Eiríkur Rögnvaldsson

Mál og tölvur

1. Tölvur og texti

1.1 Máltölvun

Á undanförnum árum hefur færst mjög í vöxt að nota tölvur við vinnu að ýmsum málfræðilegum verkefnum, s.s. gerð ýmiss konar tíðniskráa og orðstöðulykla, orðabókagerð o.s.frv. Þetta eru verkefni sem í sjálfu sér væri hugsanlegt að vinna án aðstoðar tölvunnar, en yrðu flest hver aldrei unnin vegna þess hversu umfangsmikil þau eru. Þetta er nefnt máltölvun. Oft er gerður munur á máltölvun annars vegar og tölvufræðilegum málvísindum hins vegar. Með hinu síðarnefnda er átt við samningu mállýsingar sem tölvur geta nýtt t.d. í málgreiningu, sem er aftur undirstaða þess að hægt sé að nota tölvur við vélrænar þýðingar, lemmun, talgreiningu o.fl.

1.2 Gagnamálfræði

Gagnamálfræði er mállýsing sem byggist á athugun gagnasafna, textaheilda. Hún er því ekki ein af undirgreinum málfræðinnar, heldur aðferð til að rannsaka tungumál. Þar er t.d. skoðuð tíðni einstakra orða og orðasambanda, hlutfall milli orðflokka o.fl. Gagnamálfræði er oft stillt upp sem andstæðu fræðilegra málvísinda, þar sem megináhersla er lögð á að setja fram kenningar og prófa þær síðan á tungumálinu sjálfu. Gagnamálfræðin byrjar hins vegar á því að skoða textana vandlega og setja fram lýsingu að þeirri skoðun lokinni.

Gagnamálfræði er í sjálfu sér hægt að stunda án nokkurra hjálpartækja. Á seinni árum hefur hún þó tengst máltölvun mjög sterkum böndum. Niðurstöður gagnamálfræðinnar verða því traustari og gagnlegri sem textaheildin sem unnið er úr er stærri. Úrvinnsla úr stórum textaheildum er hins vegar nánast óframkvæmanleg nema í tölvu. Tölvutækum textum fjölgar ört og tölvur verða stöðugt öflugri, þannig að umhverfi gagnamálfræðinnar hefur gerbreyst á fáum árum.

Þær niðurstöður sem aflað er með gagnamálfræði koma síðan að gagni á ýmsum sviðum, s.s. í málgreiningu, vélrænum þýðingum, orðabókagerð o.s.frv.

1.3 Textaheild

Með textaheild er átt við mengi texta af ýmsu tagi sem sett er saman eftir ákveðnum reglum og í ákveðnu augnamiði. Reglurnar geta t.d. varðað lengd textanna, tegundir

þeirra (bókmenntatextar, fræðitextar o.s.frv.), höfunda (aldur, kyn, uppruni) o.fl. Þessar reglur eru settar til að hægt sé að halda því fram að textaheildin gefi raunsanna mynd af því sem hún á að bera vitni um, en sé ekki tilviljanakennt samsafn texta sem engar öruggar ályktanir verði dregnar af.

Nú á dögum eru textaheildir nær undantekningarlaust settar saman úr tölvutækum textum, og á seinustu árum hafa verið byggðar upp viðamiklar textaheildir á ýmsum tungumálum. Þessar textaheildir hafa verið hagnýttar á ýmsum sviðum, ekki síst í orðabókagerð. Textaheildir koma einnig að gagni við gerð málgreiniforrita, þar sem þær auðvelda mönnum að sjá hvað er algengt og hvað sjaldgæft í málinu. Slíkar upplýsingar er t.d. hægt að nýta til að auka möguleikana á réttri greiningu tvíræðra orða eða orðasambanda.

Íslensk orðtíðnibók sem Orðabók Háskólans gaf út 1991 er byggð á textaheild sem sett var saman í þessum tilgangi. Sú textaheild er þó mjög lítil (500 þúsund orð) og í hana vantar ýmsar tegundir texta; t.d. er þar ekkert talmál.

1.4 Orðstöðulyklar

Orðstöðulykill tiltekins texta er skrá þar sem öll dæmi um sérhverja orðmynd textans eru sýnd í samhengi, þ.e. með nokkrum næstu orðum á undan og eftir, þannig að hvert dæmi fær sérstaka línu. Yfirleitt fylgja hverri línu upplýsingar um það hvar í textanum dæmið er að finna. Samhengið sem sýnt er getur verið mislangt, en er oft u.þ.b. 40-50 bókstafir í hvora átt. Orðstöðulyklar eru ýmist lemmaðir eða ekki.

