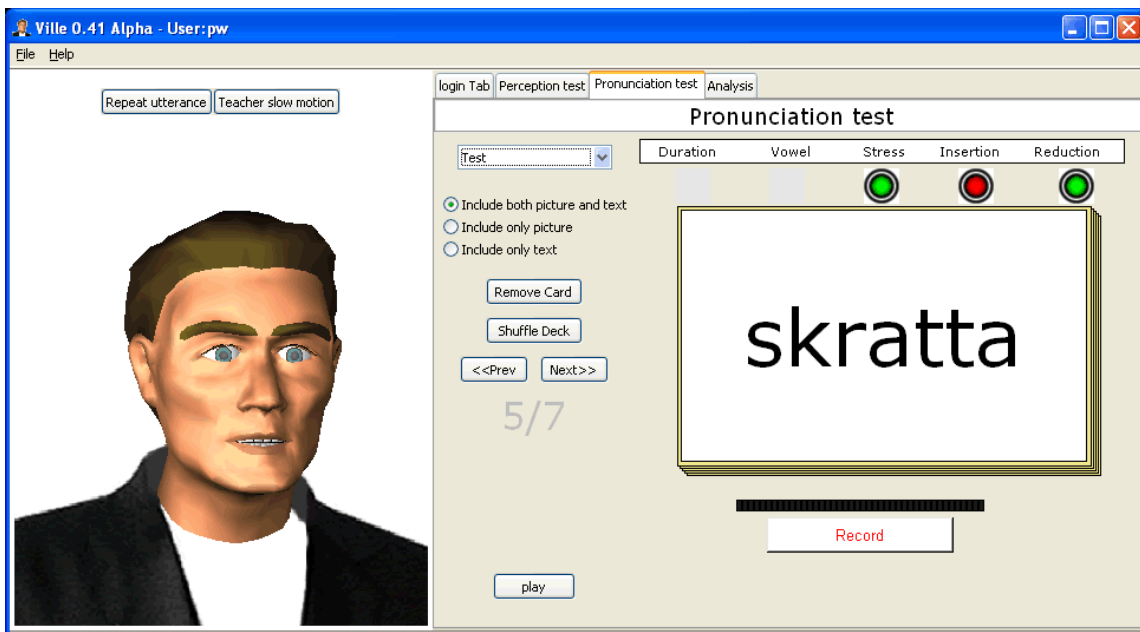
Vokabulärövningar är också en del av Villes repertoar. Samma princip som i ljudkontrastövningarna kan tillämpas för att träna på nya ord. Bildobjekt är grupperade i olika kategorier (till exempel djur, frukter, kroppsdelar) och presenteras slumpmässigt på skärmen. När man trycker på en bild uttalar Ville ordet. Vill man sedan testa sina kunskaper gör man en övning där Ville slumpmässigt säger ett av orden på skärmen, och studentens uppgift blir att klicka på den bild som korresponderar med ordet. I alla perceptionsövningar kan studenten välja att antingen gå igenom alla orden i en vald lista, eller se hur många han/hon hinner gå igenom på en förutbestämd tid, (till exempel en minut). Man har också ett val mellan övning och test. "Övning" innebär att Ville ger återkoppling på om studenten valt rätt eller fel, och väljer man fel får man försöka igen tills man hittat rätt. "Test" innebär att man bara får en chans per ord och att Ville inte kommenterar om huruvida man valt rätt eller fel, utan endast redovisar antal rätt och fel vid slutet av testet.



En skärmbild från vokabulärövningsdelen, där 16 djur presenteras slumpmässigt på skärmen.

