Doctoral Dissertations (All Dissertations, All Years)　　　　　Electronic Theses and Dissertations

2001-04-24

VIL: A Visual Inter Lingua

Neil Edwin Michael Leemans
Worcester Polytechnic Institute

Follow this and additional works at: https://digitalcommons.wpi.edu/etd-dissertations

VIL: A Visual Inter Lingua

by

Neil Edwin Michael (Paul) Leemans

A Dissertation

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Computer Science

by

_____

April 2001

APPROVED:

_____
Dr. Lee A. Becker, Major Advisor

_____
Dr. David C. Brown, Committee Member

_____
Dr. Norman Wittels, Committee Member, Department of Civil and Environmental
Engineering

_____
Dr. Stanley S. Selkow, Committee Member

_____
Dr. Micha Hofri, Head of Department

Abstract

As the world becomes smaller through advances in telecommunications, the need for communication between speakers of different languages becomes greater. Concerns about cultural and economic hegemony argue against the use of any natural language, and machine translation is not yet perfected and available to speakers of all languages. With the technological developments of the last decade, such as powerful computers, graphical interfaces, and the World Wide Web, an excellent opportunity has been created for a computer-mediated visual interlingua to meet this need. An iconic language could be designed to take advantage of the technology. People would be able to communicate with an iconic language without the need to draw pictures themselves, since they could choose these pictures from the screen.

This dissertation describes VIL, an iconic visual interlingua based on the notion of simplified speech. Similar to pidgins, languages arising from the prolonged contact between people speaking two or more languages, VIL utilizes features that are in the 'greatest common denominator' of features in different languages. This allows its complexity to be significantly reduced; for example, it has no inflection, no number, gender, or tense markers, and no articles. VIL has no linear order. This is possible because it was designed as a visual language, in contrast to written languages which are the result of a transfer to visual modality of spoken language, which evolved in the context of auditory modality where sequencing and ordering is critical.

After reviewing previous research on universal languages that are artificial, non-artificial, and visual, VIL is described in detail, including its parts of speech, its grammar, and its organization for verbs, nouns, and adjectives. Throughout the discussion a set of principles is proposed, some of which are relevant to any universal language, others specific to visual or iconic languages. The development of a set of icons is also presented. Finally, the evaluations of the icons, language, and the system itself are described.

Acknowledgments

First I would like to thank my advisor Dr. Lee A. Becker for all his support and guidance during my studies and stay at WPI. In times of difficulty, he was always the one to say that everything was going to workout fine. Somehow, this always gave me renewed strength to continue with my research. I also thank the other members of the committee for reading and commenting on my dissertation, especially Dave Brown, who, with his critical notes now and then, helped me strive for perfection.

I really enjoyed my 4-year stay in the US and have grown quite fond on the country. The people I met were all very friendly and helpful, especially the Computer Science staff.

I am also greatly thankful to Rishi Thakoerdat from the Rotterdam Graphical College for helping me out with designing most of the icons and the VIL website. Many thanks to all employers that I had while doing my PhD. I started out as a teaching assistant at WPI, and then went to work for CLARiiON (Data General) in Southboro. In 1997 we moved back to the Netherlands and I began to work for Exact Software, nowadays I work as a manager at KPMG. Without their help I could not have funded my research.

Finally and definitely most importantly, I want to thank my wife Larissa for always being there for me, encouraging me, and giving her love and support. She helped me studying for the qualifying exams, helped me designing and drawing icons, and most of all gave me three beautiful children, Anfernee, Yentl, and Timothy. Without her none of this would have been possible. I love you!

1. Introduction

In 2002 the EURO becomes the currency that is recognized in all countries in the European Community. By means of this currency, people have, even in the smallest and most remote places, funds for immediate use. The intrinsic value of their national currency is just as great as it ever was, but because of the difficulties and delays of exchange, an international currency medium, which is immediately available and which passes at a definitely understood and agreed upon rate, is a great time saver and aid to travel.

Does this situation parallel that in languages? By means of an international auxiliary language, man could have the benefit of an exchange of ideas without the delay and inconvenience of translation. The exchange of ideas would then become immediate and in terms which would be commonly understood.[1]

The urge for a simple language, easy to learn, and the desire to communicate with people speaking a foreign language made itself distinctly felt about three hundred years ago. Since then people have speculated on the possibility of creating and adopting a universal language.

This dissertation is concerned with designing such a language.

1.1 A Brief History

In the past, several attempts have been made to create a universal language. Most of these languages were artificial languages, but also some existing languages have been proposed as a universal language. The attempts for universal languages can roughly be divided into three groups: artificial textual languages, non-artificial languages, and (artificial) visual languages.

1.1.1 Artificial Languages

It has been suggested that the best plan to get a world language adopted would be to have some great living language, e.g. English, to absorb all the other languages, since English is already the native tongue of hundreds of millions of people. It might also be Chinese, German, Russian, Italian, Portuguese, or perhaps French, which served as

---

[1] **Head, W. D.** *"Needed: A Spare Tire Language,"* Rotarian, 50:28-30, March 1937.

an international diplomatic language for about 500 years. Some even tried to resuscitate Sanskrit. However, each of these languages has its peculiar difficulties. English, for example, has a fairly simple grammar, but its pronunciation and spelling are notoriously irregular. Some internationally minded people have suggested that the leading nations of Europe agree on a common auxiliary language, but nationalism and racial pride are insuperable obstacles.[2] For example, the French are purposely limiting the amount of English on French television, because they want their people exposed to more French.

"The adoption of any modern national language by the common consent of the chief nations is therefore unlikely, as it would confer undue advantages and excite jealousy, however impartial the promoters of the language might be."[3]

Another plan was to create a language ad hoc, a plan already suggested by Descartes in a letter of November 20, 1629, but a practical start of large proportions did not begin until Schleyer initiated Volapük in 1879. It spread rapidly and soon there were hundreds of Volapük clubs, a dozen or more periodicals, and 300 to 400 publications in ten countries. Three international congresses were held and, according to Schleyer, two and a half million people were interested in it, but the boom suddenly collapsed. Schleyer remained stubborn to any suggestions of reform and his followers left him; however, a committee proceeded to revise Volapük, and under the leadership of the Russian engineer Rosenberger, it became "Idiom Neutral," which in turn became "Reformed Neutral," which was for the most part Neo-Latin. Thus Volapük was born in 1879, reached its peak in 1888, and was practically dead in 1890. It has an ingenious grammar, but an artificial vocabulary. Some other artificial languages, which either corrected or exaggerated defects of Volapük, were Spelin, Dil, Dilpox, Balla, Tal, and Pankel.[4]

The best known and most successful of the artificial languages is Esperanto, instituted in 1887 by the young Jewish oculist Louis L. Zamenhof, a native form Bialystok, Poland, where Jew and German, Pole and Russian jostled each other on the street and spoke mutually unintelligible languages and differed further in religion and custom. His brochure published in 1887 was signed "Dr. Esperanto" meaning Dr. Hopeful. Esperanto spread rapidly throughout Europe and the world; about four thousand books and one hundred periodicals were published in it. Twenty-seven European broadcasting stations were using it in 1927, and several world conferences were held including one in Washington at which it was spoken. However, World War II gave it

---

[2] **Kamman, W. F.** *"The Problem of a Universal Language,"* Modern Language Journal, 26:177-182, March 1942.

[3] **Johnson, J. E.** *"Excerpts,"* from Basic English, The Reference Shelf. Vol. 17, No. 1, The H. W. Wilson Company, N. Y., 1944, p. 58-68.

[4] **Kamman, W. F.** *"The Problem of a Universal Language,"* Modern Language Journal, 26:177-182, March 1942.

a serious setback. Critics say it has too many circumflexed or hooded letters, too many sibilants or hushing sounds, and unnecessary accusative endings, and is too Russian in character. Furthermore, many well-known words are mutilated. Others say it looks like spoiled Latin or a parody on Italian.[5]

Besides Esperanto, there are four other artificial languages that have won public attention. They are Ido, Esperanto-II, Occidental, and Latin without inflections. Like Esperanto all have regular grammars and vocabularies based on Indo-European languages. The main descendant of Esperanto, Ido, is an international scientific language based on the leading languages of Europe. Its main players were Dr. Couturat and M. Beaufront. Ido corrected the main defects of Esperanto, but it began to degenerate in 1914. In that same year, Dr. Couturat died and no competent leader arose to fill his shoes. Dr. Talmey eventually revised it and it became Arulo, and later Gloro. It has ceased to prosper. About 200 schemes and projects appeared thereafter, but none has attained a great and lasting popularity or importance.[6] Esperanto-II is a scheme for adapting Esperanto and was made widespread by Dr. Rene de Saussure from Switzerland.

While Esperanto, Esperanto-II, and Ido bear an obvious family likeness, Occidental, advocated in 1922 by Estonian Edgar von Wahl, grew out of efforts to construct a more "natural" simplified language. Latin without inflections, propagated in 1903 by the Italian Prof. G. Peano, is what its name implies. It uses classical Latin stem for forms, supplemented by a certain number of modern international words.[7]

1.1.2 Non-Artificial Languages

Besides the artificial languages there have also been attempts to simplify existing languages. These are the Basic languages, of which Basic English has attracted the most popularity. Although, as mentioned, Basic English is not considered an artificial language, it was still designed by someone (C. K. Ogden), and did not grow naturally.

Resembling simplified English, but quite different, are the (English) Pidgins. These are languages that come into existence in multi-lingual areas. They are not made up by any one individual, but develop naturally where people with different linguistic backgrounds come into contact with each other.

---

[5] **Kamman, W. F.** *"The Problem of a Universal Language,"* Modern Language Journal, 26:177-182, March 1942.

[6] **Kamman, W. F.** *"The Problem of a Universal Language,"* Modern Language Journal, 26:177-182, March 1942.

[7] **Morris, A. V.** *"A Language Cut to Fit the World,"* Rotarian, 53:36-39, November 1938.

Pidgins and Basic English will be discussed in more detail in chapter 2 in order to show that it is quite possible to communicate using a restricted grammar and a limited vocabulary.

1.1.3 Visual Languages

Lastly, there are the visual communication languages. These are languages that do not use words, but solely rely on signs, symbols, pictures, or icons. The Chinese writing system started out as a pictorial language (but too complex to serve as a universal language). The ancient Egyptians with their hieroglyphs and the American Indian also communicated through a pictorial language. We will argue for a universal computer-based iconic language. As we will see, visual languages overcome most of the problems that written languages encounter. Furthermore, with the technological developments of the last decade, such as powerful computers, graphical interfaces, and the World Wide Web, an excellent opportunity has been created for a computer-mediated visual interlingua to meet the need for communication between speakers of different languages, which becomes greater as the world appears smaller due to advances in telecommunications. This iconic language should be designed to take advantage of the technology. People will be able to communicate with an iconic language without the need to draw the pictures themselves, since they can choose these pictures from a screen.

1.2 The Organization of the dissertation

The dissertation consists of seven chapters and a bibliography. Chapter one is this introductory chapter. The four remaining chapters can roughly be divided into two parts. Part one only consists of chapter two and concentrates on previous work. Part two consists of chapters three through seven, and concentrates on iconic communication and especially VIL. Throughout the dissertation a set of principles are proposed, some of which are relevant to any universal language, others specific to visual or iconic languages. These principles make explicit much of the design rationale of VIL, and it is hoped that they will serve to stimulate further discussion and debate on these issues. The principles will be designated in the following manner:

Principle X.Y: *Name of the Principle*

> ....
> ....
> ....

where X stands for the chapter number the principle occurs in, and Y stands for the number of the principle in that chapter.

Part I: Universal Languages

In this part of the dissertation the non-artificial languages and visual languages, from section 1.1.2 and 1.1.3 respectively, will be discussed. For the non-artificial languages, the emphasis will be on Basic English and Pidgins and Creoles. We will study these in order to show that it is very well possible to communicate using a restricted grammar and a limited vocabulary. The rest of part I is about visual communication languages. The languages studied are not all attempts at creating a universal language, e.g. hieroglyphs and Chinese, but are discussed to explore the world of pictorial languages. Traffic signs, which form a language of their own will be addressed as well, and various examples of traffic signs are given. The last part of the chapter discusses icons and computer-based iconic languages, since this is the area that is of most interest to us, and describes several iconic systems, some are not computer-based (like IICS, Isotype, and Blissymbolics), others are more recent and are computer-based (like The Hotel Booking System, CD-Icon, and The Elephant's Memory).

Part II: VIL: Visual Inter Language

This part of the dissertation focuses on the development of our iconic communication language. Chapter three discusses the semantics of the language. Various principles for VIL are stated and VIL's grammar is described in detail. The classification of verbs, nouns, and adjectives are researched and the resulting hierarchies are given.