Flettiorðum orðstöðulykils er raðað í stafrófsröð, en tveir möguleikar eru á innbyrðis röðun dæma um hvert orð. Annar er sá að raða dæmunum eftir stöðu þeirra í textanum, þannig í fyrstu línunni komi fyrsta dæmið um orðið í textanum, og svo koll af kolli. Þessi röðun er yfirleitt viðhöfð í orðstöðulyklum að Biblíunni. Hinn möguleikinn er sá að raða dæmunum eftir umhverfi þeirra, án tillits til stöðu í textanum. Þá er yfirleitt miðað við stafrófsröð þess sem kemur á eftir.

Báðar aðferðirnar hafa sína kosti og galla. Með því að raða dæmum eftir stöðu í texta má t.d. sjá hvort tíðni eða notkun orðs er breytileg milli einstakra hluta textans, en erfiðara er að greina mynstur í textanum. Með því að raða dæmum eftir umhverfi er aftur á móti auðveldara að koma auga á ýmis endurtekin mynstur, s.s. orðastæður en erfiðara er að átta sig á mun milli einstakra hluta textans. Í tölvutækum orðstöðulyklum er oft auðvelt að kalla dæmin fram í hvorri röðinni sem er.

Gagnsemi orðstöðulykla er margvísleg í ýmsum fræðigreinum, en málfræðingum nýtast þeir helst til að átta sig betur á notkun og merkingu einstakra orða, orðasamböndum, orðastæðum og slíku. Þegar öll dæmin um tiltekið orð eru skoðuð saman koma oft í ljós ýmis atriði sem annars væru hulin. Orðstöðulyklar eru því m.a. ómetanleg hjálpartæki við orðabókagerð.

1.5 Lemmun

Lemmun texta felst í því að flokka saman þær myndir sem tilheyra sama flettiorði, og greina sundur samhljóma orð sem tilheyra mismunandi flettiorðum. Þannig þarf t.d. að færa saman myndir eins og ‘á’, ‘eiga’, ‘ætti’ o.fl. af so. ‘eiga’, en greina sundur dæmi um ‘á’ eftir því hvaða flettiorði þau tilheyra[1].

Nú á dögum fer lemmun oftast fram í tölvum, enda er mjög mikið verk að lemma texta “í höndunum”. Til að vélræn lemmun skili góðum árangri verður hún að byggjast á málgreiningu. Vélræn lemmun verður þó sjaldnast fullkomlega rétt, og því er oft farið yfir hana eftir á og hún leiðrétt eftir þörfum.

1.6 Töggun

Með töggun er átt við það að setja ákveðin tákn inn í tölvutækan texta. Þessi tákn geta verið af ýmsu tagi og haft margvíslegt gildi, en hlutverk þeirra er yfirleitt að merkja tiltekin atriði eða einingar í textanum sem ýmis forrit geta síðan túlkað á mismunandi hátt. Sem dæmi má nefna merkingu orðflokka, setningarliða, leturgerða o.m.fl.

Töggun farið fram ýmist á handvirkan eða vélrænan hátt, en oft er vélræn töggun leiðrétt handvirkt eftir á. Við málgreiningu fer yfirleitt fram vélræn töggun af einhverju tagi; greiningarforritið les textann, túlkar hann, og merkir einingar hans s.s. orð, setningarliði og setningar á ákveðinn hátt. Önnur forrit geta síðan nýtt sér þessa merkingu. Það er t.d. algengt að fyrst sé sett inn orðflokkstöggun, þ.e. orðflokkur allra orða merktur, en síðan sé textinn settur í setningagreiniforrit sem greini hann í setningar með hjálp orðflokkagreiningarinnar.

Ýmis töggunarkerfi eru til, en það þekktasta og útbreiddasta er SGML (Standard Generalized Markup Language). Það er alþjóðlegur staðall sem tekur til mjög fjölbreytilegra merkinga í texta, bæði hvað varðar innihald hans og útlit. HTML-málið (Hypertext Markup Language), sem stjórnar útliti vefsíðna, er upprunalega einfölduð gerð af SGML. Undanfarið hefur önnur “mállýska” af SGML, svokallað XML (Extensible Markup Language) verið í örri þróun og virðist eiga mikla framtíð fyrir sér á mörgum sviðum.

