

Holsanova, J. (1996). När säger en bild mer än tusen ord? Om samspelet mellan text och bild i gränssnitt. In: Löwgren, J. (red.) *Teman i människa-datorinteraktion*. (pp. 65 - 76). Institution för Informatik, Lunds universitet.

När säger en bild mer än tusen ord? Om samspelet mellan text och bild i gränssnitt.

Jana Holsanova

Att konstruera multimediala och multimodala gränssnitt är en komplex process som kräver en analys av en rad aspekter: karakteristik av den information som ska förmedlas, en karakteristik av målen och syften med representationen, valet mellan olika medier och modaliteter och koordinationen mellan de olika medierna. Hänsyn måste även tas till användarens fysiska och kognitiva förmågor (såsom uppmärksamhet, minne) och preferenser (Maybury, 1996).

När man har bestämt vilken typ av information som ska förmedlas till adressaten, måste man som nästa steg bestämma i vilket medium det ska ske. Valet är inte enkelt. Några medier och modaliteter är mer effektiva än andra för vissa uppgifter, användare och kontexter. Kartor lämpar sig t.ex. för att förmedla spatial information, medan talat språk är bäst för att kontrollera apparater i kontexter där händer och ögon är fullt upptagna. I andra fall stödjer en kombination av olika medier en mer naturlig, effektiv eller exakt interaktion (t. ex. 'lägg det där' som blandar talat språk och gester). Liksom talat och skrivet språk kan användas för olika kommunikativa funktioner (att identifiera, referera till, förklara, ändra fokus), har också gester olika funktioner. Även blicken som traditionellt använts som ersättning för musen eller deiktiska indata, kan användas i flera syften: markera användarens uppmärksamhetsfokus, regulera turtagning, visa emotionella tillstånd, attityder mellan personer, expertisnivåer mm (Maybury, 1996).

Jag ska dock snäva in antalet medier och modaliteter och koncentrera mig på språk och bild. En designer måste välja mellan verbala och piktoriala representationer varje dag och den allmänna åsikten är att bilder är roligare och mer intressanta än texten. Men *vilka representationer uttrycker bäst den typ av information man vill förmedla? Vilka styrkor och svagheter finns hos verbala och bildliga representationer och hur fungerar kombinationer av dem i användargränssnitt?*

INFORMATIONSMAPPNING OCH MODALITETSTEORI

Bernsen (1994a,b,c, Bernsen and Lu, 1995, Lu and Bernsen 1995, Bernsen and Verjans, 1995) utgår ifrån att olika modaliteter har olika inherenta egenskaper och därmed är lämpliga för att uttrycka olika typer av information. Han har studerat informationsmappning i *externa representationer*. Man kan förmoda att ju bättre informationsmappning där finns mellan det som representeras och dess representation, desto lättare för användaren att kommunicera med systemet. Bernsen har inom sin *modalitetsteori* utvecklat en generativ heltäckande *taxonomi* av modaliteter för att fördjupa den

teoretiska förståelsen för mappningsprinciper som optimerar användbarhet i gränssnittet. Talat språk, skrivet språk och analoga statistiska grafer är olika exempel på enskilda modaliteter som även kan kombineras till multimodala representationer. Taxonomin grundas på ett litet antal *oppositionella särdrag* som hjälper att avgränsa modaliteterna från varandra: språklig/icke-språklig, analog/icke-analog, arbiträr/icke-arbiträr, statisk/dynamisk. Förutom modaliteter urskiljer han olika uttrycksmedier (grafisk, auditiv och taktil). Vi kommer att titta närmare på hur Bernsen (1994) bedömer egenskaper av naturligt språk versus egenskaper av analoga grafiska representationer.