In chapter four, the focus is on the representation of the system. The chapter begins by discussing icon design, and gives various examples of good and bad icons along with principles to improve icon design. At the end of the chapter the representation for the grammatical entities (e.g. verbs, nouns, adjectives, etc.) will be discussed (i.e. what will be represented visually and what will be represented iconically). We will talk about how to represent abstract versus concrete meaning, and how to represent categories as opposed to instances of a category.

Chapter five explains in detail the computer program that was designed for VIL. It describes the techniques used and how the program works. Chapter six, contains an evaluation of the system. It describes how the various criteria can be tested. It also describes how the icons, developed for VIL, and the program itself were tested together with the results of the test. The last chapter, chapter seven, contains a conclusion.

1.3 Contributions

Our research into universal languages and iconic communication, together with our program that implements an iconic language, makes several contributions:

1. We have proposed a set of principles, some of which are relevant to any universal language, while others are specific to visual or iconic languages. These principles can be used by others in the field as guidelines or at least can serve as a starting point for discussion and debate. Many of these principles, however, cannot be evaluated, but have to be argued for.
2. We have designed a language and an internationally recognizable set of icons to represent both abstract and concrete meanings for both category icons and icons that represent a single item (terminal icon).
3. We have implemented an iconic communication language, VIL.
4. VIL will demonstrate the viability of a universal visual inter language even if every aspect of VIL is not perfect.
5. We will identify those areas, which need additional work by means of our evaluation, for example intelligent aids to help the user.

Before we start by describing some universal languages, we will first give some conditions and criteria to which such a language must conform.

1.4 Criteria for a new language

Richards[8] mentions three conditions that have to be met in order for a world language to be accepted:

1. **Political conditions**: They include free adoption, absence of all threats of domination of any type, protection for primary languages, and symbolization of supranational (international) aims. The political considerations seem to argue against any existing national language as a candidate.
2. **Psychological conditions**: The root criticism of any revived or artificial language, however well designed, is that the immediate incentive, which would make enough people learn and use one is lacking. If you are to go through the trouble of learning a language you need to feel that you will get a return for your toil this very year.
3. **Linguistic conditions**: Languages are shaped by use rather than by design. Long-established languages have been hammered and bent, broken and remade, in

---

[8] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

countless ways, which only the biggest and best of dictionaries can show in detail. No artificial language can acquire a part of such richness of interdependence (of our words), without centuries of wide and varied use. Ogden stated that with a simplified form of a living language these discrepancies are reduced (arguing for Basic English).

Besides the previously mentioned conditions, there are four principles that we consider important when designing a new visual language:

Principle 1.1: **Learnability**
>   The language and the use of its delivery system should be easy to learn.

Principle 1.2: **Encodability**
>   It should be easy to compose messages using the language and its delivery system.

Principle 1.3: **Decodability**
>   Messages composed in the language should be easy to comprehend, i.e. should let the user decode the messages without much effort.

Principle 1.4: **Extensibility/Evolvability**
>   The language should have the possibility to be extended and to evolve.

2. Universal Languages & Linguae Franca

This chapter mainly focuses on other people's work. Pidgins and Creoles and Basic English are researched to show that it is possible to communicate with a restricted grammar and a limited vocabulary, and ways are researched in which these restricted languages can be extended. Pidgins and Creoles will be the topic of section 2.1 and Basic English will be the topic of section 2.2.

In section 2.3 visual communication languages are described. The section starts with an explanation of why a visual language is a desirable approach and then gives a brief background on semiotics and icons. It will then describe some pictorial and iconic languages. Chinese, hieroglyphs, and traffic signs are divided into three groups: concrete, abstract, and composite signs. Then some attempts at universal iconic languages that did not use computers are described, like IICS, Isotype and Blissymbolics. Finally, some computer-based iconic systems are described, like The Hotel Booking System, CD-Icon, and more recently The Elephant's Memory. Section 2.3.4 on CD-Icon, is of great interest to us since it is also an attempt to create a computer-based iconic communication language. CD-Icon was developed at the University of Exeter and is based on Schank's Conceptual Dependency theory.

2.1 Pidgins and Creoles

Unlike Esperanto or Ido, which are universal languages that are designed a priori by certain individuals, pidgins and creoles are communication languages that develop naturally and come into existence out of economic necessity in multilingual areas where people do not speak each other's language.

This section will concentrate on pidgins and creoles with the emphasis on pidgins. It will try to show that it is possible to communicate with a drastically reduced grammatical structure and lexicon. This will support the design of VIL, which, much like pidgins, uses the greatest common denominator of several languages. Some of the simplifications used in pidgins (e.g. uninflected forms, no articles, no copula) might be appropriate in a visual communication language like ours.

2.1.1 Simplified Speech

Every speech community has forms of speech for use with people who are regarded (for some reason) as unable to understand the normal speech of the community. These might be babies, foreigners or deaf people. These forms of speech, like *baby talk*, are simplified versions of the language, and are thought of as being easier to understand and as being the way children speak. Baby talk is a linguistic subsystem regarded by a speech community as being primarily appropriate for talking to young children. It consists of a special set of lexical items. The special lexical items typically number between 25 and 60 and cover kin names and appellations, bodily functions, certain simple qualities (e.g. dirty, pretty, hot, cold), and vocabulary concerning animals, nursery games, and related items. Baby-talk words typically contain stops, nasals, and a limited selection of vowels, are frequently duplicated, and often have a diminutive suffix characteristic of baby talk in that language.[9]

Another form of speech, which is widespread and maybe even universal, is *foreigner talk*. It is used by speakers of a language to outsiders, who are felt to have very limited command of the language. Particular features of pronunciation, grammar, and lexicon are used in this situation. For example, a speaker of Spanish who wishes to communicate with a foreigner will typically use the infinitive of the verb or the third singular rather than the usual inflected forms. He/She will use *mi* 'me' instead of *yo* 'I' and will omit the definite and indefinite articles: *mi ver soldado* 'I see the soldier'.[10]

Sometimes a speech community deliberately uses foreigner talk to outsiders, because they do not want them to acquire the language. Dutton[11] reports how the Motu deliberately withheld knowledge of their real language in dealing with foreigners; instead they produced a kind of simplified Motu foreigner talk. The Chinook also have a haughty attitude towards their language, even within their own speech community. The Chinooks believed that non-Chinooks should not properly learn true Chinook.[12] Foreigner talk plays an important role in pidginization (pidgins have a much reduced range of linguistic functions to perform in comparison with native language).

---

[9] **Ferguson, C. A.** *"Baby Talk in Six Languages,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 113-133.

[10] **Ferguson, C. A.** *"Absence of Copula and the Notion of Simplicity: A study of Normal Speech, Baby Talk, Foreigner Talk, and Pidgins,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 277-292.

[11] **Dutton, T.** *"Police Motu: iena sivarai (its story)."* Port Moresby: University of Papua New Guinea Press, 1985.

[12] **Hymes, D.** Commentary. In A. Valdman and A. Highfiled (eds.) *"Theoretical orientations in language studies."* New York: Academic Press, 1980.

Besides baby talk and foreigner talk, there are more forms of simplified speech. Forms whose limitations are irrelevant, and that are used only to save space, time, or money. Telegrams resemble baby talk and foreigner talk in omitting definite article, prepositions, and copula, and the resemblance of these usages to early childhood language behavior has been noticed.[13] Other examples include instructions printed on packages where there is plenty of space, and newspaper headlines.

The notion of simplicity in language is important, since it may be related to theories of language universals and language acquisition. Jacobson and others have assumed that the simpler of two comparable features is likely to be more widespread among languages of the world, and also the earlier acquired in child language development.[14]

Ferguson[15] posed the following hypothesis: "If a language has an inflectional system, this will tend to be replaced in simplified speech such as baby talk and foreigner talk by uninflected forms (e.g. simple nominative for the noun; infinitive, imperative, or third person singular for the verb)." Also, studies of child language development seem to show that children first make equational clauses without a copula and only later acquire the construction with the copula.

2.1.2 Pidgins and Creoles

One term that comes to mind when discussing multilingual contact situations and pidgins is the term "*lingua franca*", so let me start by giving some definitions.

- A "*lingua franca*" stands for any language that is used as a medium of communication among people, who have no other language in common. The language used is not native to any of the groups involved. So, when a Telegu-speaking Indian and a Hindi-speaking Indian converse together in a bus station using English, then English serves as a lingua franca.[16]

---

[13] **Brown, R. W. and Bellugi, U.** *"Three processes in the child's acquisition of syntax."* Harvard Educ. Rev. 34, pp. 133-151 (Repr. in "New directions in the study of language," Lenneberg, E. (ed.), Cambridge, Mass., 1964).

[14] **Ferguson, C. A.** *"Absence of Copula and the Notion of Simplicity: A study of Normal Speech, Baby Talk, Foreigner Talk, and Pidgins,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 277-292.

[15] See note 14.

[16] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

It often happens that, to communicate with each other, two or more people use a language whose grammar and vocabulary are very much reduced and which is native to neither side. This is what constitutes a *pidgin*.

- A *"pidgin"* is a contact language that is a mixture of linguistic elements of two or more languages. It arises in social and economic transactions between at least two groups, who speak different languages, by a process of restriction and simplification of one of the languages of these groups, usually the one in a social superior position.[17]

At least two languages must be involved in the creation of a pidgin: that which supplies the bulk of the pidgin language's vocabulary and is spoken by a socially superior group (the superstrate language) and that which contributes less obviously and is spoken by a socially inferior group (the substrate language). There is normally only one superstrate language (e.g. English, Dutch, Spanish), but commonly multiple substrate languages. [18]

By definition, a pidgin is no one's native language. Two conditions must be met to call a language a true pidgin. The first one is that its grammatical structure and its vocabulary must be sharply reduced. For example, when a Dutchman speaks to Turkish immigration workers- *Dit mooi schilderij Rembrand zijn* (This be beautiful picture Rembrand)- grammatical particle is lost, the infinitive is used instead of an inflected form of the verb, and the preposition *door* "by, of" is not used before Rembrand), and the addressees speak to him in the same kind of Dutch. The second condition is that the resultant language is not native to any of those who use it. So the pidgin Dutch is not native to the Dutch man speaking it (i.e. he should not be leaving out inflection, particles, and prepositions in his every-day speech with other Dutchmen).

The syntactic structure of a pidgin is less complex and less flexible than the structure of the languages that were in contact, and though many pidgin features clearly reflect usages in the contact languages, others are unique to the pidgin. Pidgins have discarded many of the *inessential features* of the standard variety (see table 2.1).[19]

---

[17] **Foley, W.A.** *"Language birth: the processes of pidginization and creolization,"* Chapter 9.

[18] **Foley, W.A.** *"Language birth: the processes of pidginization and creolization,"* Chapter 9.

[19] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

French	English	Neo-Melanesian		Cameroon Pidgin	
Je vais	I go	mi	go	a	go
tu vas	you go	yu	go	yu	go
il/elle va	he/she/it goes	em	go	I	go
nous allons	we go	yumi/mipela	go	wi	go
vous allez	they go	yupela	go	wuna	go
ils/elles vont		ol	go	dem	go

Table 2.1

All natural languages have some degree of redundancy. As can be seen from the table, English has less verbal inflection than French, but both pidgins have an invariable form. Furthermore, in many European languages, plurality is marked in the article, the adjective, and the noun, as well as by a numeral (sometimes). Consider for example the sentence "L*es deux* grand*s* journ*aux*". Here there are four evident markers of plurality (in written form, three in spoken form). English is a little less redundant than French, but still in "The *two* big newspaper*s*" there are two markers of plurality. The pidgins Neo-Melanesian (The pidgin English of Papua New Guinea) and Cameroon pidgin are less redundant still. They only mark plurality by the numeral, *tupela bikpela pepa* and *di tu big pepa* (resp. *two big newspaper* and *the two big newspaper*).

A *creole* arises when a pidgin becomes the mother tongue of a speech community. The simple structure of the pidgin is carried over into the creole, but since a creole, being a mother tongue, must be capable of expressing the whole range of human experience, the lexicon is expanded and frequently a more elaborate syntactic system evolves.

- A *creole* is a pidgin, which has become the native language of a speech community. In the process of becoming nativized, the pidgin undergoes extension and elaboration.