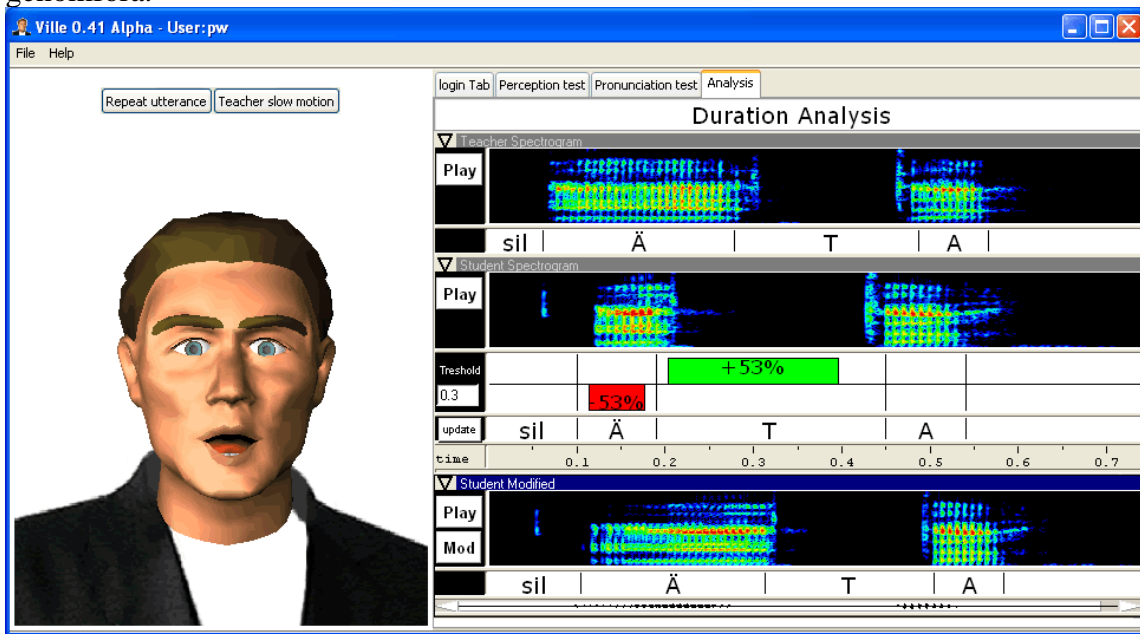
Produktionsövningar

Här har studenten möjlighet att själv tala och få sitt uttal analyserat av Vile. Även här är lektionerna indelade i olika feluttalstyper, så att en lektion till exempel fokuserar på kvantitet (lång och kort vokal) medan en annan lägger fokus på ordbetoning. Gränssnittet till denna del av programmet kan liknas vid en kortlek, där varje ord ligger på ett eget kort (s.k. flash-cards). Klickar man på det översta kortet så uttalar Vile ordet. Trycker man på inspelningsknappen och säger det som står på kortet kommer Vile utföra en uttalsanalys av yttrandet.



En skärmbild från produktionsdelen av programmet med tre detektorer aktiva.

Vi har i denna version av programmet valt att ge visuell återkoppling i form av röda eller gröna cirklar för de detektorer som är aktiva. Vi hade till att börja med en verbal återkoppling från Ville men den upplevdes efter en stund som lite tjatig. Det är inte lätt för en virtuell lärare att veta när det är passande att ge en viss sorts information och när det är bäst att hålla tyst. (Vi återkommer till frågan om lämplig återkoppling i ett senare avsnitt). Fördelen med ett visuellt återkopplingsystem är att studenten snabbt får en översikt över hur hans yttrande blev graderat. Om studenten önskar mer detaljerad information om varför han fick en röd cirkel kan han fördjupa sig i felets natur genom att klicka på den, och då komma till en ny sida med en mer detaljerad beskrivning. (se figur nedan). Om studenten däremot har ett återkommande fel och redan har tagit del av denna information, kan han helt enkelt reflektera över det: ”javisst ja, jag ska ha en kort vokal i det ordet” och sedan försöka igen, utan att behöva lyssna till information som efter några genomgångar kan verka överflödig. Möjligheten att ge information på ett diskret sätt utan att avbryta studenten i hans arbete är något våra teststudenter har uppskattat. Det är ju inte så att verbal återkoppling nödvändigtvis är den bästa i alla lägen. Kanske detta är ytterligare ett exempel på att Ville ibland har vissa fördelar jämfört med en mänsklig lärare. Man kan ju se för sig en mänsklig lärare göra samma sak, och vifta med olika flaggor som representerar olika typer av fel, men i praktiken är det nog svårt att genomföra.