1.7 Málgreining

---

[1] ‘á’ getur verið forsetning, eins og í ‘Ég bý á Íslandi’; nútíð so. ‘eiga’, eins og í ‘Stelpan á þessa bók’; nefnifall, þolfall eða þágufall kvenkynsorðsins ‘á’, eins og í ‘Þessi á er straumhörð’; og þolfall eða þágufall kvenkynsorðsins ‘ær’, eins og í ‘Hún gaf mér flekkótta á’. Auk þess getur verið um að ræða heiti bókstafsins ‘á’, upphrópunina ‘á’! sem táknar sársauka og e.t.v. fleira.

Málgreining fer yfirleitt fram í tölvu, og felst í málfræðilegri greiningu texta. Slík greining er undirstaða ýmiss konar tungutækni, s.s. vélrænna þýðinga, málfræðiforrita o.fl.

Greiningin getur verið misjafnlega nákvæm. Stundum er textinn aðeins greindur í orðflokka, en einnig getur verið um að ræða greiningu í setningarliði og greiningu á venslum þeirra, þ.e. formgerð setninga. Slík greining verður sjaldan fullkomlega rétt, en þó er hægt að ná mjög góðum árangri ef greiningin byggist á vönduðu gagnasafni. Þar er annars vegar um að ræða orðmyndaskrá sem forritið leitar í til að finna upplýsingar um einstakar orðmyndir, og hins vegar reglusafn sem hefur að geyma upplýsingar um leyfilega gerð setninga og setningarliða í málinu.

Einn meginvandinn í málgreiningu felst í greiningu tvíræðra orðmynda. Orðmyndin ‘á’ í íslensku getur t.d. táknað býsna margt, og til að greina hana rétt verður að fara eftir samhengi. Ef ‘á’ stendur næst á eftir sögn en á undan nafnorði, eins og í ‘Ég bý á Íslandi’ eru allar líkur á að um forsetningu sé að ræða, því að sagnir og nafnorð standa varla í þeirri stöðu. Standi ‘á’ hins vegar næst á eftir nafnorði en á undan ábendingarfornafni, eins og í ‘Stelpan á þessa bók’, er líklegast að um sé að ræða sögn, því að sjaldgæft er að finna forsetningar í þessari stöðu, hvað þá nafnorð. Í þessum dæmum dugir setningafræðileg greining yfirleitt, vegna þess að um er að ræða mismunandi orðflokka. Til að greina milli orðmynda af sama orðflokki getur þurft að grípa til merkingarlegrar greiningar[2].

Sáralitlar tilraunir hafa verið gerðar með málgreiningu íslenskra texta.

2. Hagnýting tölvutækni í tengslum við tungumál

2.1 Tungutækni

Með tungutækni, sem einnig hefur verið nefnd tungumálaverkfræði, er átt við hvers kyns hagnýtingu tölvutækninnar í tengslum við mannlegt mál – og öfugt. Undir þetta fellur t.d. smíði talgervla sem líkja eftir mannsrödd og lesa upp ritaðan texta, gerð ýmiss konar leiðréttingarforrita sem lagfæra stafsetningu, beygingar, orðanotkun og stíl, þýðingarforrit, forrit til talgreiningar og fjöldamargt annað.

Hagnýting tungutækninnar byggist á viðamiklum málrannsóknum af ýmsu tagi. Þær rannsóknir flokkast einkum undir tölvufræðileg málvísindi eða máltölvun og textamálfræði eða gagnamálfræði.

---

[2] Í dæmunum ‘Ég sá straumharða á’ og ‘Ég sá flekkótta á’ stendur ‘á’ í sams konar setningafræðilegu umhverfi, en lýsingarorðin sýna að í fyrra dæminu er átt við no. ‘á’ en í því seinna no. ‘ær’. Í setningunni ‘Ég sá þessa á í gær’ er hins vegar útilokað að greina um hvort orðið er að ræða. Að vísu er hugsanlegt að næstu setningar á undan eða eftir leysi málið (t.d. ef næst kæmi ‘Hún var kolmórauð eftir leysingarnar’ eða ‘Hún hafði týnt öðru lambinu’); en óvíst er hvort málgreiniforritið gæti nýtt sér slíkar upplýsingar.