Consequently, creoles are normally contrasted with pidgins by their greater functional and structural complexity and their stability in usage.[20]

A creole can develop in two ways. First of all, speakers of a pidgin may be put in a position where they no longer can speak their mother tongue (e.g. in slave trade, people from the same area where kept apart from each other to prevent plotting). Consequently, children born in this situation acquired the pidgin as their first language and so a creole came into existence. A creole, however, is not always the result of people being deprived the opportunity to use their mother tongue. A pidgin

---

[20] **Foley, W.A.** *"Language birth: the processes of pidginization and creolization,"* Chapter 9.

can become so useful as a community lingua franca that it may be expanded and used even by people who share a mother tongue. In this way children may acquire it as one of their first languages. In theory, the distinction between a pidgin and a creole is that a pidgin is no one's first language whereas a creole is. [21]

From a structural viewpoint, the essential characteristic of a pidgin language is that it is sharply reduced in its grammatical structure and vocabulary. In general, this reduction is in the direction of whatever features are common to the languages of all those using the pidgin, for mutual ease in use and comprehensibility, thus arriving at a kind of **greatest common denominator**.[22] In the development of a creole out of a pidgin, on the other hand, the main change is in the direction of re-expansion of both structure and vocabulary.

Todd[23] distinguishes between 'restricted' and 'extended' pidgins. A restricted pidgin is one which arises as a result of marginal contact such as for minimal trading, which serves only a limited purpose and which tends to die out as soon as the contact which gave rise to the pidgin is withdrawn. An extended pidgin is one which, although it may not become a mother tongue, proves vitally important in a multilingual area, and which is extended and used beyond the original limited function, which caused it to come into existence.

Pidgins generally have the following two functions: [24]

- *Directive function*; to get people to perform desired actions within a plantation or trading context.
- *Referential function*; to describe a concrete situation to bring about a goal.

They usually do not serve the following functions:

- *Interactional function*; to promote social cohesion.
- *Expressive function*; to express abstract ideas and inner states.
- *Metalinguistic function*; to talk about language.
- *Poetic function*; to create language based art forms.

Much of the complexity of native languages is associated with the expression of these latter functions.

---

[21] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

[22] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

[23] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

[24] **Foley, W.A.** *"Language birth: the processes of pidginization and creolization,"* Chapter 9.

2.1.3 How does the simplification come about?

In his work on foreigner talk, Ferguson[25] points out that it is part of the communicative competence of adult speakers in many speech communities to know how to simplify their language to aid comprehension by non-fluent speakers.
Table 2.2 displays the features of simplified language forms in contrast with those of the source language.[26]

The relationship between form and meaning should be as transparent as possible. There is no need to cope with variations and irregularities.

More complex or unsimplified linguistic structure	Simpler or simplified linguistic structure
Lexicon	
Larger vocabulary in a given domain or overall	Smaller vocabulary, generic terms rather than specific
Compounds and morphologically complex words	Monomorphemic words, paraphrases of complex words
Syntax	
Subordinate clauses	No subordinate clauses
Variable word order, conditioned by syntax (e.g. negation)	Invariable word order
Presence of copula, pronouns, function words	Absence of copula, pronouns, function words
Morphology	
Extensive inflectional systems	Heavily reduced or no inflections
Allomorphy, including of stems	No allomorphy, invariant stems (e.g. full forms as opposed to contradictions)
Phonology	
Polysyllabic words	Monosyllables

Table 2.2.

So, by now it should be clear that pidgins and creoles often fail to show many of the familiar categories of grammatical inflection, especially number, case, gender, and tense. This was one of the major types of simplification that speakers of European languages considered necessary in order to make themselves understood by the natives with whom they came into contact. But these particular categories are far from universal and essential, and pidgins and creoles develop categories of their own.

---

[25] **Ferguson, C. A.** *"Absence of Copula and the Notion of Simplicity: A study of Normal Speech, Baby Talk, Foreigner Talk, and Pidgins,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 277-292.

[26]**Ferguson, C. A.** *"Simplified registers and linguistic theory,"* in L. Obler & L. Wise (eds.) Exceptional language and linguistics, New York: Academic Press, 1982.

- **Number**: There are no synthetic plurals of the 'man/men' or 'computer/computers' type. Instead in pidgins and creoles nouns are invariable, like 'sheep' (see table 2.3). Plurality is normally implicit in the context.

English	Neo-Melanesian	Cameroon Pidgin
one man/person	wanpela man	wan man
ten men/people	tenpela man	ten man

Table 2.3.

- **Gender:** In pidgins and creoles gender distinctions are reduced or eliminated in both nouns and pronouns.

- **Inflection:** There is no concordial agreement between subject and predicate in pidgins and creoles. Both noun and verb are invariable, where the singular is used for nouns and the imperative for verbs. Since the verb form is invariable, distinctions relating to time and continuity of action are either understood from the context or are indicated by adverbials or a set of free morphemes, which precede the verb.

- **Possession:** The only way of indicating possession in Neo-Melanesian is by PREPOSITION + OBJECT:

haws bilon mi "my house".

- **Relative pronouns:** Neo-Melanesian has no relative pronouns, adjectives, or adverbs. So "the meat (which) you ate" is *abus ju kajkaj*.[27]

- **Questions:** Question transformations are absent. Neo-Melanesian interrogatives all maintain the same word order:

Yu wokim haus?   "You building house?"
Yu mekim wanem?   "You doing what?"[28]

- **Negation:** In Haitian Creole, negation is done by placing *pa* in front of whatever needs negated:

u pa-gaso   "you are not a regular fellow"[29]

---

[27] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

[28]**Schumann, J.** *"The Pidginization Process: A Model for Second Language Acquisition,"* Newbury House Publishers, 1978.

- **Verbs and tense:** In Haitian Creole all contrasts of number, person, and tense have been lost in the verb, and every verb has only one form used in inflection:

> *gade* "keep, kept, will keep, etc."
> *domi* "sleep, slept, will sleep, etc."[30]

In English-Japanese pidgin, adverbs are used to indicate time:

> *asta* "tomorrow"
> *kino* "yesterday" [31]

Haitian Creole has developed a new set of inflectionally bound prefixes. These prefixes do not show tense, but aspect, i.e. the type of action that is involved: whether it is going on and not yet finished (imperfective, durative), or is finished and completed (perfective, completive). The imperfective-durative prefix *apr-* indicates that an action is going on, continuing, not yet complete, or future:

> *m-apr-ale* "I am going"
> *n-ap-sate* "we are sitting"[32]

The perfective-complete prefix *fek-* means "to have just":

> *m-fek-rive* "I have just arrived"[33]

2.1.4 Extending the Vocabulary

There are a number of ways by which the vocabulary can be extended.

- **Reduplication** is a very simple method of extending the vocabulary of a pidgin.

  Reduplications in pidgins serve three purposes (see table 2.4) [34]:

---

[29]**Schumann, J.** *"The Pidginization Process: A Model for Second Language Acquisition,"* Newbury House Publishers, 1978.

[30] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

[31]**Schumann, J.** *"The Pidginization Process: A Model for Second Language Acquisition,"* Newbury House Publishers, 1978.

[32] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

[33] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

1. to reduce the number of homophonous forms.
2. to extend the meaning of the simple form.
3. as intensives, this type being confined to the adjective/verb class.

Atlantic	Pacific

1. san   - sun
   sansan   - sand
   was   - wash
   waswas   - wasp
	pis - peace
	pispis - urinate
	sip - ship
	sipsip - sheep

2. ben   - bend
   benben   - crooked
   wan   - one
   wanwan - one by one
	sing - sing
	singsing - ritual singing
	was - wash
	waswas - go to wash

3. krai   - cry
   kraikrai - cry continuously
   fain   - lovely
   fainfain - very lovely
	tok - talk
	toktok - chatter
	bik - big
	bikbik - huge, very big

Table 2.4.

- **Word compounding** is another way to extend the pidgin's vocabulary (see table 2.5) [35].

Atlantic	Pacific
bigman (big+man) important person daiman (die+man) corpse bigai (big+eye) greed draiskin (dry+skin) thin	bigsonde (big+Sunday) feastday daiman (die+man) corpse bigmaus (big+mouth) conceited draibon (dry+bone) tough

Table 2.5.

One of the compound types is NOUN + NOUN, in which the first noun tells some characteristic or feature of what is referred to by the second noun. There is, however a difference between one compound word (consisting of 2 nouns) or those same two nouns not compounded:

hawsboj (one word, meaning "houseboy") and
haws boj (two words, meaning "house for the boys") [36]

Another compound type is ADJECTIVE + NOUN, and it has a transferred or figurative meaning:

bigmaus "conceited" vs. *bigfela maus* "large mouth"

---

[34] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

[35] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

[36] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

blaekman "Malesian" vs. *Blaekfela man* "dark-colored man" [37]

- **Deverbal nouns**: Almost any verb can be used as a noun. Thus *maze* "eat" can also mean "food, meal".

As the vocabulary is expanded, the pidgin becomes more useful in a wider set of communication situations.

2.1.5 Universals of Contact Situations

dat bele we rebren komot, na di tifman komot
(that belly which reverend come out, be thiefman come out)
i.e. very different exteriors can hide similar origins. [38]

Todd poses a hypothesis that there are universal patterns of linguistic behavior appropriate to contact situations. People of different linguistic backgrounds adjust their language behavior in similar ways. This suggests that the behavior is rule-governed and may be the result of linguistic universals. Children, sailors, traders, and Portuguese adventurers might all have been responding to linguistic situations according to an innate behavioral 'blueprint'. He further poses that if it is possible to show that human beings have predetermined propensities for acquiring language, then it may well be that the capacity for linguistic simplification and accommodation - the process that produces pidgins - is also innate and universal. Then it is also likely that there are linguistic universals common to all languages, irrespective of their surface manifestations. The Krio proverb quoted above, underlines the importance of remembering that languages differ in form, not in fundamentals. [39]

When acquiring language, children produce patterns, which are regular for them, but irregular for adults. In such cases children cannot be said to be imitating adults. Studies in unrelated languages like English, Congolese Luo, and Finnish suggest that a child's first negatives are regularly 'negator plus sentence', e.g. 'no' plus 'I go'. There is a rule here, and since it is not the adult rule, it would appear to be the child's own contribution.

Furthermore, many parents automatically simplify their language when speaking to very young children. They cut down on sentence subordination, use simple tenses, and repeat new words. And they do this without being fully aware of what they are

---

[37] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

[38] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

[39] **Todd, L.** *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

doing and how they accomplish it. In other words, their behavior seems to respond to a built-in simplification mechanism.

Smith states that both the child in early native language acquisition and the pidgin speaker reduce and simplify language to which they are exposed into a set of primitive categories, which are undoubtedly innate. These primitive categories emerge in speech as utterances relatively unmarked by inflections, permutations, and functions.[40]

Corder maintains a similar position, but argues that simple codes used in child talk, foreigner talk, and pidgins are not simplified. Instead, they represent a basic language, which is expanded and complicated in the process of learning. Simple codes are nearer to the underlying structure or 'inner form' of all languages.[41]

Schumann states that the codes may result from a regression to a set of universal primitive linguistic categories that were realized in early first language acquisition. Then under conditions of social and/or psychological distance, the pidginized form of speech persists.[42]

2.1.6 Summary

Contrasted with the "major" languages of Western Europe, pidgins and creoles have relatively little morphological variation in their structure. Many grammatical concepts that, in our familiar languages, are expressed by morphological (especially inflection) features, find their expression in syntactical combinations.

From a structural point of view, a pidgin represents the very first stage of rudimentary language learning.

A pidgin serves a function that no other means of communication could duplicate: that of facilitating mutual understanding on a limited range of topics.[43]

---

[40]**Smith, D. M.** *"Pidginization and language socialization: the role of marking,"* Unpublished paper, Georgetown University, 1973.

[41]**Corder, S. P.** *"Simple codes' and the source of the second language learner's initial heuristic hypothesis."* Paper presented at the Colloque 'Theoretical Models in Applied Linguistics' IV, Universite de Neuchatel, 1975.

[42]**Schumann, J.** *"The Pidginization Process: A Model for Second Language Acquisition,"* Newbury House Publishers, 1978.

[43] **Hall, R.** *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

A pidgin, however, would never be accepted as a universal language, because a lot of people think that pidgins are socially inferior. It does however show that it is very well possible to communicate with a drastically reduced grammatical structure and vocabulary. Also an attempt has been made (see section 2.1.5) to show that reduced language is closer to our universal roots and more innate than any of the modern languages.