Figuren visar hur analysen av durationsdetektorn blir presenterad för studenten när han önskar mer detaljerad information. Den översta tredjedelen visar lärarens (Villes) yttrande i form av ett s.k. spektrogram, och en tidsmarkering av varje fonem i ordet (äta). Om man vill kan man lyssna på yttrandet genom att klicka på den översta play-knappen. Den mittersta tredjedelen är ett spektrogram av studentens eget yttrande där de ljud som är i fokus (vokalen /Ä/ och den efterföljande konsonanten /T/ i denna övning) är markerade med en röd eller grön rektangel. Den nedersta tredjedelen visar en tidsmodifierad version av studentens yttrande, där man förändrar längden på studentens /Ä/ och /T/ så att de liknar lärarens yttrande. Att låta studenten höra sin egen röst med ”rätt uttal” är något vi tror kan hjälpa studenten att lättare korrigera sitt uttal. Detta är ytterligare ett exempel på hur talteknologiska hjälpmedel kan vara till nytta i uttalsundervisningen.

Det talande huvudet

En annan mycket central del av programmet är naturligtvis det animerade tredimensionella huvudet som personifierar den virtuella läraren. Teknologin som är utvecklad vid CTT (Beskow, 2003) ger möjligheten att synkronisera tal med munrörelser. Talet kan antingen skapas syntetiskt (text-till-tal), eller komma från färdiginspelade ljudfiler. För att språkstudenterna ska få en bra förebild till sitt eget uttal, har vi hittills valt att spela in alla Villes yttranden, istället för att använda syntetiskt tal. Tal är multimodalt, det är lättare att förstå då man både ser och hör talare. Dessutom kommunicerar vi med våra ansikten mycket utöver det verbala. De talande huvudena kan också röra på andra delar av ansiktet, såsom till exempel ögonen och ögonbrynen. Detta används för att signalera t ex uppmuntran, bekräftelse, ifrågasättande, turtagning och vad som är viktigt i det man säger.

Framtidsutsikter och reflektioner

Uppenbart fattas det fortfarande en hel del i den virtuella läraren, men som framgår av ett test vi beskriver i slutet av kapitlet är de studenter som provat på den nuvarande versionen mycket positiva till programmet. Här vill vi peka några områden som kräver mer utveckling, och som kommer att göra den virtuella läraren mer användbar.

Taligenkänning

De som överhuvudtaget hört talas om talteknologi associerar det ofta med taligenkänning och talsyntes. När talteknologi ska appliceras på CALL är det ofta taligenkänning man i första hand tänker sig ska användas. Taligenkänning som den vanligtvis används kan vara ganska problematisk tillsammans med CALL av flera skäl.

Som människor är vi otroligt flexibla i vår mänskliga taligenkänning, och lyckas ofta adaptera till stora individuella variationer så att vi förstår vad någon säger, trots att det skiljer sig avsevärt från normen. ”Jak inte fesstå va do säga” jaha, han menade ”Jag förstår inte vad du säger”. En taligenkännare söker igenom sin språkmodell efter ord den har lagrat. Antingen blir det rätt eller så blir det fel. Säger man saker lite fel (vilket ju en andraspråksinlärare ofta gör) så är risken att taligenkännaren inte accepterar det och man får ’fel’ som återkoppling, eller så blir igenkänningen rätt trots uttalsfelet, och man får ett resultat som inte heller är tillfredställande, eftersom man ju önskar någon form av återkoppling på vad man gjort fel. Att få en taligenkännare att värdera uttalsfel på samma sätt som en människa är dock fortfarande ett olöst problem. Gör man absolut inga fel så pratar man redan så bra svenska att man inte behöver programmet, och gör man fel så fångas de inte upp av taligenkännaren. För en som precis har börjat lära sig ett språk är automatisk taligenkänning egentligen inte vad man önskar, utan hellre automatisk feluttalsigenkänning.