SPRÅKLIGA REPRESENTATIONER

Både talat och skrivet språk är enligt Bernsen (1994a, 199b) icke-analogt (d.v.s. beror på en konventionell relation mellan representationen och det den representerar) och icke-arbiträr (d.v.s. är beroende av ett existerande system av betydelser). Språket har *allmän karaktär* men *saknar specificitet*. Med språket kan man *fokusera* på en aspekt av det som representeras och lämna tolkningsutrymmet öppet. Därför går det bra att uttrycka abstrakta tillstånd, relationer och koncept med hjälp av språket (t.ex. 'sanning', 'rättvisa'). Men det går t.ex. inte att grafiskt framställa abstraktionen 'färg' eftersom grafiska medel är inherent specifika. Även 'grön färg' är en abstraktion. Med analoga grafiska medel kan man bara framställa en samling av specifika instanser (färgspektrum). Med språkets hjälp kan vi däremot fokusera på den rätta abstraktionsnivå.

Språket innehåller viktiga *logiska och epistemiska operatörer* ('icke', 'eller', 'om/så'). Visserligen kan en negation även framställas med standardiserade ikoniska former (en cigarett överstruken med ett kryss) men dessa grafiska representationer får därmed samma icke-analoga och icke-arbiträra karaktär som naturligt språk. Språkliga representationer har jämfört med analoga representationer en *begränsad uttryckskraft*.

Nya betydelser i kontexten kan skapas genom att manipulera skriftspråk grafiskt (färgas, roteras, belysas, ändras i storleken) eller talspråk auditivt (ändringar i rytmen, intonationskonturen, styrkan).

Maybury (1996) refererar till några mediala preferenser kopplade till olika typer av information och rekommenderar att använda språk för att uttrycka temporal överlappning och temporal kvantitet ('alltid'), temporal förändringar ('tre dagar senare'), spatial eller temporal sekventialitet och kausalitet. Texten lämpar sig för semantiska relationer som orsak/följd, händelse/resultat, problem/lösning och villkor, samt för kvantifiering ('de flesta', 'flera', 'några').

ANALOGA GRAFISKA REPRESENTATIONER

Den viktigaste egenskapen hos analoga representationer är enligt Bernsen deras *specificitet*. Det finns många sätt att grafiskt framställa en vinkel på 60 grader, men språkligt har vi bara en möjlighet. "However abstract a piece of analogue graphics is, the meaning it expresses is always more specific than that of the corresponding linguistic expression." (Bernsen 1994a:6). En annan egenskap hos grafiska representationer är att de sluter tolkningsutrymmet och att de *inte kan fokusera*. Här medger Bernsen att mycket beror på den abstraktionsnivå analoga representationer har: Mycket abstrakta analoga representationer har väldigt lite dimensionalitet och närmar sig konventionella representationer. Av specificiteten hos de grafiska representationerna följer att de är uttömmande, direkt mätbara, erbjuder en ungefär föreställning om egenskaper och kvaliteter av det som representeras, kan användas som ersättning för direkt perceptuell erfarenhet och möjliggör en omedelbar identifikation (det är bl.a. därför polisen föredrar ett foto på rånaren hellre än verbala beskrivningar). Dessutom är en ny entitet lättare att introducera i en grafisk representation.

Grafer behandlas av Bernsen som en speciell typ av analoga representationer. Här har all bildlig likhet med verkligheten försvunnit, men dimensionaliteten är bibehållen. Grafer behöver å ena sidan få sina mappningsprinciper förklarade men representerar å andra sidan strukturerade data som har strukturella likheter med verkligheten. Grafer står alltså mellan analoga och icke-analoga representationer.

Till skillnad från den traditionella uppfattningen att ikoner utgör en singular 2D statisk grafisk representation, kan ikoner enligt Bernsen skapas i varje generisk modalitet. Jag kan föreställa mig att han menar att 'nej', 'ojoj' eller 'uhm' i talat språk ingår i dialogen med datorn och har en ikonisk funktion där. Ikoner är alltså oberoende av medium och existerar därför tvärs över alla distinktioner i hans taxonomi. De utgör enligt Bernsen inte en modalitet utan snarare en modalitetsstruktur. Andra modalitetsstrukturer är t.ex. diagram eller listor som kan bestå av ord, texter, bilder eller animationer.