2.2 Basic English

When we think of artificial languages, we think of languages like Esperanto and Ido. Basic English is not considered to be an artificial language. However, in my opinion Basic English is an "artificial language", in so far that is was designed by someone (C. K. Ogden) and did not develop naturally. It is not really artificial like Esperanto, because it is a restricted version of the existing and certainly not artificial language, namely English. When we hear "a simplified version of English", some people might confuse it with English pidgins. The only thing they have in common is that they are both restricted languages, and that they are both no one's native tongue. The main difference, however, between Basic English and English based pidgins is that at least two languages must be involved in the creation of a pidgin: that which supplies the bulk of the pidgin language's vocabulary and is spoken by a socially superior group (the superstrate language) and that which contributes less obviously and is spoken by a socially inferior group (the substrate language). Basic English, however, is just a simplified version of English. Another difference is that pidgins developed naturally in multilingual areas, whereas Basic English was designed for foreigners as a simplified way to learn and communicate in English. It was even proposed to be a universal language.

Basic English is English made simple by limiting the number of words to 850 and by cutting down the rules for using them to the smallest number necessary for the clear statement of ideas. This is done without change to the normal order and behavior of these words in everyday English.

2.2.1 Why Basic English?

Richards[44] mentions three conditions that have to be met in order for a universal language to be accepted. These are political, psychological, and linguistic conditions. Some of the necessary political conditions are free adoption, absence of all threats of domination of any type, protection for primary languages, and symbolization of supranational (international) aims. The political considerations seem to argue against

---

[44] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

any existing national language as a candidate. Let us turn to the psychological conditions.

The idea of a universal language has kept the minds of people busy for centuries. Several artificial languages have been proposed. The root criticism of any revived or artificial language, however well designed, is that the immediate incentive, which would make enough people learn and use one is lacking. If you are to go through the trouble of learning a language you need to feel that you will get an immediate return for your toil.

As for linguistic considerations, languages are shaped by use rather than by design. Long-established languages have been hammered and bent, broken and remade, in countless ways, which only the biggest and best of dictionaries can show in detail. No artificial language can acquire a part of such richness of interdependence (of our words), without centuries of wide and varied use. With a simplified form of a living language these discrepancies are reduced. Its meanings are held in place by the extent of the common use its words have been put to.

So according to those people in favor of Basic English, if there is to be a common language it must be a simplified form of one of the world's existing major languages. That would give the learner immediate access to innumerable speakers. It would lead into the parent language and give the learner admission to a vast literature. Thus, two of the three requirements would have been met (and people in support of Basic seem to think that these outweigh the political condition).

English is able to cover a vast range of needs with a very small vocabulary and a very simple set of constructions. Three times as much Spanish, it is estimated, would be needed to cover comparable ground. The conjugation of the verb by inflection does, it is true, make personal pronouns as subjects unnecessary except in an occasional ambiguous situation. Yet this economy of pronouns is far outweighed by the increased learning load of the verb changes themselves.[45] English has no genders for nouns, and an adjective takes the same form whether applied to a male or a female.

2.2.2 What Basic English is

Ogden was struck by the fact that whatever you are defining, certain words keep coming back into your definitions. Define them, and with them you could define anything. This suggests that there might be some limited set of words in terms of which the meanings of all other words might be stated. If so, then a very limited

---

[45] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

language (limited in its vocabulary but comprehensive in its scope) would be possible.[46] This led Ogden to the creation of Basic English.

On examining 400,000 words of writing by 2500 Americans, Dr. L. P. Ayres found that the 50 commonest words accounted for more than half the total number of words used, that 250 more accounted for another 25 percent, and that 1000 accounted for 90 percent. Ogden argued that the language could be spoken intelligibly with even less than 1000 words.[47]

His selection of words was not based on the question: "How frequently is this word used?" but "In what ways will this word take the place of what other words?"[48] They are the ingredients of all the ideas for which we normally use all the language. They are words with which all other words in English may be defined or explained. High frequency suggests that the word may be important for a limited vocabulary. It does not prove that the word is necessary.[49]

He claimed that even with such a small word list and such a simple structure (as Basic English has) it would be possible to say in Basic English anything needed for the general purposes of everyday existence - in business, trade, industry, science, and medical work.

In the end three principles came out clearly. First, that Basic English must be an all-purpose language and serve trade, commerce, technical education, as well as news, politics, general knowledge, and discussion at simple levels of all the common affairs of man. Secondly, that it must conform to current English usage. There must be nothing in it which would have to be unlearned by those going on from it to a more complete mastery of English. Thirdly, it was to be as limited in vocabulary and as simple, intelligible, and regular in syntax as is compatible with these other aims.[50]

---

[46] **Ogden, C. K.** *"Basic English: A General Introduction with rules and grammar,"* London: Kegan Paul, Trench, Trubner & Co. Ltd., Broadway House, Carter Lanes E.C., 1930.

[47] **Mencken, H. L.** *"The American Language,"* Copyright, Alfred A. Knopf, New York, 1936, p. 603-607.

[48] **The Orthological Committee**, *"Notes on Basic English,"* Number One, A short Account of the System, October 1940.

[49] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[50] **Richards, I. A.** *"Basic English and Its Applications,"* Royal Society of Arts. Journal (London), 87:735-755, June 2, 1939.

What resulted were the wordlist of 850 words and the ordered system, which restricted their use and idioms to a limited range. Most of the 850 words can be taught in their first and root senses by pictures, pointing, or going through the motions.[51]

The 850 words in Basic English are divided into three main classes. There are 600 names of things, of which 400 are general things and 200 are things that may be pictured. There are 150 names of qualities ('adjectives"), mostly given in pairs of opposites, such as "first-last," and "early-late." Basic English encourages teaching words in logical groupings - it is possible to carry grouping by opposites into the nouns (front-back) and the prepositions (before-after, under-over), and the verbs (come-go, etc.). Then there are 100 "operations" (words that put the others into significant relationship with one another; the acts and directions).

The most striking feature of Basic English is its limitation of the verbs to sixteen, with the auxiliaries *may* and *will*, which are auxiliaries of possibility and permission, and of futurity. Twelve of them are verbs of simple bodily or manual action - *give*, *get*; *take*, *put*; *come*, *go*; *keep*, *let*; *make*, and *say*, *see*, and *send*. The other four are *do*, *have*, *be*, and *seem*. Other verbs are covered by the use of one of these verbs with some limiting word, mainly the name of a direction or position (the verb "to enter" becomes "to go in").[52]

Ogden came to this small set of verbs, because all English verbs may be broken down into the operations and directions they are talking about. For example, to *immerse* your hands in water is to put them under water; to *illustrate* your arguments is to give examples; if you *dust* the table, you take dust off it. This made it possible for Basic English to have only 18 verbs (the words for these necessary physical motions and *have* and *be*), in addition to the names of directions (*to*, *from*, *on*, *in*, *under*, and so on), and the names of the things to which or with which the act is done.[53]

These operators, in combination with other Basic English words, translate more than four thousand verbs of full English. For example:

abandon, abdicate, abjure, cede, desert, desist, forgo, forsake,..............,relinquish, renounce, resign, vacate, withdraw, and *yield* are all replaced by *give up*.

---

[51] **The National Commission on the Enrichment of Adult Life**, *"A General View of the System of Basic English,"* Committee Report on Basic English for Adult Education, Washington, D. C. mineo p. 2-19.

[52] **Richards, I. A.** *"Basic English and Its Applications,"* Royal Society of Arts. Journal (London), 87:735-755, June 2, 1939.

[53] **The Orthological Committee**, *"Notes on Basic English,"* Number One, A short Account of the System, October 1940.

These 18 verbs are used as in normal English, undergoing whatever changes of form are necessary in different relations. Thus *give* is the headword, under which might be listed in a full table *gives*, *gave*, *giving*, and *given*. The same thing is true of the pronouns (*I*, *he*, *you*, *this*, and *that*). Similarly, *who* covers *whom*, *whose*, *which*, and *what*. The Word List is not a manual of Basic English, but the briefest, most compact possible specification of a language.[54] As in normal English, the addition of *s* is made for plural, and *-er* and *-est* is added to names of qualities as a sign of degree. Adverbs are formed by putting -ly at the end of names of qualities. Opposites are formed by putting *un-* in front. Three hundred of the names of things may take the endings *-er* (gives the name of the agent), *-ing* (gives the nouns for the action and corresponding adjectives), and *-ed* (gives the past particles and the passives of the 300 words)[55]

In their central uses, the 20 directives (*about* , *across*, ... , *up*, *with*) all have to with position or direction in space, see Figure 1, taken from Richards, 1943.
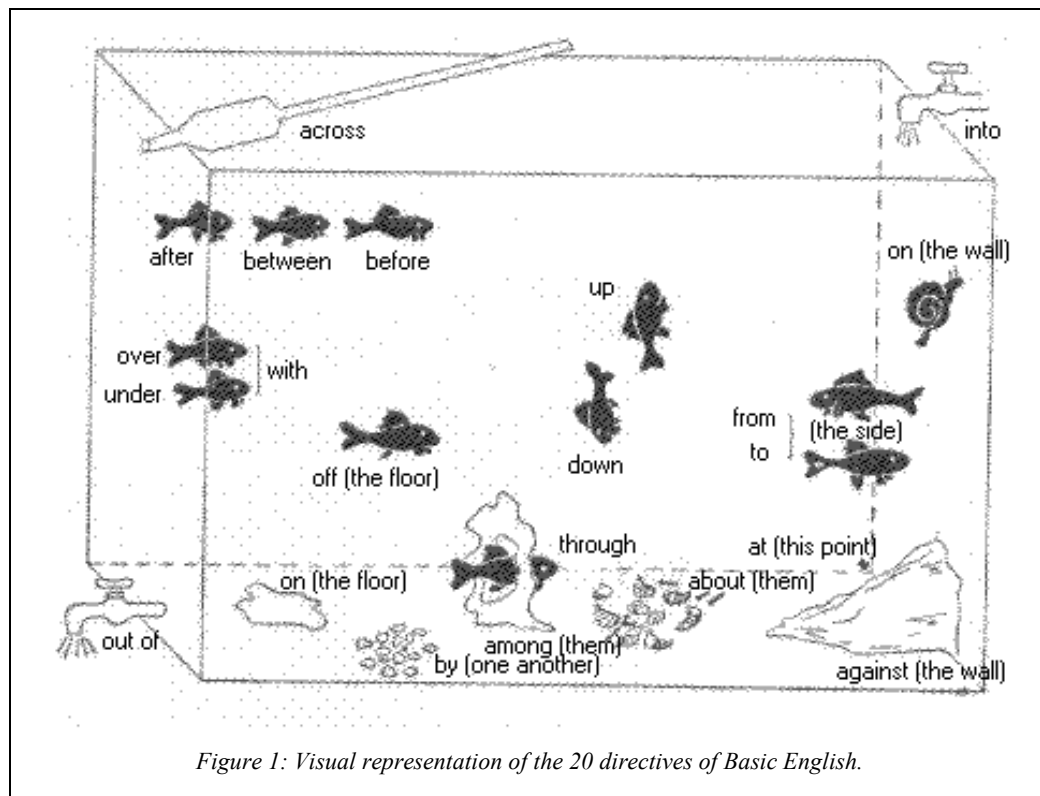
Figure 1: Visual representation of the 20 directives of Basic English.

---

[54] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[55] **Ogden, C. K.** *"Basic for Science,"* London: Kegan Paul, Trench, Trubner & Co. Ltd., Broadway House, Carter Lanes E.C., 1942, 514p.

The foreign learner can see with his/her eyes and his mind at once what these little, easily confused words mean in their essential senses.[56] This is evidence for the clarity, unambiguity, and usefulness of a visual language, such as the one we will propose. If it is easier for a language to be learned by looking at pictures (see section 2.2.4 on Visual Aids in Language Learning), then why not have a visual communication language? Visual communication will be discussed in section 2.3.

Among the 600 names of things are many that at first sight may be taken to be verbs (*act, attack, attempt, change, fall, etc.*). In Basic English they are nouns. The operators allow such words to be used in phrases that make a verb use of them unnecessary (e.g. not *attempt* something but *make* an attempt).

In writing for those trained in science, the wordlist is increased by another 100 words covering the general language of science, and 50 more for the needs of any special branch. In addition, if any other new words are needed, they may be made clear by an account in Basic English words, or by a picture.

In addition there are the (50) international words that Ogden regards as so common that they are not included in the Basic English word list. Among them are *bar*, *piano*, *restaurant*, and *telephone*, the names of and some of the terms in the chief sciences, and titles such as *president*.

2.2.3 Arguments against Basic English

The foregoing discussion might have given an ideal picture of Basic English, however, there are several reasons of why Basic English is not ideal as a universal language.