Det finns dock andra, kanske mer oortodoxa sätt att använda taligenkänning på, som vi önskar utforska som en del av den vidare utvecklingen av Ville.

Den kan till exempel användas tillsammans med de uttalsanalysmetoder som beskrivits ovan. Taligenkänningen skulle kunna ligga i bakgrunden och övervaka uttalet. Om taligenkänningen ger en hypotes på vad som blev sagt som överensstämmer med det som förväntades, så säger det något om uttalets kvalitet. Om det som sagts är förståeligt för en taligenkänning, kanske det betyder att det också är det för en människa? Om

taligenkännings resultat under en längre tids interaktion överensstämmer med det förväntade, kan man tänka sig att det betyder att studenten klarat av första delen av språkkursen, och pratar så pass bra svenska att man kan försöka sig på svårare, mer komplexa övningar där man har någon form av interaktiv kommunikationsträning med Ville.

Dialogsystem

Dialogsystem används idag framför allt i vissa typer av telefontjänster som till exempel SJ's röststyrda bokningssystem. Man får då frågor upplästa av typen: "Var vill du åka?" och "När vill du åka?", och man svarar med naturligt tal "Till Göteborg" eller "I morgon".

En taligenkännare tolkar yttrandena och de tilldelas sedan en semantisk tolkning så att systemet vet vilken typ av information som har getts och kan utifrån kontexten bestämma vilken typ av fråga som är lämplig för att föra dialogen vidare. CTT har en lång tradition av att bygga dialogsystem och vi tittar nu på möjligheterna för att använda dialogsystem även som ett språkinlärningsverktyg. Det innebär att en student och Ville skulle kunna föra ett samtal inom en begränsad domän, lite i samma anda som de klassiska parlörerna. Teman såsom "på restauranten", "på sjukhuset", "på tågstationen" osv. skulle kunna simuleras så att studenten får möjlighet att träna på situationer som han/hon med stor sannolikhet kommer att ställas inför. Man kan även se på hela Villes interaktion med studenten som ett dialogsystem på en metanivå, alltså samtal med *läraren* Ville i domänen språkundervisning, inte Ville när han låtsas vara receptionist eller kypare.

Lektionsplanerare

Eftersom det i den nuvarande versionen av Ville finns få möjliga lektioner, serveras allt innehåll på ett smörgåsbord, så att studenterna själva kan välja vad de vill träna på. Vi strävar efter en utveckling där Ville är kapabel att ta större ansvar och kontroll över lektionerna så att personmetaforen får större slagkraft i interaktionen. För att komma dit måste Ville innehålla en väl utvecklad dialoghanterare, men även kunna fatta en hel del beslut kring vilka övningar som är mest gynnsamma för just den här studenten i just den här fasen av hans språkutveckling. En lektionsplanerare kan också ge innehållsleverantörer (språklärare, andra pedagoger och fackkunniga personer) möjligheter att utforma en läroplan oberoende av programmet.

Studentprofil

Genom att logga studenternas interaktion med Ville, kan vi skapa en individuell profil över studentens interaktionshistoria. Lektionsplaneraren måste basera sina beslut på bland annat studentens svårigheter och framsteg. Sådan information är även värdefull för att kunna ge relevant återkoppling.

Återkoppling

En av de viktigaste egenskaperna hos en lärare, vare sig den är virtuell eller reell, är att ge korrekt och relevant återkoppling till studenterna.

