

LEMMATISERING AF ISLANDSK:

Sidste sommer ville man forsøge at experimentere med automatisk lemmatisering af islandsk. Til rådighed havde man 1-2 måneders tid (juli/august), en PDP 11/60 (Fortran og Basic) og tre medarbejdere Baldur Jónsson, Höskuldur Þráinsson og Björn Ellertsson. Vi diskuterede grundigt de forskellige muligheder vi havde for at kunne komme frem til nogen resultater og besluttede til sidst at bruge et system med ordbogskonsultation. For også at have en mulighed for at få et system implementeret, blev det Basic-programm, som Martin Kay fremviste i Bergen 1973, benyttet. Det var i forvejen klart at det ikke kan analysere sammensatte ord (medmindre disse findes i ordbogen), men vi håbede på, at med mindre ændringer og tilføjelser skulle det kunne klare størstedelen af den islandske morfologi. Oprindeligt kunne det klare genkendelse og ændringer af præfikser og suffikser, udføre ordbogssøgning, addere oplysninger til en identifikationsstræng om analyseordet, benytte subrutiner og udføre analysen på en ikke-deterministisk måde. Det var altså klart i forvejen, at uændret kunne programmet ikke klare den overordentlig hyppige u-omlyd (a til ö, barn - börn), og ikke uden besvær synkope (bortfald af a, i, u i endelser), f.eks. akur - ökrum, samt andre fenomener af lignende art (fár - færri; assimilation af stam og endelse (muligvis med stamforandring) osv.) Med tilføjelse af instruktioner for at genkende og ændre infikser (indlysende komplettering til præfikser og suffikser), en instruktion, som med omvendte transformationer gør u-omlyd tilbagegængelig, samt mulighed for ordbogsopsøgning af ord med synkope er systemet blevet temmelig komplet. Det er blevet afprøvet på størstedelen af substantivernes deklination og adjektivernes deklination og komparation. Også verbernes konjugation synes det at kunne klare (ganske vist med op til 6 stammer for stærke verber og op til 2 stammer for svage verber). Vi syntes, at resultatet var tilfredsstillende, systemet er enkelt at benytte og instruktionerne lette at indlære sig. På den negative side står, at programmet selv interpreterer hver enkelt grammatik-instruktion hver gang den skal udføres (uden f.eks. at indsætte run-time-adresser), hvad der kan accepteres (med fordel?) på et experimentelt system, men ellers er uacceptabelt. Det samme gælder iøvrigt for benyttelse af Basic, medmindre man ikke overfører systemet til en egen home-computer. Sammenfattende kan man altså sige, at det drejer sig om et beskedent system, som også kan klare kompliceret flexion, dens benyttelse er enkel og nem at tilegne sig (ideel for undervisning), det kan let udvides, men til produktion skulle det omskrives til et kompileret sprog (Fortran eller lign.) og koden skulle også optimeres, hvis det ikke drejer sig om grammatikken som stadig bliver ændret.

Reykjavík, den 17. sept. 1979

For gruppen

Björn Ellertsson

Björn Ellertsson.