1. It is easier to learn for people that have some related language as their mother tongue. This is considered an injustice to speakers of languages remote from English.
2. Its dependence upon verb phrases may confuse rather than help the foreigner, whose difficulties with prepositions are notorious. These phrases are often extremely idiomatic, and have to be learned as though they were compound words.[57]
3. Basic English lets in all the difficulties and complications of the English verb while claiming to eliminate verbs through the elimination of the simple, easy, natural present and past tenses. One cannot say, "he attempted the flight", but

---

[56] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[57] **Walsh, C.** *"The Verb System in Basic English,"* American Speech, 8:137-143, December 1933.

must say, "the flight will have been attempted" (attempt is a thing and can be extended by -ed).[58]

4. It produces pseudo-simplicity by printing compound verbs separately, for example:

get up in the morning,
get out of the window,
get over an illness,
get your hair *cut*.[59]

These are essentially different words. The *get* stem does not make them one word. They will not make learning easier, but rather confuse it.

5. Some may believe that having to substitute *move myself through water by motions of my arms and legs* for the verb *"to swim"* very awkward. In terms of encodability and decodability this would certainly score low.

6. It is still English. It is conceivable that there are a lot of people in the world who do not want to speak English, no matter how easy it is made, and who especially do not want to speak it in negotiation with those who know more English than they do. By this argument, artificial languages, which do not have base languages, would be superior because they are only impartial.[60]

7. The spelling and pronunciation have practically no relation to one another. It is impossible for the foreigner to guess in advance the pronunciation of an immense number of English words. Consider the diversity in pronunciation of the words in the next poem:

When the English tongue we speak,
Why is "break" not rhymed with "freak"?
Will you tell me why it's true,
We say "sew" but likewise "Jew"?
....
"Beard sounds not the same as "heard";
"Cord" is different from "word";
"Cow" is cow, but "low" is low,
"Shoe" is never rhymed with "foe."
And since "pay" is rhymed with "say,"
Why not "paid" with "said," I pray?[61]

---

[58] **Aiken, J. R.** *"Little English,"* Journal of Adult Education, 1:11-14, June 1936.

[59] **West**, **M. P., Swenson, E. et al.** *"Vocabulary Limitation,"* Journal of Education, London, 71:582-584, September 1939.

[60] **Johnson, J. E.** *"Excerpts,"* from Basic English, The Reference Shelf. Vol. 17, No. 1, The H. W. Wilson Company, N. Y., 1944, p. 58-68.

[61] Poem by **Lord Cromer**, in the Spectator, August, 9th, 1902.

2.2.4 Visual aids in Language learning

According to Ogden in 1930, Basic English can be learned easier and faster by the aid of motion pictures (back then they did not have television, let alone computers), because it relies on verbs of motion and on visual modes of explanation. The Basic English operators and the directives are visualizable in their key senses in the fullest measure, and these senses may be illustrated for the eye in countless ways. The eye is the most comprehensive and the subtlest of our organs.

The more work we can hand over to the eye, the better will the whole work be done and the more clearly will it be remembered. Our thoughts, we say, "take shape." We learn most easily by seeing. For example, we understand by *seeing the point*, things become *clear* to us, *reveal* themselves to us.[62]

Richards further argues that the proportion of the meanings of a language that can be visually presented is an enormously important factor in determining the ease with which it can be learned and retained.

Ogden has had in mind a pictorial guide to the Basic English words, presenting their root senses, expansions, and specializations systematically.

The actions in pictures take the place of references to any mother tongue. By and large pictures are universal. People in the remotest corners of the earth and living in cultures least akin to ours learn our pictorial dialects with surprising ease and become thereby accessible to what we find to say with our pictures.[63]

Visual languages are even more important when we bear in mind that something like one half of the world's population, or more, are either unable to read at all or unable to read our alphabet.

2.3 Visual Communication Languages

Visual communication is the process of communicating with pictorial symbols (predominately icons). Sentences constructed visually could be an excellent form to convey messages, if they represent the deep structure of a sentence, thus bypassing

---

[62] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[63] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

the problems posed by its expression in natural language, such as ambiguity and diversity of form. Consider the diversity of form in the following sentences,

> Jim gave a book to Mary.
> A book was given to Mary by Jim.
> Mary was given a book by Jim.

They all have the same meaning, although the surface structures are different and they may differ in emphasis or focus. They would have the same representation if represented by pictographic symbols. This section, and the rest of the dissertation, will mainly focus on *iconic* communication.

In iconic communication people communicate by transferring pictures from one human mind into another. These pictures correspond, for human beings, naturally to types of things.

2.3.1 Why a Visual Interlingua?

One may think that it would be easier to have the whole world learn the same spoken language. This could be an artificial language or a natural language. Artificial languages like Esperanto have met little success, largely due to the initial effort required to learn them (by a significant number of people). As discussed in section 1.4, a natural language, such as English, would not work, because it would never be accepted by some countries, who might worry about cultural domination. For example, in France they are limiting the amount of English on television, because they do not want people to learn too much English or American culture. Besides this political problem, there are some other disadvantages of learning a spoken language to write in.

1. They have a linear structure, which can cause ambiguity (you can not see the parse tree).

2. They are designed for speaking and hearing, not to be perceived by the eye. This implies that they do not have the multidimensional capability a visual language would have.

3. Even in one spoken/written language there are a variety of ways people use the language. Even in one language there are inconsistencies (e.g. dialects).

4. Sentences with the same deep structure can be expressed with quite a few different surface structures. Consequently, there are a lot of ways to say the same thing.

An artificial language that is based on deep structure, would express every element exactly in one way.

Furthermore, easier and faster learning is achieved through visual recognition (see also section 2.2.4 on Visual aids in language learning). We believe that it is easier to recognize visual representations than to come up with words without system assistance, like when one learns a new (spoken) language. Keep in mind that whenever people do not understand each other, they try to make themselves clear by use of mime and gestures. The user only has to recognize the icons to construct sentences. Icons are also easier to remember. It is a well-known aid to associate a thing to be remembered with a simple object to which it has a defined relationship. One drawback with visual communication is the difficulty of producing visual symbols representing abstract meaning. These symbols are usually highly conventional and have to be learned. For symbols representing concrete objects, this poses fewer problems.

Figure 2: Comparison of natural language vs. Iconic Signs (Feierbaugh, 1993).

Figure 2 shows several iconic signs that are becoming fairly standardized in international usage, along with an approximation of their meaning in English. Note how the signs convey information in a culture- and language-free mode. Note also how the eye is instinctively drawn to the iconic symbols and only later browses back

37

to the natural language interpretation. The iconic symbols often relate at a "primitive" level to one or more of our other senses such as hearing, touch, and smell which also requires low-level cognitive processing. Meaningful alphanumeric symbols, on the other hand, are "compound" symbols (words and sentences) comprised of strings of more elementary symbols (letters) that require a higher level of cognitive processing to extract their meaning.

If iconic communication could be shown to be more efficient, quicker, easier, and less ambiguous than conventional communication means, like speaking or writing, then it might be a solution to the problem of multi-lingual communication (although it does not have to be all of these to be useful).

In addition, powerful computers and advanced user interface techniques (e.g. animation and interactive response to the user) provide an excellent opportunity for a computer-based iconic communication language. Another reason why now is the time for a computer-based universal language is the explosive growth in people using the Internet. People from any country now have access to the Internet. They send each other e-mail, read news, or use the World Wide Web (WWW). Many more people would have access (mentally) to the Internet and be able to communicate if they had a common language. This is a perfect opportunity for an iconic language.

2.3.2 Background on Semiotics and Icons

"Every blink of the eye brings a picture to the human mind." [64]

Images can be recognized quickly and committed to memory with surprising persistence. Many people "never forget a face". According to Mullet and Sano , images are particularly important in two areas:[65]

1. **Identification**: when serving as representations of concrete, real-world objects, images make identification easy. We learn the names of things at an early age, but we must first learn to recognize the image of the named object.
2. **Communication**: pictorial representations cross social and linguistic boundaries with ease when the objects that are being represented are relatively constant across cultures.

---

[64] **Frutiger, A.** *"Signs & Symbols: Their Design and Meaning,"* New York: Van Nostrand Reinholt, 1989.

[65] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

Representation provides the basis for all communication. We can convey ideas about things that are not materially in our presence only by calling forth an appropriate mental representation. The manner in which such representations are interpreted by participants in a communication system to create shared meaning can be fully understood only within the interdisciplinary context of semiotics. Semiotics is the study of signs and symbols.

The process of representation depends on establishing a clear relationship between a representamen and its object. Peirce identifies three forms this relationship can take:
[66]

1. **Icon**: An icon denotes its object by virtue of its own likeness or resemblance of that object, on the basis of some quality or characteristic inherent in the icon itself.
2. **Index**: An index refers to its object indirectly, by means of an association based on contiguity rather than on resemblance.
3. **Symbol**: Either an icon or an index may, over time, develop into a symbol, which denotes its object by convention alone and which thus depends on agreement between the parties in communication.
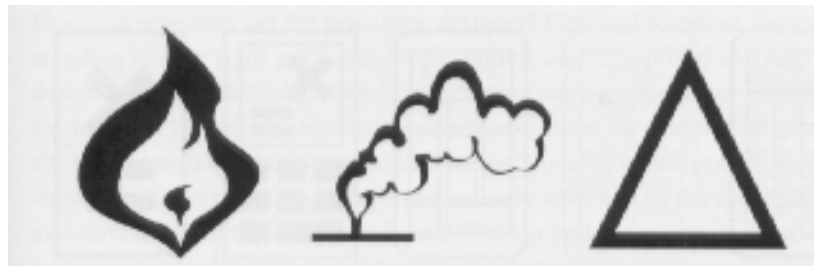
Figure 3: The concept of "fire" can be represented visually as an <u>icon</u> (a) through visual resemblance to flames, as an <u>index</u> (b) through visual suggestion of smoke (sound, temperature, or smell would work as well, in this case), or as a <u>symbol</u> (c) such as the Medieval alchemists conventional notation (Mullet & Sano, 1995).

Pure icons rely initially on recall of a previous visual experience on the part of the user, with enough distinction to make their use in a particular context clear. They may be used in conjunction with conventionalized symbols, e.g. a diagonal bar through an icon could indicate negation.[67] In common usage, the term *icon*, has come to denote any small raster image appearing in a graphical user interface (GUI) display.

---

[66] **Peirce, C.** *"Collected Papers,"* Cambridge, MA: Harvard University Press, 1931 (excerpted in Buckler, Justus, ed., Philosophical Writings of Peirce, New York: Dover, 1955).

[67] **Mealing, S** and **Yazdani, M.** *"A Computer-based Iconic Language,"* Intelligent Tutoring Media, 1(3), pp. 133-136, 1992.

"Iconic communication deals mainly with non-verbal communication between human beings by the use of visual signs and representations (such as pictures) that stand for an idea by virtue of resemblance or analogy to it in contrast to symbolic communications where the meaning of a symbol is entirely nominal (such as English text describing a picture)."[68]

In the next sections we will discuss some systems that are based on visual representations. Some of the systems are iconic and some of them are graphical, i.e. non-iconic. The earlier attempts at iconic communication were systems in which the icons were drawn on paper. The computer provides an ideal device for the implementation of a flexible iconic communication system, since the icons do not have to be drawn, but merely recognized. Also, icons can be manipulated easier (moved around, added, removed, etc.). Computer-based iconic systems are of the most interest to us and they will be addressed in the section 2.3.4.

2.3.3 Visual Communication Systems (Pictorial and Iconic)

Even in visual communication systems, there is a division into natural and artificial languages. The Chinese writing system was originally pictorial. However, the simplification of pictorial symbols by scribes to facilitate writing tended to turn them back into arbitrary symbols.[69] Furthermore, both the old Egyptians (hieroglyphs) and the North American Indian also communicated through pictures. In addition, there are existing signs, symbols, and icons that are understood internationally, but were not intended as a language. These are for example arrows to indicate direction and overlaid diagonals to indicate negation, most of which are found in traffic signs. A study of these can be useful in the design of our icons.

In the past, a number of iconic languages have been proposed; some examples are Isotype, Semantography (Blissymbolics), and Worldsign, a language created for mentally handicapped children.

The most recent attempts in iconic communication are computer based. Examples of this approach are the Hotel Booking System, which allows communication on a restricted domain (that of hotel booking), CD-Icon, which is based on Schank's conceptual dependency theory and designed as a pure person-to-person communication system, and the Elephant's memory, a computer iconic environment

---

[68] **Huggins, W. H.** and **Entwisle, D. R.** *"Iconic Communication: an annotated biography,"* The John Hopkins University Press, 1974.

[69] **King, A. J.** *"On the possibility and impossibility of a universal Iconic Communication System,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

which allows the user to build a visual message by combining symbols from a vocabulary of approximately one hundred and fifty separate graphic elements.