En pedagogisk och samvetsgrann lärare använder förmodligen all sin intelligens, känslighet och kreativitet i sin återkopplingsstrategi. Det är något som många lärare använder åratals på att utveckla och förbättra och som vissa lärare lyckas bättre med än

andra. En av de största utmaningarna i konstruktionen av den virtuella läraren är därför att utveckla modeller som på ett trovärdigt sätt återspeglar de komplexa processer som ligger till grund för en bra lärares val av återkoppling. I likhet med andra egenskaper hos den virtuella läraren kan man bygga ut och förbättra återkopplingen stegvis. Den mest grundläggande delen av återkopplingen - som för en mänsklig lärare kanske kan te sig som trivial, - innebär att signalera "rätt" när studenten gör något rätt, och "fel" när studenten gör något fel. Det är oftast inget enkelt problem för en dator.

Uppgiften liknar problemställningar man stöter på i många andra forskningsområden. Svårigheten ligger i att bara påpeka de fel som faktiskt har begåtts (precision), samtidigt som man inte ska låta några fel passera obemärkt (recall). Analogt, i till exempel ett sökprogram på Internet (t ex Google) innebär perfekt precision och recall, att man hittar varken mer eller mindre än det man önskar hitta. Man önskar att bara relevanta länkar dyker upp och att inga relevanta länkar saknas. Är man tveksam om en länk är relevant eller inte är det förmodligen bättre att ta med den än att utelämna den. I en inläringssituation är det tvärtom förmodligen bättre att fria än att fälla.

Återkoppling från en lärare fungerar som ett styrmedel för studentens framsteg, och tar ofta långt mer komplexa former än detta. Det handlar om att variera beröm, uppmuntran, och kritik i lagom doser. Frekvens, intensitet och typ av återkoppling beror på en mängd faktorer och är i högsta grad kontextberoende. Återkopplingen är beroende av om uppgiften bedöms som svår eller lätt – antingen i absoluta termer (globalt) eller för just en student (lokalt). Man ger mer beröm till en student som äntligen lyckas tillägna sig något han har haft svårt för, än om det samma utförs av någon första gången de försöker. Återkoppling varierar även mellan individer eftersom olika människor reagerar olika på kritik/beröm och behöver därför bemötas olika för att fungera optimalt. Ibland kan återkopplingen tonas ner på grund av tidigare återkoppling som redan getts på samma företeelse, medan man ibland måste förstärka återkopplingen av precis samma skäl. Hur den här typen av komplext resonerande skall kunna modelleras i en virtuell lärare på ett trovärdigt sätt är en utmaning.

Inläring och spel:

Inläring kan ske på grund av olika bakomliggande motiv. En önskan om integrering – att vara och tala som en infödd – kan vara en motivationsfaktor som driver, i alla fall många barn, att sträva efter perfekt uttal. En mer instrumentell motivation – att kunna använda sig av språket för att kunna uträtta saker, är en annan typ av motivation som får som konsekvens att andra aspekter av språket lyfts fram som viktiga. Man kan också motiveras att lära sig något endast för lärandets skull osv. Motivationen som driver någon att i timmar i sträck, spela ett dataspel, utan att någon direkt nyttofunktion är kopplad till de färdigheter som utvecklas under spelets gång, är exempel på ett beteende som verkar drivas av en annan typ av motivation. I ett väldegnat spel är det glädjen av att klara av uppgiften som står framför en som driver lärandet framåt, och sen spelar det ingen roll om det är till synes "meningslösa" kunskaper man tillägnat sig. Denna, låt oss kalla den uppgiftsrelaterad motivation, är spelutvecklare väl medvetna om. Motivationen är att komma längre än man gjorde förra gången och där det finns en relation mellan svårighetsgrad och motivation. Är spelet för svårt blir det frustrerande och tråkigt, och är

