

Neurala korrelerat till fyllda pauser

En fMRI-studie av disfluensperception

Robert Eklund

Biträdande Professor i Språk, Kultur och Kommunikation
Institutionen för Kultur och Kommunikation, Linköpings Universitet
& Docent i Datorlingvistik, Institutionen för Datorlingvistik
Linköpings Universitet
<http://roberteklund.info>

Artikeln presenteras här i förkortat skick. Hela artikeln finns på www.rostforum.se

Introduktion

Mänskligt, spontant producerat talspråk kännetecknas av att inte vara helt "flytande". Jag sätter ordet inom citationstecken eftersom det råder delade meningar om huruvida "oflyt" i själva verket underlättar såväl talproduktion som talperception. Den vanligaste termen för detta är disfluenser, men även denna term är inte helt etablerad. En annan sak att hålla i minne är att den alternativa stavningen dysfluenser förekommer, speciellt engelskspråkig litteratur.

Disfluenser har studerats i över ett sekel, och en introduktion följer nedan. Denna artikel redovisar resultaten från en unik fMRI-studie av den mest "speciella" av de olika disfluentstyperna, det som ofta(st) benämns "fyllda pauser", som (i svenska) "eh" eller "öh". Notera att även denna term inte är etablerad.

Icke-patologiska disfluenser

Ungefär 6 % av spontant tal utgörs av något slags disfluens, dvs vi "stakar oss" på ett eller annat sätt på (i snitt) ungefär vart 20:e ord. Denna siffra har visat sig vara tämligen stabil mellan de språk som varit huvudsakligt fokus för disfluensstudier.

Formella studier av disfluenser påbörjades på 1930-talet med som huvudsakligt syfte att reda ut orsaker till och prediktion av stamning, men studerades de kommande decennierna, med andra syften, inom andra discipliner, och med banbrytande studier främst inom psykoterapi.

Disfluenser har även studerats inom en stor mängd andra discipliner som neurovetenskap, psykologi, sociolingvistik, filosofi, medvetandefilosofi, teckenspråk, språkteknologi, paralingvistik, tvåspråkighetsforskning, biologi, etc. För en omfattande redogörelse för och sammanfattning av forskningen upp till år 2004 hänvisas till Eklund (2004:51–171).

Rörande exakt vilka språkliga fenomen som räknas som disfluenser så varierar även detta inom littera-

turen, kanske främst rörande huruvida företeelser vilka i andra sammanhang går under namnet diskursmarkörer, som *va*, *asså*, *ba'*, liksom etc skall inkluderas under paraplytermen disfluenser. Eftersom samtliga sådana fenomen har "riktiga" ord som bas – de kan t ex slås upp i ordböcker – har jag valt bort sådana företeelser, och räknar enbart följande fenomen, med ungefärliga proportioner av det totala antalet disfluenser angivna i procent (siffrorna är tagna från Eklund (2004): ofyllda pauser (tystnader, UP), >50 %; fyllda pauser, ~25 %; prolongationer (segmentförlängningar), ~6 %; trunkerar (avhuggna ord), ~6 %; feluttal, ~1 %; och "explicit redigering", som "hopp-san", "nej, det var fel" (etc), <1 %.

Fyllda pauser

Som vi såg ovan så utgör fyllda pauser (FP) den vanligaste formen av vokaliserade disfluenser. Ofyllda pauser undviks tämligen ofta i analyser eftersom det är svårt att dra en gräns för när en tystnad i spontant talspråk, "spontalt", faktiskt är en disfluens. Dessutom används ofyllda pauser ofta på ett medvetet, retoriskt sätt, där durationen i sig kan utgöra ett medvetet funktionellt grepp. Mitt favorit-exempel ges av ståuppkomikern Jerry Seinfeld, som konstaterade att "mitt-tystnaden" i

frågekonstellation "Could I ask you a favour ... [tystnad i N sekunder]... could you X", där X är den tjänst man ber om och N är tystnadens duration. Ju högre N, desto större X. Lingvistiskt mycket insiktsfullt, anser jag.

FP utgör ungefär en fjärdedel av samtliga disfluenser, vilket innebär att ungefär 3,5 % av alla "ord" vi yttrar är ett "eh". En sammanställning av FP-frekvens, på ordnivå, presenteras i Tabell 1.