2.2 Leiðréttingarforrit

Til eru ýmis forrit sem lesa tölvutæka texta og benda á villur eða hugsanlegar villur í þeim. Einföldustu forritin af því tagi leita að stafsetningarvillum. Slík forrit hafa þá innbyggt safn rétt ritaðra orðmynda. Þau lesa textann, orð fyrir orð, og bera saman við orðasafnið. Ef þau finna orðmynd sem ekki er í orðasafninu stansa þau og vekja athygli notandans á þessari orðmynd. Stundum er um að ræða rétt ritað orð sem ekki er í safninu, og þá er gefinn kostur á að bæta því í safnið; en sé um rangt ritað orð að ræða er hægt að leiðrétta það. Sum slík forrit, t.d. Púki[3], búa líka yfir upplýsingum um hvernig orð geti verið til í málinu. Leiðréttingarforrit af þessu tagi byggjast yfirleitt ekki á málgreiningu og finna því ekki villur þar sem leyfileg orðmynd er notuð á röngum stað.[4]

Til eru leiðréttingarforrit fyrir íslenska stafsetningu, t.d. Púki. Erlendis hafa einnig verið skrifuð ýmis forrit sem skoða málfar og stíl. Þau geta t.d. gert athugasemdir við orðaröð, orðanotkun o.fl., allt eftir því hversu fullkomin þau eru. Málgreining er hins vegar forsenda forrita af þessu tagi, og engin slík eru til fyrir íslensku.

Af svipuðum toga eru orðskiptiforrit sem skipta orðum milli lína í samræmi við reglur viðkomandi tungumáls.[5] Orðskiptiforrit geta þó ekki unnið eingöngu eftir almennum reglum, heldur verða líka að byggjast á orðasafni þar sem talin eru upp helstu orð sem víkja frá reglunum.

2.3 Vélrænar þýðingar

Sú hugmynd að nota tölvur til að þýða texta af einu tungumáli á annað er næstum því jafngömul og tölvutæknin sjálf. Gífurleg vinna og fjármagn hefur verið lagt í tilraunir

---

[3] **Púki** er stafsetningarleiðréttingarforrit sem Friðrik Skúlason er höfundur að. Forritið þekkir ekki aðeins leyfileg íslensk orð, heldur kann það líka reglur um íslenska orðmyndun. Þannig samþykkir það orð sem það hefur aldrei "séð" áður, ef þau samræmast íslenskum orðmyndunarreglum.

[4] Í setningunni 'Ég hitti Þórarinn' er 'Þórarinn' í þolfalli og á aðeins að hafa eitt 'n'; en vegna þess að 'Þórarinn' með tveimur 'n'-um er leyfileg mynd (rétt nefnifallsmynd) gerir forritið ekki athugasemd. Í setningunni 'Vatnið síður' er seinna orðið rangt ritað; þar á að vera 'sýður' (af so. 'sjóða'). Vegna þess að 'síður' með 'í' er leyfilegt orð (atviksorð) gerir forritið heldur ekki athugasemd í þessu dæmi. Til að svo mætti vera þyrfti frekari greiningu textans.

[5] Orðskiptiforrit byggjast yfirleitt á ákveðnum reglum um það hvar í stafasamböndum megi skipta orðum. Í íslensku verður t.d. að vera a.m.k. eitt sérhljóð í hvorum hluta; 'skrjóðs' má hvorki skipta 'skrj-óðs' né 'skrjó-ðs'. Ekki má heldur flytja aðeins eitt sérhljóð milli lína; og seinni hlutinn á yfirleitt að hefjast á sérhljóði. Undantekningar frá síðastnefndu reglunni eru þó fjölmargar, einkum í samsettum orðum.

með tölvuþýðingar eða vélrænar þýðingar síðan um miðja öldina, en árangurinn hefur verið misjafn.

Þær kröfur eru venjulega gerðar til þýðingar að hún skili málfræðilega réttum og eðlilegum texta, og merking frumtextans haldi sér. Góðir þýðendur geta fullnægt þessum kröfum, en nauðsynlegt er að hafa í huga að til að gera það þurfa menn að búa yfir góðri kunnáttu í bæði málinu sem þýtt er af og málinu sem þýtt er á. Til að tölvur geti leikið þetta eftir þarf að mata þær á mjög miklum upplýsingum um bæði tungumálin. Þar er að sjálfsögðu um að ræða einstök orð og merkingu þeirra, en einnig reglur um orðasambönd, orðaröð, setningagerð o.m.fl. Forsenda þess að vélræn þýðing takist er rétt greining á frumtextanum, þar sem leyst er úr allri tvíræðni í merkingu orða og setninga. Ekki dugir að þýða orð fyrir orð, heldur þarf að laga setningagerð að reglum málsins sem þýtt er á.