SPECIFICERINGSMEKANISMER I SPRÅK, FOKUSERINGSMEKANISMER I ANALOG GRAFIK

Enligt Bernsen har både språket och analoga representationer utvecklat en rad mekanismer för att komma över sina egna nackdelar. Genom att använda korta, klara och innehållsrika instruktioner, sammanfattningar, eller upprepningar av viktiga nyckelord kan man öka språkets specificitet. Språket blir också mer specifikt om vi använder oss av indexikal referens, dvs. kopplar språket till den icke-språkliga erfarenheten, pekar och använder väl definierade beskrivningar ('pennan ligger där på bordet i vardagsrummet'). Att lägga till exakta metriska egenskaper hjälper till att reducera tolkningsutrymmet. Och slutligen tycks metaforer och analogier stärka specificiteten hos språket (beskrivningen 'han som ser ut som mullvad i ansiktet' hjälper effektivt för att plocka fram en person i skaran). Precis som språket utvecklat specificeringsmekanismer för att kompensera sina handikapp, har också analoga grafiska representationer utvecklat möjligheter till fokusering (Bernsen 1994a,b). Genom en svagare relation mellan representationen och det som representeras, t.ex. ett selektivt avlägsnande av specifika aspekter, minskar deras analogkaraktär och samtidigt höjs deras fokuseringsförmåga. Som fokuseringsmedel fungerar också konturer, färgsättning, foregrounding, zooming, förstoring, blinkning, användning av pilar samt repetition vilka betonar vissa aspekter i den grafiska framställningar.

PIKTORIALA REPRESENTATIONER

Mullet & Sano (1995) väljer i kapitel 6, Image and Representation, en *semiotisk* utgångspunkt och beskriver huvudsakligen bildliga representationer som piktogram, logogram, ikoner, index, symboler, logotyper, grafiska applikationer och diagramatiska representationer. Deras teckenkaraktär är produkt av en trefaldig relation mellan a) representation, b) det som representeras och c) dess mentala uttolkare (interpreter). Och eftersom varje person har en unik social, kulturell och erfarenhetsbakgrund, blir tolkningarna olika. Mullet & Sano illustrerar och analyserar en rad designlösningar och beskriver de vanligaste felen. De använder sig av Peirce's uppdelning i *syntax* (representationens internstruktur), *semantik* (relation mellan representationen och det som representeras) och *pragmatik* (effekter av syntaktiska och semantiska aspekter i relation till en specifik tolkare och hans/hennes psykologiska kontext), och betonar att hänsyn måste tas till alla tre nivåer för att kunna bedöma tecknets lämplighet för ett kommunikativt syfte och för att utveckla visuella representationer som fungerar effektivt för användarna. Slutligen formulerar de regler och principer för praktisk tillämpning i designprocessen.

Enligt Mullet & Sano har bildliga representationer följande egenskaper: Bilder är *dominanta*, *lätta att känna igen* och komma ihåg och har *stora uttrycksmöjligheter*. Om de piktoriala representationerna är relativt konstanta, kan man kommunicera med dem tvärs över sociala och språkliga gränser. En bra och effektiv bildlig representation har *omedelbar perceptuell verkan* och lämpar sig för situationer där identifikationen och en snabb och pålitlig reaktion är viktig: som varumärken, för trafiksäkerhetsskyltar och markering av farligt riskmaterial. Till skillnad från Bernsen, tillskriver Mullet & Sano även bilder en *allmän karaktär*. Den hänger ihop med bildens olika abstraktionsnivåer. I abstraktionsprocessen bevaras bara objektets mest karakteristiska delar. Skillnaden mellan t.ex. foton och schematiska representationer är att schematiska representationer tack vare en selektiv utelämnning av detaljer blir mer kanoniska och generiska och lämpar sig som representationer av

bredare klasser av objekt. Enligt Mullet & Sano kan alltså bildliga representationer fokusera på någon aspekt av objektet som i fig. 1 från Yellow Pages där den visuella representationen betyder 'målarverktyg', inte 'målarroll nr. 4'.



Fig. 1.

En förutsättning för en bra bildlig representation är att man måste bevara *kohesion* i uppsättningen av representationer (samma stil, orientering, liknande element). Ett annat villkor är en *effektiv karakterisering* av objektet. Det är bl.a. viktigt att välja de viktigaste egenskaperna i rätt perspektiv för att understryka det man vill förmedla (se fig. 2).



Fig. 2.