Tabell 1: FP-frekvens i olika källor, angivet om procent av det totala antalet ord. Tabellen är en sammanställning av tabeller i Eklund (2010). Legend: H–M: Människa–Maskin-konversation. H–H: människa–människa-konversation. "Push-to-talk": talarna pratade inte förrän de var redo, och tryckte på en knapp alldeles innan de började tala, och hade alltså antagligen förberett sitt tal i förväg. WOZ: Wizard-of-OZ: en datainsamlingsmetod där man låter försökspersonerna tro att de pratar med ett automatiskt system (en dator), vilket i själva verket utgörs av en eller flera människor som "bakom draperiet" låtsas vara systemet; för en utförlig beskrivning av denna metod, se Eklund (2004:179–180); Allwood och Haglund (1992).

Källa	FP/ordnivå	Korpusstyp
Maclay & Osgood (1959)	3.9 %	Konferensdata
Laljee & Cook (1969)	2.8 % 3.4 %	"High pressure" (hög stressnivå) "Low pressure" (låg stressnivå)
Laljee & Cook (1973)	1.9 % 4.3 %	Kvinnor Män
Lutz & Mallard (1986)	3.6 %	Konversation
Eklund & Shriberg (1998)	3.4 % 2.6 % 0.3 % 3.0 %	H–M WOZ H–H H–M "push-to-talk" H–H
Bortfeld et al. (2001)	2.6 %	Konversation
Eklund (2004)	3.0 % 4.1 % 2.2 % 4.4 %	H–M WOZ H–M WOZ H–H H–M
Moniz, Mata & Viana (2007)	2.3 %	Presentationer
Eklund (2010)	7,5 %	Autentiska kundsamtal till Telia 90 200 med riktiga ärenden. Talarna var intemedvetna om att de spelades in.

Som visas i Tabell 1 så är FP-frekvensen tämligen stabil i olika studier. De två resultat som sticker ut är den låga siffran 0,3 % i Eklund och Shriberg (1998), vilken kan förklaras med att just den korpusen utgjordes av "push-to-talk"-talspråk, dvs försökspersonerna pratade inte förrän de "var redo", och först då tryckte på en knapp för att tala, samt den höga siffran 7,5 % (Eklund, 2010), som intressant nog är den enda av de rapporterade korpusarna som baseras på taldata där försökspersoner dels inte utförde artificiella uppgifter (typ experiment) utan autentiska ärenden, dels inte var medvetna om att samtalen spelades in.

Men det riktiga intressant med FP är att det tidigt visade sig att FP uppvisar såväl en egen distribution som beteende jämfört med alla andra disfluentstyper. Under de senaste decennierna har ett antal olika hypoteser rörande orsaken till och/eller funktionen för FP föreslagits:

Floor-holding hypothesis

Maclay & Osgood, (1959) var antagligen först ut med att föreslå att FP kan användas för att förhindra att man blir avbruten av en annan talare, dvs man "behåller golvet" i en konversation, även om man vid vissa tillfällen inte har nästa ord omedelbart färdigt för leverans. Det är helt enkelt svårare att avbryta någon som säger "eh" än någon som blir helt tyst. Denna hypotes föreslogs även av Livant (1963).

Help-me-out hypothesis

En alternativ hypotes som går ut på motsatsen är att "eh" utgör ett slags kommunikativ "flagga" att talaren behöver hjälp av interlokutören/na. En mildare form av denna hypotes är att FP helt enkelt signalerar att talaren för tillfället har problem med sin talproduktion. Så, när en talare söker för ett ord – eller helt enkelt för ett sätt att fortsätta prata – så är "eh" ett tecken på att "hjälp önskas".

Self-monitoring/error detection hypothesis

Levelt (1989) föreslog att FP är ett tecken på intern feldetektion, en tråd som följdes upp av Christenfeld och Creager (1996) som var av meningen att allting som resulterade i problematisk talproduktion kunde resultera i emitterade FP. Ett exempel är "choking under pressure" Baumeister (1984).

Many-options hypothesis

Lounsbury (1954:99) föreslog att FP "correspond to the points of highest statistical uncertainty in the sequencing of units in any given order", dvs att man vid början av yttrande – där entropin är som högst – innan talaren ännu har "committat" sig till något, och där talproduktionen således fortfarande är som mest öppen och därmed komplicerad uppvisar en större sannolikhet för att producera ett "eh".