det för lätt blir det också fort tråkigt. Man måste misslyckas/förlora en viss andel av sina försök för att upprätthålla lusten att spela vidare, och spelets svårighetsgrad måste öka proportionellt med att spelarens skicklighet ökar. Det ska vara lätt att komma igång med spelet, men svårt att bemästra. En låg ingångsnivå och en hög utgångsnivå ger förhoppningsvis upphov till många timmars spelglädje. Detta är faktiskt precis de förutsättningar som gäller för språkinläring. Om man kunde kombinera en sådan uppgiftsrelaterad motivation som finns i spel med ett inlärningsprogram som Ville, skulle språkkunskaperna på sätt och vis kunna komma som en sidoeffekt av spelandet. Genom att leka, och spela spel blir inläringen lustfylld och därmed förhoppningsvis också mer effektiv. Eller som Åsa Moberg uttrycker det: Lustpunkten är en punkt inne i huvudet som lärandet måste nå om inläringen ska ta fart (Moberg 2006).

Normerat tal – Vad bör normen vara?

I skriftspråket finns det mycket tydligare gränser för vad som är rätt och vad som är fel än vad det finns i talspråket. Det finns många dialektala variationer som avviker mer från rikssvenskan än vad en utlandsbrytning kan göra. Dessa avvikelser räknas ändå som acceptabla medan en utländsk brytning kanske inte gör det. För en språklärare kan det vara svårt att bestämma sig för vilka normer som ska sättas för acceptabelt uttal. Är det angeläget att man som vuxen invandrare ska lära sig prata svenska med så lite brytning som möjligt? Är det förståelighet man önskar uppnå eller 'perfekt svenska'? Är det en myt att vuxna inte kan lära sig 'perfekt uttal'? Att veta vilka variationer som uppfattas som avvikande och vilka som inte gör det är ingen lätt uppgift.

Många andraspråksinlärare hävdar dessutom att de inte *vill* bli av med sin brytning. Att deras identitet på något sätt är bunden till deras uttryckssätt och uttal. Olle Kjellin hävdar att detta för de flesta är en bortförklaring, och ett försvar mot att behöva erkänna att man inte kan så bra som man egentligen skulle vilja. En brytning kan hindra kommunikation, förståelse, vänskap och integration. En stark brytning gör kanske att lyssnaren bara hör felen och inte ser människan bakom dem. (Kjellin 2002)

Att lära sig rätt uttal är lite som att spela teater, något som barn ofta klarar av galant utan en tanke på att det skulle vara pinsamt att göra sig till för att låta som en infödd. Som vuxen däremot blir det ofta så att man behåller det rytm och intonationsmönster man tillägnat sig från sitt modersmål, med resultatet att man får en brytning i det nya språket. Kjellin och flera andra hävdar att huvudskälet till att vuxna andraspråksinlärare ofta får en brytning inte är biologiska hinder på grund av hjärnans fossilisering eller att någon kritisk period har passerats. I stället kan det vara oviljan att sätta sig i en situation som kan upplevas som pinsam i kombination med brist på vägledning i t ex prosoditräningen (rytm och satsmelodi).

En av de kanske viktigaste potentiella fördelarna med ett dataprogram som Ville, är därför att det ger personer möjlighet att träna på de aspekter av det nya språket som till en början känns ovant och pinsamt, i en miljö som inte alls känns hotande på samma sätt som ett klassrum kan kännas. Till och med hos en privatlärare kan det för många kännas lite genant att träna på detaljer i uttalet.

Ett praktiskt prov med Ville

Under sommaren 2006 hade Ville sin första testperiod med riktiga studenter genom ett samarbete med språkenheten vid KTH. Språkenheten ansvarar för språkundervisning av såväl svenskar som ska lära sig ett främmande språk, som utländska studenter som ska lära sig svenska. De hade påbörjat en Web-baserad kurs i svenska och behövde ge uttalsundervisning som en del av programmet.

Studenterna som valdes ut till att delta i försöksprojektet var alla masterstudenter från utlandet som inte hade fått plats på de lärarledda klassrumsbaserade språkkurserna. De var unga, motiverade, välutbildade, och tekniskt kunniga, personer. Studenterna hade en språkligt varierande bakgrund, de kom från Peru, Ryssland, Spanien, Pakistan, Grekland, Kina, Bangladesh, Indien och Ghana.