När det gäller bildernas *kommunikativa kraft*, måste man ta hänsyn till den fysiska, konceptuella och kulturella kontexten i vilken representationer blir tolkade. Istället för den amerikanska ikonen för brevlådan borde man - med tanke på de internationella mottagarna - föredra ett kuvert vilket vore mer universellt. Författarna varnar dock mot alltför många metaforer. Metaforer kräver nämligen två tolkningsnivåer: a) förhållandet mellan representation och det som representeras samt b) förhållandet mellan det som representeras och systemets funktion. Mullet & Sano rekommenderar att använda bilder för representationer av kända konkreta objekt som användare kommer i kontakt i vardagen, för representationer av inherent spatiala begrepp med kort presentationstid och dåliga villkor för visuell perception (t.ex. trafiksymboler). Maybury (1996:8) föreslår att använda grafiska medel för fysiska händelser. Bildliga representationer är däremot inte lämpliga för abstrakta begrepp, operationer och processer.

SAMPELET MELLAN SPRÅK OCH BILDER

Bernsen (1994a) anser att kombinationen av naturligt språk och analog grafik är optimal för många syften: Grafiken är specifik och språket fokuserar. Multimodala kombinationer av språkliga och analoga representationer kompletterar varandra och deras kombination kan ha större expressiv styrka. Detta har även visat sig i experiment som Velichkovsky (1995) genomfört: Visuellt och språkligt information blev mycket effektivare överfört i kombination än med antingen språk eller visuellt. Bernsen nämner tre fall hur kombinationen av båda modaliteter kan fungera. a) Språket *annoterar* analoga representationer som kartor, diagram, foton, bildserier. Grafiska medel är i detta fall primära, naturligt språk fokuserar bara en viss aspekt i ett kommunikativt syfte eller definierar tolkningen av de grafiska medlen. b) Analoga representationer *illustrerar* och specificerar språket som är central i representationen. Illustrationen är viktig för att en motsvarande mental modell skapas. c) En tredje möjlighet är att båda representationer är centrala och *disambiguerar varandra* (jämför också Preece m fl., 1994).

Mullet & Sano (1994) analyserar också vissa exempel på en kombination av verbala och piktoriale representationer. De kommer till följande slutsatser:

- När bild och text kolliderar, dominerar bilden. Den är allmänt förståelig till skillnad från språket, som därmed hamnar i underläge som i det klassiska exemplet i fig.3.



Ceci n'est pas une pipe.

Fig. 3.

- Språket uttrycker jämfört med bilder mycket effektivare *nyanser*. Detta påstående står i motsats till Bernsens påstående om språkets allmänna versus grafikens specifika karaktär. Men bedöm själv bilderna och språket i fig. 4.



Fig. 4.

De språkliga beteckningarna 'bar', 'snack', 'selfservice', 'restaurant' är omedelbart förståeliga med alla konnotationer och detaljer. Bilderna kan i gruppen vara rätt förståeliga, men i isolation kan det vara svårt att urskilja vilken som betecknar vad. Flera bilder motsvarar mer än en av de språkliga beteckningarna.

- Logogram är tecken som ska läsas som bilder fastän de består av bokstäver eller ord. De är konventionella och inte så universella som piktogram eftersom de baseras på stavning i ett visst språk. Detta är en potentiell fara som framgår av fig. 5.



Fig. 5.

- Bilden utgör kontexten, språket uttrycker det primära budskapet. Genom att ta bort detaljer associerade med det enskilda, kan en designer använda de allmänna kvaliteterna (en förenklad solfjäder) för att etablera en kontext i vilken meddelandet är inbäddat och kan tolkas (fig. 6).



Fig. 6.

- Vissa generella, konventionella representationer (t.ex. play, stop, pause, cue, review, stop, record) som tidigare kombinerats med språkliga beteckningar, används numera utan ord eftersom de har blivit så förståeliga att de kan stå för sig själva. Det tyder på en dynamisk utveckling.
- Det uppstår en tävling mellan ord och bild som verkar distraherande om avstavningen är fel som i fig. 7.



Fig. 7.

I motsats till Bernsens föreställning om att de båda modaliteterna disambiguerar varandra i kombinationen kan det uppstå en tvetydig relation mellan ord och bild. Det är inte klart om fig. 8 betyder 'Exit to Windows' eller 'Exit Windows'.