Men den kanske mest iögonfallande – och min favorit – bekräftelsen av denna hypotes är den studie som Schachter et al (1991) utförde, vilka insåg att olika discipliner uppvisar olika grad av inherent entropi (oordning). De beslöt sig således för att studera föredrag

inom tre olika discipliner, men olika grad av "oordning":

(1) naturvetenskap, där det är lättare att förutsäga vad "nästa ord" kommer att vara, eftersom det finns en begränsad mängd sätt att avsluta t ex en påbörjad mening om en planetbanas karakteristik, eller utkomsten av en kemisk reaktion;

(2) samhällsvetenskap, med en "mellanivå" av förutsägbarhet vad rör en diskussion kring tänkbara svar på hypoteser;

(3) humaniora, med en mycket hög grad av oförutsägbarhet vad gäller hur en föredragshållare tänker avsluta en påbörjad mening. Som ett exempel ber författarna fundera på hur följande mening avslutas:

"The reason Jackson Pollock put the patch of red in that corner of the canvas was..." Mycket riktigt producerade föredragshållare inom naturvetenskapliga discipliner först FP, samhällsvetenskapliga områden lite fler, men föredragshållare inom humaniora flest FP.

Attention-getting signal

Lalljee & Cook (1974) skapade att antal experiment för att test floor-holding-hypotesen, men fann inget stöd för denna hypotes, vilket förelidde författarna till att förstå att FP i stället, och helt enkelt, utgör ett sätt att få uppmärksamhet. Detta skulle även förklara den höga incidensen av FP i yttrande-initial position ("här är jag, nu är det min tur att prata"). Emellertid påpekar Lalljee och Cook även att FP mycket väl kan fylla flera olika funktioner, och att ett experiment som är designat för att undersöka en (datorlingvistik- och logikjargon för "en och endast en) funktion mycket väl kan missa andra funktioner.

För att summera denna – ofullständiga och förenklande – listning av olika föreslagna funktioner som FP kan tänkas ha är det viktigt att tänka på att FP mycket väl kan ha flera funktioner i talspråk – vilket Lalljee & Cook påpekar – samt att dessa mycket väl kan vara "sanna utan att lyckas".

Men som slutord kan man inte komma ifrån att many-options-hypotesen har rönt en oerhörd stor mängd stöd i de studier som utförts. Det verkar helt enkelt vara som Christenfeld (1994:192) konstaterar: more options did produce more filled pauses".

Fyllda pauser: bra eller dåliga?

Vi har ovan sett att FP har studerats i detalj och över decennier utifrån deras funktion ut ett talproduktionsperspektiv. Varför produceras FP? Den uppenbara och följande frågan är vilken effekt FP har på lyssnaren, dvs vad har studier av perception av FP uppvisat? Jag ska i det följande kortfattat redogöra för ett antal studier som har rapporterat resultat från perceptionsstudier av disfluent tal (snarare än FP specifikt).

Studier som har visat att disfluenser kan ha negativa effekter på lyssnare inkluderar: McCroskey och Mehrley (1969), vilka fann att tal som innehöll disfluenser resulterade i "mindre övertygande tal". Duffy, Hunt Jr. och Giolas (1975) fann att disfluenser negativt påverkade lyssnarens uppfattning av talarens kompetens. Fox Tree (1995) fann att "false

starts" (talaren avbryter sig och får börja om från början) försvårade talförståelse, men att repetitioner (upprepade ord) inte hade denna effekt. Lickley och Bard (1996) fann att specifika ord kan vara svårare att förstå i allmänt disfluenta yttrande än i flytande yttranden. Christenfeld (1995) rapporterar att även om flytande tal ger bäst intryck så upplevdes det mycket bättre om talaren sade "eh" än om talaren blev tyst (ofyllda pauser). Exakt samma slutsats drogs av Bortfeld et al. (1999).

Ergo: vid tvekan, tystna inte utan säg "eh"!

Studier som har visat på positiva effekter inkluderar: Fox Tree (2001) fann att "uh" fick lyssnare att förstå kommande ord snabbare än i flytande tal – medan den amerikanska varianten "um" inte hade denna. Slutligen visade Arnold et al. (2003) att FP fick lyssnare mer predisponerade mot perception av inte tidigare nämnda objekt, medan flytande tal fick lyssnare mer predisponerade mot tidigare nämnda objekt. Detta bekräftades i en unik EEG-studie av Corley, MacGregor och Donaldson (2007), som visade att den EEG-komponent som kallas N400 (EEG uppvisar en negativ "spik" efter ca 400 ms), och som är välkänd vid lyssning på tal som innehåller semantiska anomalier, i hög grad undertrycktes om den semantiska anomalin följde en FP.