2.3.3.1 Division of Hieroglyphs and Traffic Signs into three groups

Liu Shu is a fundamental classification of Chinese characters into six groups. Kummel[70] tried to group the Egyptian hieroglyphs according to these groups. He also used these groups to divide up traffic signs. I will only mention the first three groups, because these are the ones that occur most clearly both in hieroglyphs and in traffic signs. The three groups are:

Group1: Xiang Xing, or Group of Concreta Signs. Signs that fall into this group express concrete meanings, one expression unit or symbol for one meaning/concept. Examples are rain, deer, field, grain, and gate [Kummel, p. 47].

雨 rain, 鹿 deer, 田 field, 禾 grain, 門 gate.

Group2: Zhi Shi, or Group of Abstracta Signs. Signs that fall into this group express abstract meanings, one expression unit for one meaning/concept. Examples are origin, end, the upper, the lower, middle [Kummel, p. 47].

本 origin, 末 end, 上 the upper, 下 the lower, 中 middle

Group3: Hui Yi, or Group of Composite Signs. Signs that fall into this group are exclusively composed of pictorial symbols of groups 1 and 2. Examples are grove, forest, gathering, public, and rule [Kummel, p. 48]. As one can see, the sign for forest is composed of three trees.

林 grove, 森 forest, 集 gathering, 衆 public, 法 rule.

2.3.3.1.1 Hieroglyphs

According to Kummel, the ancient expression system of hieroglyphs can be completely analyzed by the Chinese Liu Shu classification rules.

---

[70]**Kummel, P.** *"Formalization of Natural Language,"* Springer-Verlag.

There are concreta signs [Kummel, p. 53]:
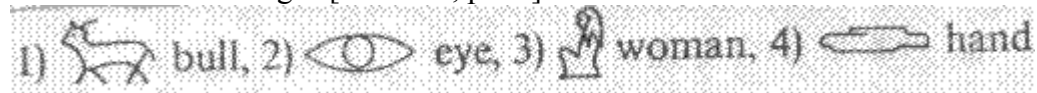
In group two, there are abstracta signs [Kummel, p. 53]:
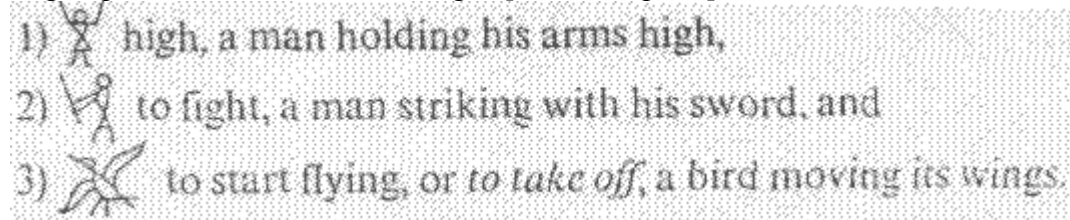
Composite signs according to group three of the Liu Shu classification are [Kummel, p. 53]:
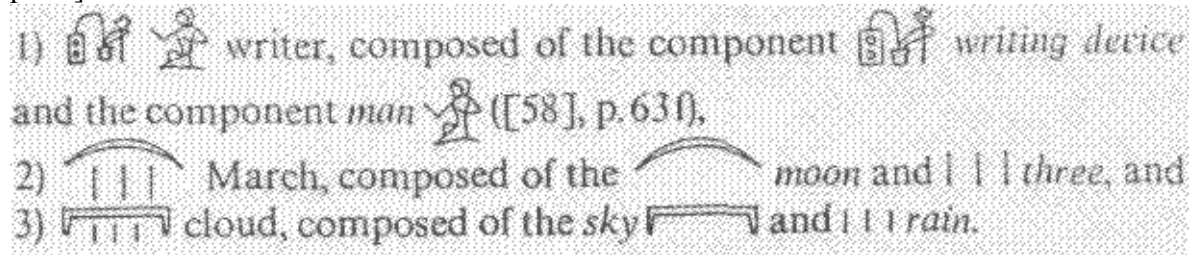
Later the hieroglyphs, just like Chinese, lost their self-explanatory features, because of simplification of pictorial symbols by scribes to facilitate writing.

2.3.3.1.2 Traffic Signs

Road Traffic Signs

Traffic symbols were created for ease of reading at higher traffic speeds. Another reason was the aim to address foreign drivers of different native languages. The frames of road traffic signs offer distinct shapes, which convey distinguished meanings. For example, the triangular frame of many signs implies the meaning *attention* or *caution*. A circular frame bears negative meaning, indicating a restriction.[71]

1) Road Traffic Signs Expressing Concreta

---

[71]**Kummel, P.** *"Formalization of Natural Language,"* Springer-Verlag.

Consider the following signs for road traffic with red frames omitted. They illustrate the meanings on the right:

 double curve right and left

 road narrows to the center

 road narrows to the right

 crossing
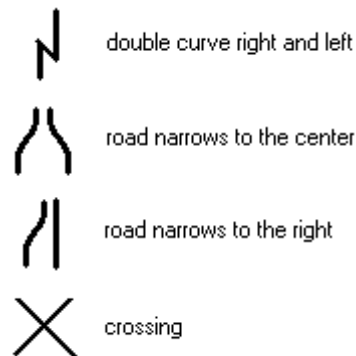
2) Road Traffic Signs Expressing Abstracta

The motion of vehicles and means of transport is expressed by abstracta signs, which incorporate time, the fourth dimension.

Examples:

 1) bicycle with no prohibitive red frame, but on a blue ground, indicating bicycle riding permitted

 2) motor cycling prohibited (indicated by the red peripheral circle)

 3) passenger cars prohibited (indicated by the red peripheral circle)

 4) caution railway crossing

 5) caution low-flying aircraft
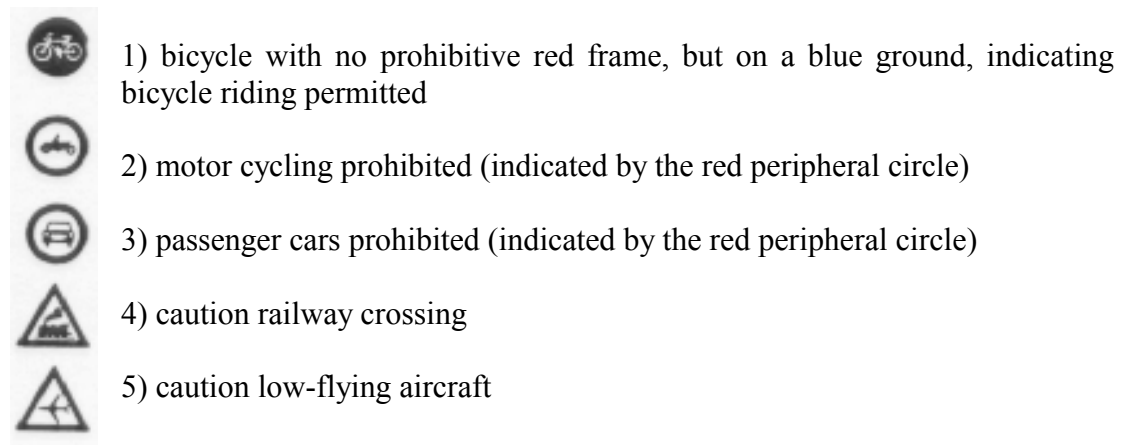
Abstract meanings are even more clearly indicated by signs representing moving animals:

 1) a horse rider on a blue ground, meaning: horse riding permitted

 2) a running and jumping deer, and

 3) a slowly walking cow, each meaning that the specific animal is a hazard
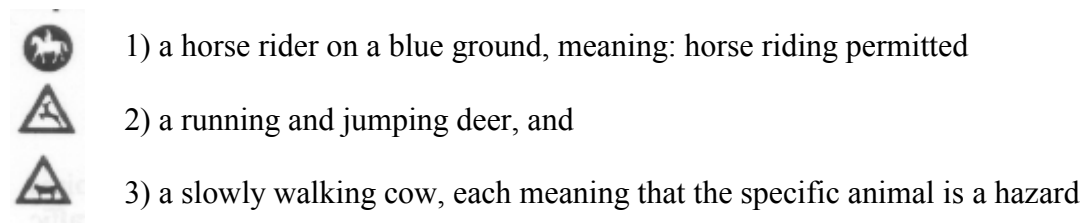
3) Composite Road Traffic Signs

A high percentage of internationally used traffic signs is represented by composite signs. At least two or even more components must exist, e.g. two children, worker and shovel, mountainside and falling rocks, automobile and tire traces, bankside and falling automobile:

1) caution: children crossing

2) caution: road work

3) caution: falling rocks

4) caution: slippery road

5) caution: river bank or quayside

Pedestrian Traffic Signs

A more recent development is pedestrian traffic signs, e.g. in the cities hosting Olympic Games or at international airports. These have square peripheries for each expression unit and are easily recognized. The motivation for their use primarily relates to addressing individuals of international origin, e.g. during festivals, congresses, and at international traffic junctions.

1) Concreta Signs for Pedestrian Traffic

Examples:
1) taxi, 2) bus, 3) telephone receiver, 4) movie camera and, 5) suitcase
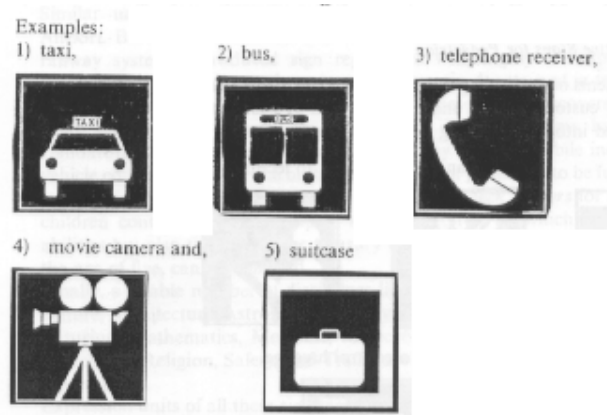
Some of these five concrete objects suggest motion. Still the expressed meanings of the five pictograms are to be understood as concrete meanings.

2) Abstracta Signs for Pedestrian Traffic

1) pass to the left, 2) pass to the right 3) escape from fire,

4) wash hands, 5) use escalators upward.
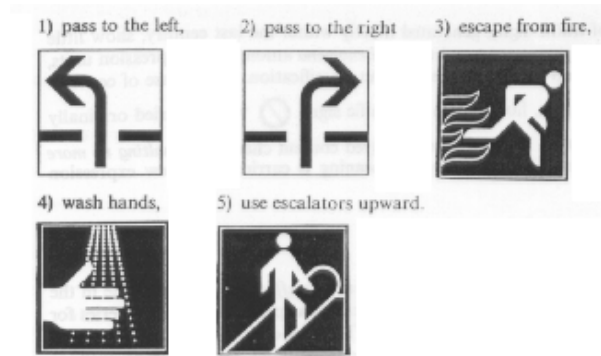
These all incorporate time as the fourth dimension. Kummel states that if the fourth dimension participates in the form of arrows for moving directions or dotted lines for displaced or removed parameters, an additional component is included in the sign morphology. Thus abstracta signs are fundamentally of a composite structure.

3) Composite Signs for Pedestrian Traffic

The components of 1) crawling swimmer and water, 2) dancing couple, 3) person and lecture desk, 4) uniformed person and passport, and 5) uniformed person and luggage are integrated into the following pictographic expression units:

1) to swim, 2) to dance, 3) to lecture,

4) to control passports and, 5) to control luggage.

Distinctive morphologies within a sign mainly serve the purpose of clearly differentiating one unit from another. Kummel states that for each meaning a self-explanatory structure could be conceived and created, thus eliminating expression ambiguities. These expression units would have the advantage of being perceivable and understandable by percipients of different particular language domains.

A sequence of 24 pedestrian traffic signs reveals information, which, Kummel claims, can be understood internationally with little or no ambiguity (see Figure 4). The signs have been selected from the pedestrian traffic signs at Rhein/Main Airport. The signs, however, are not all self-explanatory. What is, for example, the difference between the first and the fourth icon? Is the icon with the post horn, meaning post office, clear in all cultures? Furthermore, a story depicted in this manner uses sequencing. The icons need to be read from left to right and top to bottom. A truly universal system would be order independent, since not all cultures read in this manner (e.g. Chinese).