Man kan inte förvänta sig några stora förändringar av en students uttal efter en månads självstudier på sommaren. Försöksprojektet sågs därför först och främst som en möjlighet att testa och få återkoppling på programmets funktionalitet och robusthet. Vi önskade naturligtvis ändå att undersöka om vi kunde få några mätbara förbättringar av studenternas uttal efter den stipulerade träningsperioden. Studenterna fick därför göra ett diagnostiskt test både före och efter sin träningsperiod med Ville. Testet bestod av 100 utvalda ord som studenterna skulle spela in. För testet användes ett gränssnitt som liknade programmet studenterna skulle använda hemma, men de 100 ord som skulle spelas in fanns inte med i materialet de fick med sig hem. Programmet som användes under testet hade heller inga verktyg för analys av studenternas uttal, utan var i princip bara en 'bandspelare' med det animerade huvudet bredvid.

Studenterna fick sedan med sig en CD med programmet och ett headset med hörlurar och mikrofon. Instruktioner i form av en manual fanns att tillgå på projektets hemsida (Wik 2006)

Utvärderingen av försöket är inte färdig när detta skrivs, men av de kvalitativa frågorna från intervjuerna efter projektets slut framstår försöket som mycket positivt. Här följer några oretuscherade omdömen:

"Pedagogical but doesn't force you to study. But I lived in Sweden for 10 months and I didn't study Swedish. With this program I got the possibility to train Swedish, and the accent."

"I would never have studied Swedish if it wasn't for the program."

"Not so big as I expected. Despite its not so big I haven't managed to do all the exercises."

"My kid started to play with it also, especially with the animals."

"Good to work at home with accent spelling etc."

"Very useful to see the mouth expression."

"Gives a feeling of someone helping you."

"Useful so you can see how to pronounce the articulation of the words."

"I Would like to be able to interact, converse with Ville."

Det verkar som om det animerade ansiktet har en förmåga att ge användare intrycket av intelligens, trots att ingen egentlig intelligens finns. En av studenternas kommentar var att det kändes som om det fanns någon där som hjälpte till med studierna. Ett gränssnitt som kan skapa sådana förväntningar hos användarna är naturligtvis positivt när det fungerar, men kräver en noggrann avvägning för att inte bli en besvikelse om man misslyckas med att leva upp till de förväntningar man skapat. Under våra försök med de utländska masterstudenterna på KTH, var det flera försökspersoner som kommenterade reaktioner hos Ville, som överhuvudtaget inte existerade. Flera studenter sa till exempel att de hade observerat hur Ville hade rynkat på ögonbrynen när de hade uttalat något ord fel. Om detta är fördelaktigt eller ej ur en pedagogisk synvinkel återstår att se när programmet exponeras för en större publik.

Referenser:

Engwall Olov, Wik Preben, Beskow Jonas, Granström Björn (2004) Design strategies for a virtual language tutor, ICSLP 2004, vol 3, 1693-1696

Beskow, Jonas (2003). Talking Heads - Models and Applications for Multimodal Speech Synthesis. Doctoral dissertation, KTH

Wik, Preben (2006). "Ville - The Virtual Language Tutor" .Tillgänglig på [<http://www.speech.kth.se/ville/>]

Sjölander, Kåre. "The Snack Sound Toolkit" Tillgänglig på [<http://www.speech.kth.se/snack/>]

Kjellin, Olle. (2002). Uttalet, språket och hjärnan. Teori och metodik för språkundervisningen. Uppsala: Hallgren & Fallgren Studieförlag AB. ISBN 91-7382-756-8.

Moberg, Åsa. (2006), Lustpunkten : en bok om inläring utan utbildning, Natur och Kultur, ISBN: 9127083071