Fig. 8.

- Visuellt brus uppstår om man tvingar in verbala tecken i ett trångt bildformat. Då försvinner alla fördelar med en bildlig representation.
- Man måste avväga pragmatiken nog. Risken är stor att representationen kan missförstås av användaren om den är baserad på språkspel, fonetik eller stavning, humoristiska associationer, talspråksidiomer, slang eller terminologi som är kända endast för ett fåtal personer, eller om de förutsätter historiska och kulturella kunskaper. Fig. 9 skulle associera traditionen 'apple for the teacher' i amerikanska skolor, medan 'dragon' i fig. 10 skulle stå för funktionen 'drag - on'.



Fig. 9 och 10.

Kindborg (1991) tillämpade en kombination av textuella och visuella medel för att underlätta orientering och navigering i hypermedia. *Lexivisuell presentation* som metoden kallas för innefattar en rad principer och tekniker för en effektiv informationspresentation. Metoden används i de traditionella tryckta medierna som encyklopedier, serier, tidningar och instruktionsböcker och en av de viktigaste principerna är att fokusera på det primära budskapet medan man samtidigt ger en detaljerad information om kontexten. Detta reducerar risken att man förlorar orientationen. Man använder sig av grafiska tekniker såsom zoom-lådor, miniatyrkartor och pratbubblor. Fig. 11 visar en detaljerad information i sin kontext.

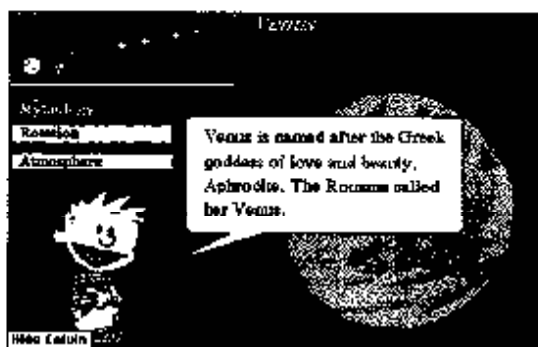


Fig. 11.

Delvis olika slutsatser när det gäller samspelet mellan ord och bild presenterade ovan reflekterar författarnas olika tillvägagångssätt. Mullet & Sano är mer praktiskt inriktade och uppmärksammar även tekniska detaljer som dyker upp i praktiken, som inte är mediuminherenta, och som alltså inte kan dyka upp i Bernsens teoretiska artiklar. Av felanalysen i Mullet & Sano följer att inte bara semantiken (som fokuseras av Bernsen), räcker till, utan att även syntaxen och pragmatiken (med t.ex. språk- och kulturspecifika konsekvenser) måste beaktas i designprocessen. Bernsens teoretiska artiklar är mer deskriptiva och tar inte så mycket hänsyn till användarens bakgrundkunskap, hans/hennes associationer, erfarenheter och interaktion med systemet. Han nämner visserligen att man inte kan undvika att ta hänsyn till *interna representationer* som framkallas hos användaren men detta påverkar inte hans taxonomi. Hela den teoretiska texten skulle kunnat bli mycket tydligare om Bernsen använde sig av principer han själv förespråkar: Användning av grafiska medel, illustrationer, konkreta exempel på modaliteter skulle ge mer specificitet och därmed förbättra förståelsen för läsaren.

KONTEXTEN

Kontexten är mycket viktig för tolkning. Vi tolkar bilden annorlunda om vi ser den isolerad eller i en serie av bilder eller om vi får veta källan eller bakgrunden. Fig. 12 hos Mullet & Sano (1995:185) har en hög abstraktionsnivå där uppmärksamheten fokuseras på några få viktiga aspekter. Vi känner likväl igen det som representeras. Men vilken funktion har den abstrakta bilden egentligen för oss? Den kan betyda både klänning, kläder, kemisk tvätt, garderob, syaffär eller damtoalett beroende på i vilken fysisk, konceptuell och kulturell kontext den befinner sig.