Figure 4: 24 pedestrian traffic signs

The content of the little story carried by the 24 pictograms might be the following (from Kummel, p. 97):

 "A traveler wants to take a trip by plane. Using car and train, he reaches the airport, where he checks in and puts the luggage on the scales. After waiting a little, he takes off, proceeds by flying and lands at his destination. Custom officials make passport and luggage checks. The suitcase is put into a locker, while the traveler passes a door and turns left to mail a letter. When the envelope has been put into the mail box, the

traveler proceeds via an escalator to a restaurant where he eats a snack and drinks a beer." [72]

As it is widely considered easier to conceive self-explanatory icons of concrete meanings by simply presenting a picture of that concrete content, abstract meanings have to be expressed more carefully. One fundamental rule is to incorporate the fourth dimension, time, into the expression structure. This can be done by integration of movements into the icon. In computer-based icons the explanation of abstract meanings could also further be aided by the use of simple animation.

The advertising pictograms in Figure 5 (left) feature three expression units carrying concrete contents and one, the second one, carrying the abstract meaning *swimming*. Kummel states that if all four pictograms had been conceived in logical relation, then number two indicating swimming possibilities should also be expressed by the pictograph of a swimming pool.

Figure 5: Commercial pictograms of abstract meanings in conventional implementations (left) and with improved structures (right).
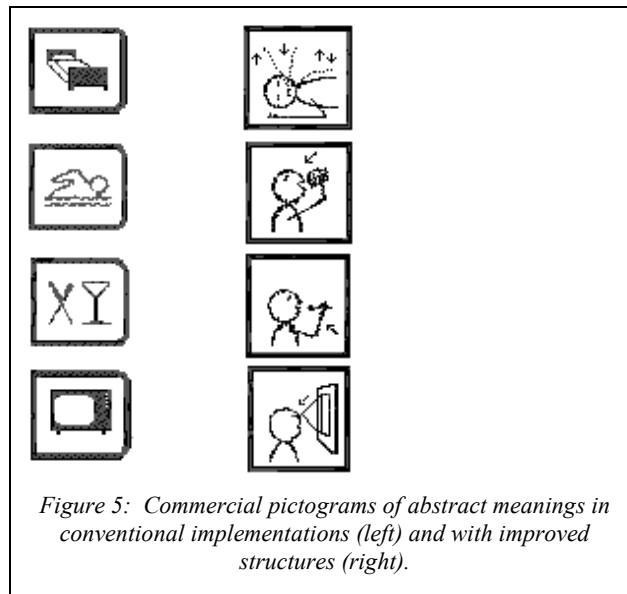
The symbols on the left could be used to express the concrete meanings, the furniture of a bed, a person swimming, a knife and fork set, a drinking glass, and a television set. The symbols on the left could presumably also be used to express the following abstract meanings:

1. sleeping or sleeping facility,

---

[72]**Kummel, P.** *"Formalization of Natural Language,"* Springer-Verlag.

2. swimming or swimming facility,
3. eating and drinking or eating and drinking facilities, and
4. television viewing or television facility,

Kummel states that in order to express appropriately, without any ambiguity, the abstract meanings to sleep, to eat, to drink, and to watch television, the pictograms on the right should be considered instead of the ones on the left. Otherwise ambiguous expressions of a furniture store could be perceived. Consequently the abstract signs on the right contain:

1. inhalation and exhalation movements of air from and to the mouth of a sleeping human.
2. arm movements with a cup in the hand
3. arm movements with a fork in the hand, and
4. light ray movements from the television screen to the watcher's eye.

Kummel concludes that expression morphologies of pictograms, or fully content-related graphemes carrying abstract meanings, must contain motion components related directly to the abstract content. This is not necessarily true, because concepts from different categories or grammatical entities could have the same icon to represent a different meaning, if the icons were marked in some way. For example, the icon of a bed would have the meaning of "bed" if we were talking about nouns, but would have the meaning of "to sleep" if we were talking about verbs. The difference between the icons could then be made clear by having different borders around the icon or different background colors for nouns and verbs (or concrete and abstract meanings).

2.3.3.1.3 Using images for abstract concepts

Images of familiar, concrete objects are easy to interpret correctly. Abstract concepts, processes, or situations, because they are less tangible, depend on less direct forms of representation in which even a well-designed image may be difficult to interpret correctly. Even when the sign's syntax is clear, its semantics can be obscured by the weak, largely conventional association between the visual elements of the sign and the abstract or temporal aspects of the situation signified. Images for abstract meanings are purely conventional - they must be learned before they can become useful. However, an initial struggle to determine the intended meaning of an icon may be a small price to pay if it is then remembered accurately.

2.3.3.2 American Indian Pictorial Language

Pictographic texts feature sequences of pictures, which are highly self-explanatory. Without any incision marking divisions or units like sentences, the information is expressed by fully isolated morphologies. Almost each pictogram of abstract content can be considered a minimum sentence. Single expression units or graphemes carrying concrete contents are not capable of forming a sentence. They have to be concatenated with abstract units. [73]

Since the pictures are drawn on a single surface, such as a cave wall, the whole message is visible at any moment and therefore any part of it is open to inspection in the context of the rest. This is very different from gesture language, which is serial in nature, and only a small part of the message is being communicated at one time. Attempts to look for context rely on memory.[74] Sequencing is used though, because the story is read by following the spiral, starting from the center.

Figure 6 shows American Indian pictographic text on leather in a spiral shape.
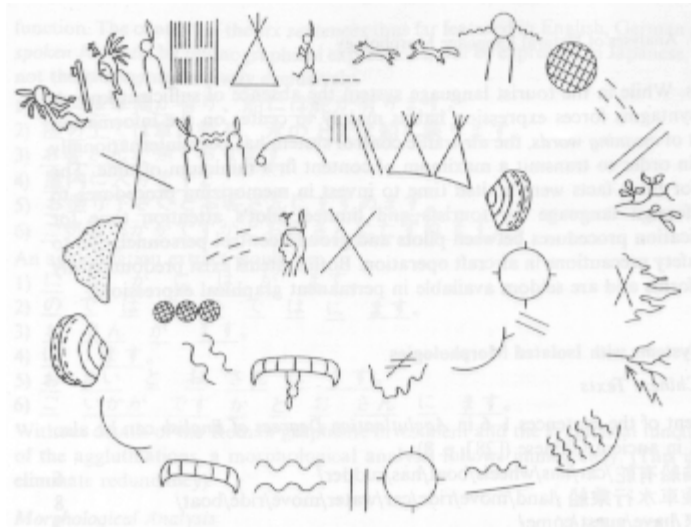
Figure 6: American Indian text on leather.

---

[73]**Kummel, P.** *"Formalization of Natural Language,"* Springer-Verlag.

[74] **Yazdani, M.** and **Goring, D.** *"Iconic Communication,"* Internal Report, Department of Computer Science, University of Exeter, England, 1990.

The interpretation of the content was given by Tomkins[75]:

"An Indian trader by the name *Little Crow* went on a journey. He traveled for three nights until he came to a river. The reason he traveled at night was because he was in enemy country. At the river, he secured a canoe, camped there that evening, and at sunrise the next morning started down the river and traveled two suns (days). He now traveled in daytime, because he was in friendly territory. He was an Indian trader in shells, which were used for wampum and ornamentation. At the end of the fifth day's travel he reached the village where shells where obtainable. He rested there for three days in conference with the chief and as a result he traded for a large amount of shells. At sunrise on the fourth day he loaded his canoe and started down the river and traveled for two days. On the second day a storm came up, with rain and lightning. He saw the lightning strike a tree and set it afire. As a result of the storm he became sick, so he searched and found some medicinal plants and waited there a couple of days until he felt better. He then traveled at night and hid away in the daytime. Hr knew that the country abounded in game because he heard foxes and wolves. He finally reached home, though some days late. Twenty braves of the tribe came out to meet him, including their chief *Standing Bear*. Their hearts were glad as a result of his safe and successful trip, and they all had a very social time."

2.3.3.3 IICS: the Iconic Interactive Communication System

IICS is an iconic system based on the pictorial language used by the North American Indians. North American Indian pictography may be described as "elemental, basic, logical, and largely idiomatic."[76] There are about 100 Indian signs in general use, and about 50 of them were implemented in IICS. American pictographs were written on skins, pottery, and cave walls. The following four pictures [Yazdani 1990] represent sentences in the iconic form of the North American Indian pictography.
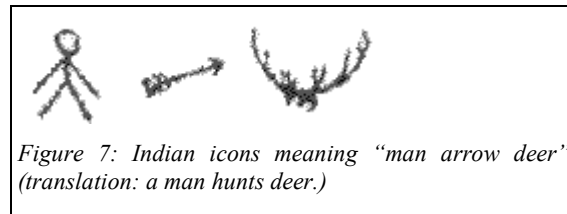
Figure 7: Indian icons meaning "man arrow deer" (translation: a man hunts deer.)

---

[75] **Tomkins, W.** *"Indian Sign Language,"* Dover Publishers, ISBN 048622029X, 1969.

[76] **Yazdani, M.** and **Goring, D.** *"Iconic Communication,"* Internal Report, Department of Computer Science, University of Exeter, England, 1990.
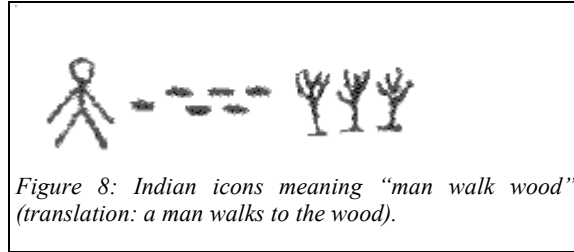
Figure 8: Indian icons meaning "man walk wood" (translation: a man walks to the wood).
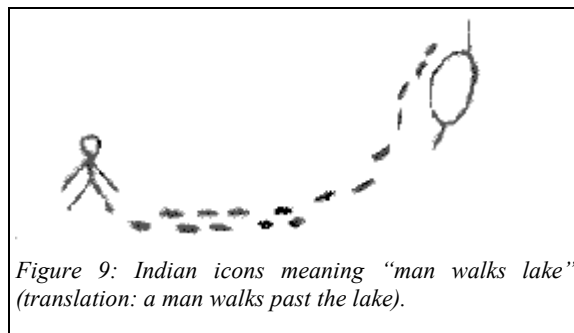
Figure 9: Indian icons meaning "man walks lake" (translation: a man walks past the lake).
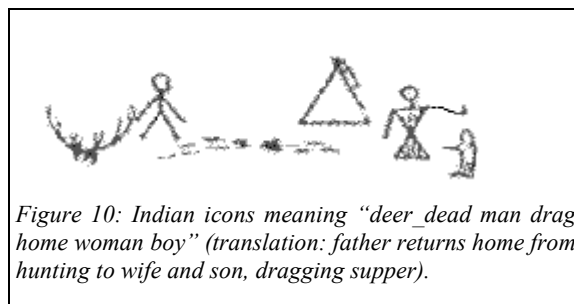
Figure 10: Indian icons meaning "deer_dead man drag home woman boy" (translation: father returns home from hunting to wife and son, dragging supper).

Yazdani found that in developing the system, structural/syntactical considerations were important. He claims that the sequence of icons for a message needs to follow the standard word ordering of the particular natural language used, and consequently the syntax of the icons must relate to the syntax of the users written language. So if you have an English person communicating with a Spanish person, then the system needs to put any message that the English user made, in the Spanish style before presenting it to the Spanish reader, and vice versa. This is, however, highly undesirable. It means that the system has a different look and feel for every language it supports and also has to have knowledge of all these languages. A truly universal system should be language independent and have the same look and feel for all

languages. It will have its own grammar and syntax, and have no linear word order, i.e. icons could be placed in any position on the screen.

2.3.3.4 Before Computers Mediated: Isotype

Isotype stands for **I**nternational **S**ystem **O**f **TY**pographic **P**icture **E**ducation and was designed by O. Neurath in 1978 at the University of Reading. Neurath tried to avoid the ambiguity and linearity of text. The system of connection between signs is far richer than in linear text. He tried to exploit the two dimensions of the page. "Several rows of connected signs are interpreted simultaneously, whereas one-dimensional text readers need to bear in mind what they have read and to make connections between dispersed elements for themselves."[77]

In Isotype there is no simple correspondence between signs and words. There is for example no sign for the word *foot* that is common to expressions such as *the foot of a man, the foot of a mountain,* and *the foot of a table*.[78] These expressions are composed of simple signs of a very different sort. Neurath suggests two rules for generating the vocabulary of an international picture language:

1. *Reduction*   - for determining the style of individual signs.
2. *Consistency* - for giving a group of signs the appearance of a coherent system.

Neurath's work includes a series of posters for an anti-tuberculosis campaign and the publication of many books and charts.