Þó að vélrænar þýðingar verði sjaldnast fullkomnar geta þær oft auðveldað þýðendum starfið. Tölvan er þá látin "hráþýða" texta sem þýðandi fer síðan yfir, leiðréttir og lagfærir. Í sumum tilvikum getur slík ófullkomin þýðing verið fullnægjandi, t.d. við upplýsingaleit í gagnabönkum á erlendum málum. Þá dugir notandanum oft að fá hráa þýðingu til að átta sig á merkingu upplýsinganna, en villur í beygingum, orðaröð og slíku skipta ekki máli. Í tilteknum verkefnum geta vélrænar þýðingar líka skilað tiltölulega réttum texta.[6]

Þróun vélrænna þýðinga milli íslensku og annarra tungumála er skammt á veg komin. Þó hafa verið gerðar tilraunir til að nota esperanto sem millimál í þýðingum. Þá er ekki þýtt beint t.d. á íslensku úr ensku, heldur er enski textinn fyrst þýddur á esperanto og sá texti síðan þýddur á íslensku. Þetta kann að virðast tvíverknaður, en kosturinn við þessa aðferð er sá að nóg er að þróa eitt reglusafn fyrir hvert mál í stað þess að hafa t.d. sérstakar reglur milli íslensku og ensku, aðrar milli íslensku og frönsku o.s.frv. Esperanto er valið sem millimál vegna þess hversu skýrar reglur þess eru, enda er það tilbúið tungumál.

2.4 Talgervlar

Sameiginlegt einkenni talgervla er að þeir líkja eftir mannlegu máli – lesa upp ritaðan tölvutækan texta með rödd sem líkist mannsrödd. Gerð talgervla er að öðru leyti mjög mismunandi. Sumir þeirra byggjast eingöngu á sérstökum hugbúnaði, en aðrir nota einnig sérhæfðan vélbúnað til að gera tölvunni kleift að mynda þau málhljóð sem um er að ræða. Nú er þó yfirleitt hægt að nýta venjuleg hljóðkort til þess. Talgæðin í talgervlum eru mjög mismunandi. Erfitt eða útilokað er að ná mannsröddinni

[6] Í Kanada eru tvö opinber tungumál, enska og franska. Þar hafa tölvur um langt skeið verið látnar þýða veðurfregnir milli þessara mála. Þá er um að ræða texta með afmörkuðum orðaforða og einfaldri setningagerð, þannig að hægt er að mata þýðingarforritið fyrirfram á öllum nauðsynlegum reglum og upplýsingum. Sömu aðferð hefur víða verið beitt í þýðingu ýmissa leiðbeiningarbæklinga milli mála.

fullkomlega, en meginatriðið er vitaskuld að talgervillinn skiljist vel og áreynslulaust, og sé sæmilega áheyrilegur.

Hugbúnaðurinn byggist upp á nokkrum meginþáttum. Í fyrsta lagi þurfa þar að vera nákvæmar upplýsingar um málhljóð þess tungumáls sem búnaðurinn er gerður fyrir, svo sem formendur sérhljóða o.fl. Sé reynt að nota hugbúnað sem gerður er fyrir annað mál, t.d. ensku, til að lesa upp íslenskan texta verður útkoman framandi því að enskt a eða enskt s hljómar öðruvísi en íslenskt a og íslenskt s.

Í öðru lagi þarf hugbúnaðurinn að búa yfir upplýsingum um samband stafsetningar og framburðar. Þannig þarf hann t.d. að "vita" að stafasambandið '-ll-' er oftast borið fram [dl] í íslensku, eins og í 'pollur'; að 'a' á undan 'ng' og 'nk' er oftast borið fram [au], eins og í 'langur'; og að 'e' á undan 'gi' er borið fram [ei], eins og í 'vegi' – svo að örfá dæmi séu tekin. Einnig koma þarna til upplýsingar um mismunandi lengd hljóða, sem fer eftir hljóðfræðilegu umhverfi.

Í þriðja lagi þurfa talgervlar helst að geta lesið textann með breytilegu tónfalli. Í íslensku er eðlilegt að tónn lækki í lok segðar en hækki t.d. í spurningum. Texti sem er allur lesinn í sömu tónhæð verður óáheyrilegur og fljótlega mjög þreytandi fyrir hlustandann. Hægt er að nota greinarmerki í textanum, einkum punkta og spurningarmerki, til að stjórna tónfallinu að nokkru leyti.