Fyllda pauser: en sammanfattning

Vi har ovan sett att FP är vanliga; att de skiljer sig från alla övriga slags disfluenser i en mängd dimensioner (som funktion och effekt); att de kan tjäna flera syften (kanske simultant); att de kan ha såväl "skadlig" som hjälpsam effekt på lyssnaren, med mera.

Fyllda pauser: vår frågeställning

En sak som nästan samtliga av ovan nämnda studier, speciellt de kognitivt inriktade, uppvisar följande drag:

- De har inte studerat FP per se, utan dessas effekt på följande, och "riktiga" ord i tal (eller liknande).
- De har i hög grad utförts på scriptat, förberett eller på annat sätt "skapat" tal (med ett notabelt undantag i Eklund (2010)).
- De "kognitiva" studier som utförts har antingen varit beteendestudier eller utförts elektrofysiologiskt/med EEG.

Vad vi (jag och Martin Ingvar) ville studera var i stället:

1. Studera den effekt FP per se har på hjärnan, alltså inte den effekt som FP kan ha på följande ord, eller liknande.
2. Vi använde fMRI (se nästa stycke) i stället för beteendestudier eller EEG.
3. I stället för att använda scriptat eller på annat sätt artificiellt producerat tal för våra stimuli använde vi data med högsta ekologiska validitet, som beskrivet i Eklund (2010).

Notera att dessa tre punkter gör vår studie unik – vilket medför såväl för- som nackdelar, som vi ska se i det följande.

Neurokognition

Jag skall inte redogöra för neurokognitionens historia men redogör för några av de första milstolparna inom detta område.

Hjärnan har inte alltid associerats med kognitiva processer, och ett belysande exempel är att man i antika Egypten vid mumifiering var noga med att bevara alla organ utom hjärnan, som snabbt avlägsnades och slängdes bort; detta är ett välkänt faktum. Att hjärnan hade med såväl kognition som personlighet blev dock alltmer uppenbart, och ett exempel som förekommer flitigt i litteraturen behandlar järnvägsarbetaren Phineas Gage (1823–60) som vid en olycka 1848 fick ett järmspett genom hjärnan, och som mirakulöst överlevde, men med en totalt förändrad personlighet.

Hjärnan är ett hungrigt organ, och trots att den i snitt enbart utgör ca 2 % av kroppsvikten förbrukar den 20 % av den totala mängden syre i kroppen samt 25 % av kroppens glukos, samt svarar som mottagare av 15 % av hjärtats "output" (Jain, Langham & Wehrli, 2010). Så den uppenbara frågan är huruvida man kan använda denna "aptit" för att objektivt mäta kognitiva processer?

En pionjär inom området var Angelo Mosso (1846–1910), som mätte hjärnaktivitet genom att lägga försökspersoner på en perfekt balanserad säng-liktande våg. När hjärnan började processa information strömmade mer blod till hjärnan (för att förse den med syre), vilket gjorde den tyngre, vilket fick sängen (vågen) att tippa över till den sida där huvudet befann sig. Detta finns beskrivet av William James (1890:98):

The subject to be observed lay on a delicately balanced table which could tip downward either at the head or at the foot if the weight of either end were increased. The moment emotional or intellectual activity began in the subject, down went the balance at the head-end, in consequence of the redistribution of blood in his system.

Det är denna ökade blodtillströmning/syre-förbrukning som utgör basen för fMRI-studier.

Hjärnabildning ("neuroimaging")

Det finns nuförtiden en stor mängd olika sätt att mäta hjärnans aktivitet, och man kan t ex använda hjärnans elektriska aktivitet (EEG), dess magnetiska aktivitet (MEG, TMS) eller, som i den här studien fMRI (functional Magnetic Resonance Imaging), där man använder sig magnetfält och radiovågor för att mäta lokalt ökad aktivitet i olika delar av hjärnan.