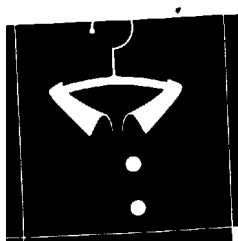


Fig. 12.

INTERAKTIONEN ANVÄNDARE-SYSTEM

Till skillnad från Bernsen (1994a:16) som påstår att "... the user is a passive recipient of information and the representer's task is one of optimising focus and specificity for given communicative purposes." anser jag med Mullet & Sano (1994) och Gärdenfors (1995:4) att tolkningen är en *aktiv* process i vilken användaren skapar tecknets betydelser genom jämförelser med sina tidigare erfarenheter och kunskaper.

Många har rätt extensivt diskuterat styrkan och svagheten hos visuella och verbala representationer och fokuserade då på frågan vilket medium är inherent överlägset. Många tror dock att svaret inte är så enkelt och att lösningen inte är *antingen* textuellt *eller* grafiskt. Den optimala representationen varierar beroende på det *kommunikativa problemet*, och då måste även en kombination av flera modaliteter prövas. Representationens lämplighet i ett visst syfte fastställs genom att den testas mot användare.

Förutom frågan om samverkan av språk och bild i gränssnitt återstår frågan om översättning mellan olika modaliteter, inte minst med hänsyn till handikappanpassning. Det finns en hypotes att samma underliggande kognitiva mekanismer i vår perception, vårt minne och vårt språk. Ur ett kognitivt perspektiv vore det alltså intressant att fråga: Vilka kopplingar finns mellan språklig och visuell information? (Jackendoff, 1989). Hur byter vi mellan olika medier? Hur översätter vi från språk till bild och vice versa? (Gärdenfors, 1995).

LITTERATUR

- Bernsen, N.O. (1994a): Why are Analogue Graphics and Natural Language both Needed in HCI? Working Papers in Cognitive Science and HCI. Centre for Cognitive Informatics. Risø National Laboratory, Roskilde University, 20s.
- Bernsen, N.O. (1994b): Foundations of multimodal representations. A taxonomy of representational modalities. *Interacting with Computers* Vol. 6 No. 4/1994, 347-71.
- Bernsen, N.O.(1994c): A revised generation of the taxonomy of output modalities. CCI Working Papers in Cognitive Science and HCI, WPCS-94-7. Centre for Cognitive Science, Roskilde University.
- Lu, S. and Bernsen, N.O. (1995): The taxonomy workbench. A multimedia database system for analysing representational modalities. In *Proceedings of the 4th International Conference on Interface to Real and Virtual Worlds*, Montpellier, June 1995.
- Bernsen, N.O. and Lu, S. (1995): A software demonstrator of modality theory. In Bastide, R. and Palanque, P. (Eds.): *Proceedings of DSV-IS'95: Second Eurographics Workshop on Design, Specification and Verification of Interactive Systems*. Springer Verlag.
- Bernsen, N.O. and Verjans, S. (1995): Designing interfaces by information mapping. A case study of the CERD design. CCI Working Papers in Cognitive Science and HCI, WPCS-95-1. Centre for Cognitive Science, Roskilde University 1995.
- Gärdenfors, P. (1995): *Human Communication: What Happens?* LUCS 43/1995. Lund University.
- Jackendoff, R. (1989): On Beyond Zebra: The Relation of linguistic and visual information, *Cognition* 26:89-114.
- Kindborg, M. (1991): Visual techniques for orientation in hypermedia structures. Department of Computer & System Science, The royal Institute of Technology, Stockholm university, 67s.
- Maybury, M.T. (1996): Research in Multimedia and Multimodal Parsing nad Generation. To appear in *Journal of Artificial Intelligence Review: Special Issue on the Integration of Natural language and Vision Processing*.
- Mullet, K, and Sano, D. (1995): *Designing Visual Interfaces. Communication Oriented Techniques*. SunSoft Press, kap. 6: Image and Representation.
- Preece, J. m fl (1994): *Human-computer interaction*, Addison-Wesley, Wokingham.
- Velichkovsky, B., M. Pomplun, and J. Rieser (1995): Attention and Communication: Eye-Movement-Based Research Paradigms. *Forschungsberichte Band 11*. Dresden University of Technology.