Í fjórða lagi er æskilegt að hugbúnaðurinn hafi aðgang að undantekningasafni; skrá um orð sem eru borin fram öðruvísi en búast mætti við eftir almennum reglum. Þannig er t.d. orðið 'pilla' ekki borið fram með [dl] eins og almenna reglan segir, heldur með [l:], þ.e. löngu 'l'. Orðið 'tangó' er líka borið fram með [a], þótt almenna reglan segi að 'a' sé borið fram sem [au], þ.e. eins og 'á', á undan 'ng'. Í hvert skipti sem hugbúnaðurinn les nýtt orð byrjar hann á að skoða hvort það er í undantekningasafninu. Ef orðið finnst þar, les talgervillinn það með þeim framburði sem þar er gefinn upp. Sé orðið aftur á móti ekki í undantekningasafninu beitir talgervillinn almennum reglum um framburð þess.[7]

Talgervlar koma víða að notum. Þeir eru ómetanleg hjálpartæki fyrir blinda og sjónskerta, og gera þeim kleift að "lesa" blöð og bækur, auk þess sem þeir auðvelda þeim að nýta sér ýmsa þætti tölvutækninnar s.s. tölvupóst og veraldarvefinn. Talgervla má einnig nota í ýmiss konar símasvörun og annars staðar þar sem koma þarf upplýsingum á framfæri í töluðu máli á sjálfvirkan hátt.

---

[7] Stundum getur sama ritmyndin haft tvenns konar framburð eftir merkingu. Ritmyndin 'galli' er t.d. borin fram með [dl] ef merkingin er 'missmíð, ókostur' en með [l] ef merkingin er 'fatnaður'. Þetta veldur venjulegum málnotendum sjaldnast vandræðum vegna þess að samhengið sker úr um það hvor merkingin á við. Hugbúnaður talgervla getur hins vegar sjaldnast greint samhengi, heldur skoðar aðeins hvert orð fyrir sig, þannig að hann getur ekki valið rétta framburðinn af öryggi. Í slíkum tilvikum er reynt að meta hvor myndin sé algengari og talgervillinn látinn nota hana alltaf; en það leiðir óhjákvæmilega til þess að sjaldgæfara orðið verður borið fram á rangan hátt.

2.5 Talgreining

Talgreining felst í vélrænni greiningu á töluðu máli. Tölva greinir þá hljóðbylgjurnar niður í sneiðar, þ.e. einstök málhljóð, og túlkar greininguna. Talgreining byggist því m.a. á málgreiningu.

Í tali eru engin sérstök skil milli hljóða, heldur er um að ræða óslitið flæði þar sem hvert hljóð bæði dregur dám af umhverfi sínu og hefur áhrif á það. Því getur verið mjög flókið að greina þetta flæði í einstök hljóð, en það er nauðsynlegt til að málgreining geti farið fram. Framburður fólks er líka mjög mismunandi, raddhæð og raddstyrkur breytilegur o.s.frv. Talað mál er líka yfirleitt óskipulagt, í því er mikið af hiki, stami, mismælum, endurtekningum og hvers kyns atriðum sem torvelda greininguna.

Talgreining er því mjög flókið ferli og er enn á frumstigi, þótt framfarir séu örar. Góður árangur hefur náðst í talgreiningu lesins texta, þar sem mörg helstu einkenni talmáls eru ekki til staðar. Hins vegar á talgreining venjulegs talmáls langt í land. Þessi lýsing á einkum við um ensku, en engar tilraunir hafa verið gerðar með talgreiningu íslensku svo að vitað sé.

Hagnýtt gildi talgreiningar í venjulegri tölvuvinnslu er augljóst. Það væri miklu fljótlegra að “tala við” tölvuna sína og láta hana skrifa talið upp en að slá sama texta inn með lyklaborðinu. Einnig mætti gefa skipanir um t.d. leturstærð, spássíur, prentun og hvaðeina annað munnlega í stað þess að smella með músinni á tiltekna staði á skjánum.

Ýmsar spár eru einnig uppi um gildi talgreiningar á öðrum sviðum daglegs lífs. Þannig má hugsa sér að ýmis heimilistæki, bílar o.fl. verði búin tölvu með talgreini, og þeim megi þá stjórna með því að “tala við” þau. Ennþá er þetta framtíðarsýn, en gæti þó verið skemmra undan en marga grunar.