Observera att man inte mäter områden där "hjärnan är aktiv" – något som antagligen till viss del ligger grund för den utbredda myten om att "man använder bara 10 % av hjärnan" (se t ex Beyerstein, 1999). Eller som Dünder och Günduz (2016) sammanfattar det:

"The idea that we use only 10% of our brain's capacity is the most common neuromyth /.../ However, science has failed to confirm any such unused region". Det man mäter med fMRI är (marginellt) ökad aktivering i vissa, specifika, områden i hjärnan.

Den aktuella studien

Den studie som beskrivs i denna artikel utfördes som ett postdocarbete på Karolinska Institutet, avdelningen för klinisk neurovetenskap, med Martin Ingvar som huvudhandledare. En stor mängd andra personer var till enorm hjälp och dessa tackas i slutet av denna artikel.

Experimentbeskrivning

Eftersom en detaljerad experimentbeskrivning egentligen kräver en tämligen god kunskap om hur fMRI fungerar hänvisar jag till (t ex) fmri4newbies (länk nedan) och undviker att här redogöra för mer tekniska detaljer av vårt experiment.

För en detaljerad sådan, teknisk, beskrivning hänvisar jag till Eklund och Ingvar (2016), samt ger i slutet av denna artikel en länk för direkt nedladdning av denna artikel (samt den poster som presenterades på Interspeech i San Francisco år 2016).

Nedan följer dock en grundbeskrivning av experimentet (utan tekniska detaljer):

Stimulusdata var ena sidan av HH-dialoger, insamlade i en WOZ-datainsamling. Det rörde sig om ena sidan ("kunden") av affärsresbeställningar, vilka producerade talet baserat på ordfria instruktionsblad (för att undvika direktkopiering från instruktionsbladet). Fyra talare användes (2 män; 2 kvinnor) och taldata valdes så att alla yttranden var helt flytande med undantag av FP och ofyllda pauser. För en detaljerad beskrivning, se Eklund (2004:187–189).

Försökspersonerna var 16 friska vuxna med åldersspannet 22–54 (medelålder 40, 3 med en standardavvikelse på 9,5). Alla försökspersoner var högerhänta enligt Edinburgh Handedness Inventory (Oldfield, 1971).

Instruktionerna som gavs till försökspersonerna var att de skulle spela rollen av resebyråagenten och noga lyssna på "kunden" sade ("jag vill boka en bil den 3 maj", etc), men att de inte förväntades svara utan enbart tyst skulle lägga uppmärksamhet på vad som sades.

För alla övriga tekniska aspekter (beskrivning av scanner och annan utrustning, experimentdesign, MRI-scanningsmetodik, data-postprocessning, etc, hänvisas till Eklund och Ingvar, 2016; länk i slutet av denna artikel).

För en grafisk beskrivning av experimentdesignen, se Bild 1 - se Röstlägets baksida.

Analys och resultat

För analys av data skapades tre kontraster, där flytande tal (FS) utgjorde baslinjen (se Bild 1):

1. FP ökad aktivitet jämfört FS
2. UP ökad aktivitet jämfört FS
3. FP ökad aktivitet jämfört UP

Givet att denna studie är den första i sitt slag tittade vi på hela. Resultaten beräknades med en False Discovery Rate med $p < 0,05$ (se Genovese, Lazar & Nichols, 2002). Tröskeln för klusternivån sattes till tio sammanhängande voxlar.

Vi fann ingen aktivitet i Brodmann Area 22 (BA22), dvs Wernickes Area, vilket förknippas med processning av semantisk information (betydelse).

För kontrasten FP > FS fann vi ökad aktivitet i Primära Hörselkortex (PAC), binauralt, i subkortikala områden (cerebellum, putamen); se Röstlägets baksida - Bild 2.

Intressantare fann vi modulering av Supplementära Motorkortex (SMA), BA6. Typisk aktivering visas i Bild 3 och Bild 4 - se Röstlägets baksida.

För kontrasten UP > FS fann vi ökad aktivering PAC, samt även (bland annat) i Heschl's Gyrus, Rolandiska Operculum. Ingen aktivering i motorkortex observerades. Typisk aktivering visas i Bild 5 - se Röstlägets baksida.

För kontrasten FP > UP fanns vi att aktiveringsmönstret till stor liknande FP > FS. Typisk aktivering (endast ett exempel) visas i Bild 6 - se Röstlägets baksida.

Resultaten visar således att såväl FP som UP påverkar vår auditiva uppmärksamhet (båda aktiverar hörselkortex), men medan UP modulerar områden för syntaxprocessning - se Eklund och Ingvar (2016) en tabulering av den fullständiga moduleringen av olika områden - så har FP inte denna effekt. FP modulerar i stället motorkortex. Jämfört med FP så verkar FS och UP ha ungefär samma effekt i det att ingen observerad skillnad föreligger i det två kontrasterna FP > FS och FP > UP.

Diskussion

Våra starkaste resultat var aktiveringen av PAC, som i hög grad moduleras av såväl UP som FP, dvs lyssnarna verkar "spetsa öronen" både när de hör UP och när de hör FP. Att ökad uppmärksamhet påverkar perception har tidigare visats (Petkov et al., 2004).

Den observerade ökade aktiviteten i PAC kan förklara (åtminstone delvis) de kortare reaktionstider till lingvistiska stimuli som följer FP som rapporterats av Fox Tree (1995, 2001).

Hypotesen som följer av våra resultat skulle således vara att även UP skulle leda till kortare reaktionstider i experiment av den typ som Fox Tree utförde.

Vad rör vår den ökade aktiviteten i motorkortex (SMA/BA6) så finns det flera tänkbara förklaringar. Den mest uppenbara av dessa är att när en lyssnare hör en talare säga "eh" så förbereder sig lyssnaren för att själv tala.

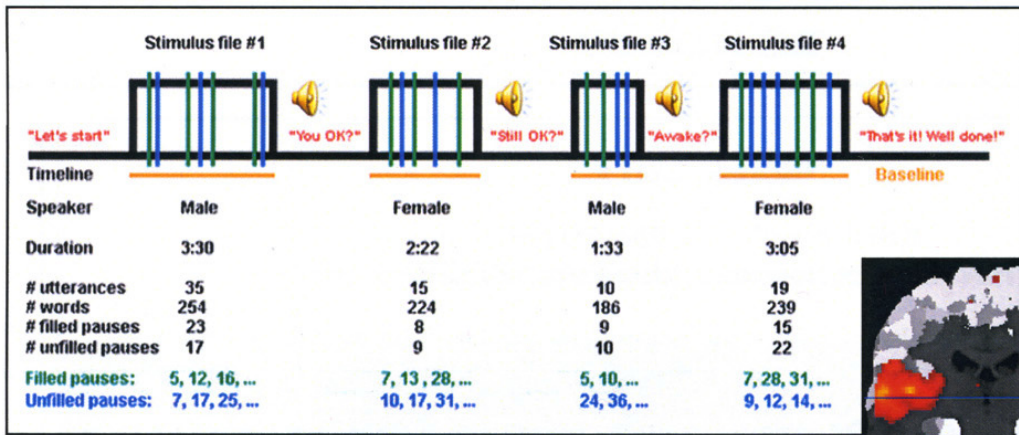
Våra resultat har direkta implikationer för två populära FP-hypoteser. Vad rör den redan 1959 föreslagna "floor-holding"-hypotesen (Maclay & Osgood, 1959), som går ut på att talare säger "eh" för att förhindra lyssnaren att ta över konversation. Detta kan mycket väl stämma, men våra resultat tyder på att produktion av FP i så fall verkar kontraproduktivt: en lyssnare som hör ett "eh" uppvisar i stället ett beteende, på neural nivå, som tyder på att talproduktionen snarare aktiveras än förhindras.

Vad rör den alternativa "help-me-out"-hypotesen (Clark & Wilkes-Gibbs, 1986) så är nyheterna bättre. Om FP aktiverar SMA hos lyssnaren så kommer en interlokutör som hör ett "eh" antagligen att reagera snabbare på den eftersökta hjälpen.

Sammanfattning

Den aktuella studien är intressant av tre huvudsakliga skäl:

1. Vi använder fMRI för att studera disfluensperception; tidigare studier har huvudsakligen använt sig av EEG med därmed associerat fokus på temporala aspekter av talperception.



Vad denna vackra palett av former, färger och tabeller innebär är inte lätt att veta, innan man tagit del av Robert Eklunds innehållsrika artikel i detta nummer av Röstläget.

Eller...äh...vad säger ni som hörde hans föreläsning på...hmm...Röstfrämjandets årsmöte i Linköping för några år sedan?

Slå upp sidan 13 och sätt igång. Utan fyllda pauser.