Isotype is used for statistics and too restricted for a computer environment, but it could be seen as a precursor of an iconic language.[79]

2.3.3.5 Before Computers Mediated: Semantography

Semantography (Blissymbolics) was developed in 1965 by C. K. Bliss[80]. In Semantography, one symbol corresponds to one word in natural language. It was

---

[77] **Beardon, C., Dormann, C., Mealing, S.** and **Yazdani, M.** *"Talking with pictures: exploring the possibilities of iconic communication,"* Association for Learning Technology Journal, 1, 1, pp. 26-32, 1993.

[78] **Neurath, O.** *"International Picture Language,"* Department of Typography and Graphic Communication, University of Reading, England, 1978.

[79] **Beardon, C., Dormann, C., Mealing, S.** and **Yazdani, M.** *"Talking with pictures: exploring the possibilities of iconic communication,"* Association for Learning Technology Journal, 1, 1, pp. 26-32, 1993.

[80] **Bliss, C. K.** *"Semantography,"* Semantography publications, Australia, 1965.

proposed as an auxiliary writing tool for communication purposes between different nations. Communication is through simple pictorial symbols. There are two kinds of symbols, symbols representing physical things using outline, and symbols representing non-physical things using geometric symbols. The first 25 symbols are already internationally accepted, they include the digits 0 through 9 and symbols like: ? . + - etc.

Bliss based his grammar on the assumption that all languages are used to describe the phenomenon of our physical world, that the main manifestations of our world can be classified into matter, energy, and mental, and that everything happens in space and time. The atomic icons, representing basic objects in the world, can be combined to form complex icons that map on to the set of words found in natural languages. A word like *writer* takes its symbolization from symbols for *man* and *pen* (see Figure 11).
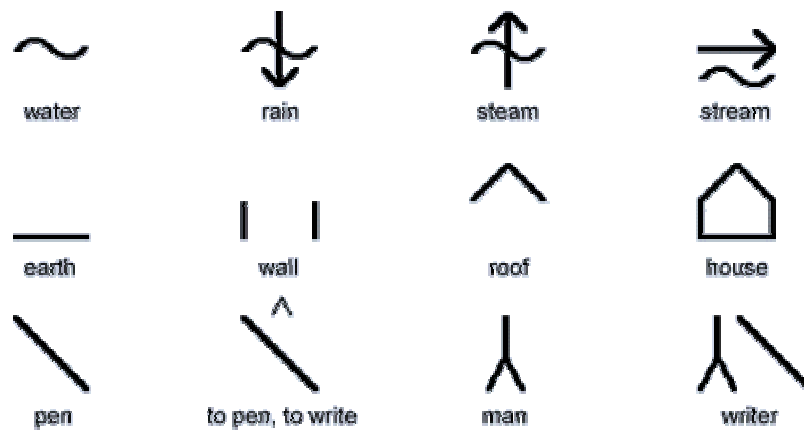
Figure 11: Examples of construction of words from primitive/basic elements.

The linear grammar of Blissymbolics relies on word order (SVO, i.e. Subject Verb Object) and is similar to the word order of European languages. Furthermore, the lexicon is not intuitive enough to make people want to learn it (try to look at the icons without the accompanying text). It takes a long time to learn. Furthermore, it was designed to be written on paper. Although there is a computer-based version of it available now, called Bliss for Windows, its basic vocabulary is limited and one still needs to draw icons by combining basic elements, instead of choosing icons from the screen, which would allow for faster encoding and recognition.

2.3.3.6 The Hotel Booking System

This system was developed by M. Yazdani at the University of Exeter, and is an initial attempt to create an interactive, iconic dialogue, using hotel booking as theme.[81] The system would allow a user to compose his complete booking requirement iconically and send the message to the hotel manager to reply to. The system is task domain dependent.

A typical scenario could find a tourist in a foreign city operating a touch screen in the window of a tourist office, or a traveler contacting a foreign town's accommodation from his terminal at home.

The compilation of the booking message is accomplished in stages. At each stage, the current domain is cued by a picture on the background of the screen (a typical hotel front, a typical hotel reception area, and a typical hotel bedroom). So, when dealing with a room, the background picture would be of a room.

The first thing the user has to do is to choose a destination (cued by a map) and desired star rating of the hotel. Then, in a new screen, the user has to specify the dates and times of arrival and departure, as well as the number of night he/she wants to stay. A third screen shows a room overlaid with icons to select the room type (number and type of occupants, number and type of beds, type of bathroom facilities, and other choices such as a television). The room screen is displayed in Figure 12.[82]

---

[81] **Mealing, S** and **Yazdani, M.** *"Communicating Through Pictures,"* from the World Wide Web, Department of Computer Science, University of Exeter, England, 1992.

[82] **Beardon, C., Dormann, C., Mealing, S.** and **Yazdani, M.** *"Talking with pictures: exploring the possibilities of iconic communication,"* Association for Learning Technology Journal, 1, 1, pp. 26-32, 1993.
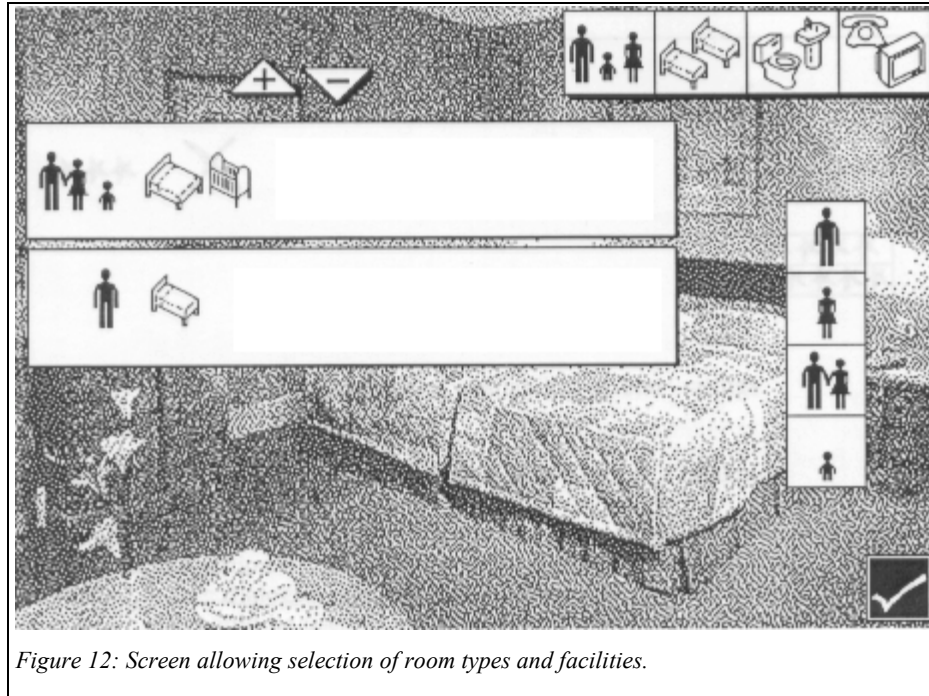
Figure 12: Screen allowing selection of room types and facilities.

When a satisfactory arrangement of rooms, occupants, and facilities has been chosen, the user clicks on the check icon. A screen is displayed that shows the complete booking requirement. If the user accepts these choices, the message is sent to the hotel. The message is revealed to the hotel in stages. Confirmation of the acceptability of each part of the message moves on to the next part of the message. If a requirement is unavailable, a range of alternatives is brought up. The final message is sent back to the customer, who will be able to accept or reject the alternatives offered.

2.3.4 CD-Icon: an Iconic Communication Language

CD-Icon was developed in 1992 by Colin Beardon at the University of Brighton. The purpose of the program is to explore unrestricted person-to-person communication. CD-Icon is an iconic language that is based on Schank's conceptual dependency theory.[83] This section assumes some knowledge about Schank's conceptual dependency theory. To better understand this section on CD-Icon, it is advisable to first read about conceptual dependency theory in section 3.2 which is about grammar.

---

[83] **Schank, R.** *"Computer Models of Thought and Language,"* W. H. Freeman, San Francisco, 1973.

CD-Icon is a means of testing the validity of conceptual dependency directly by using it as the basis of a communication system that uses only icons and no words.[84]

2.3.4.1 The System

A message is composed by selecting options from a series of interconnected screens. The message is then transmitted, also as a set of interconnected screens, without showing the options that were not selected.

Four basic types of construct can be specified from the structure of conceptual dependency diagrams:

1. A message, which can be a simple or compound conceptualization.
2. A conceptualization, which is based around a primitive action (ACT) and its related deep semantic cases.
3. A picture, which is based around a picture producer (e.g. Jim, dog, nacho).
4. A lexical entry (picture producer or picture aiders), which is an atomic icon and its semantic constraints.

The CD-Icon system reflects these divisions. It is constructed as four modules, each processing one type of construct with clearly defined interactions (see Figure 13).[85]

---

[84] **Beardon, C., Dormann, C., Mealing, S.** and **Yazdani, M.** *"Talking with pictures: exploring the possibilities of iconic communication,"* Association for Learning Technology Journal, 1, 1, pp. 26-32, 1993.

[85] **Beardon, C.** *"CD-Icon: an iconic language based on conceptual dependency,"* Intelligent Tutoring Media, Vol. 3, No. 4, November, 1992.

Figure 13: The four modules in CD-Icon.
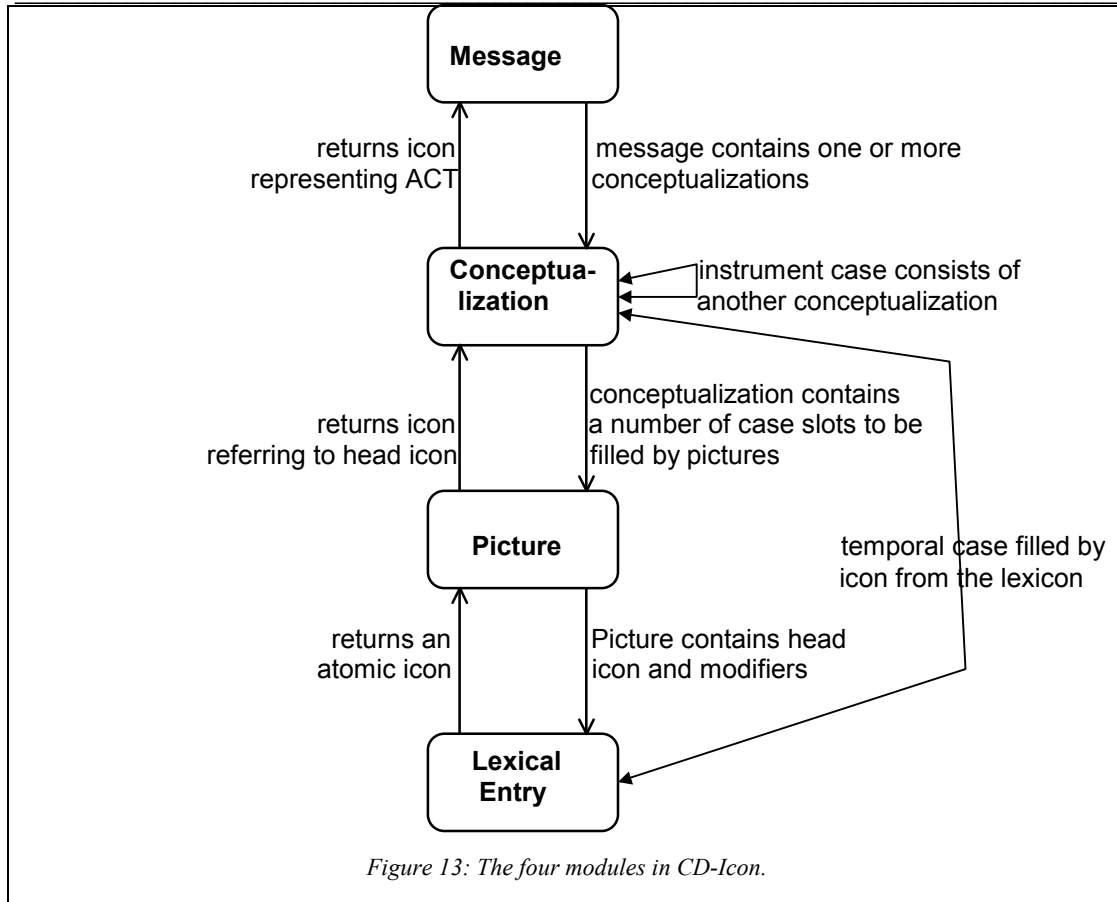
The system is entered at the message level, with each level calling the level below until the lexicon is reached, at which point objects begin to be passed back up the various levels. Let us look at an example, and compose the message "The big man went home". In a conceptual dependency diagram this message looks like:
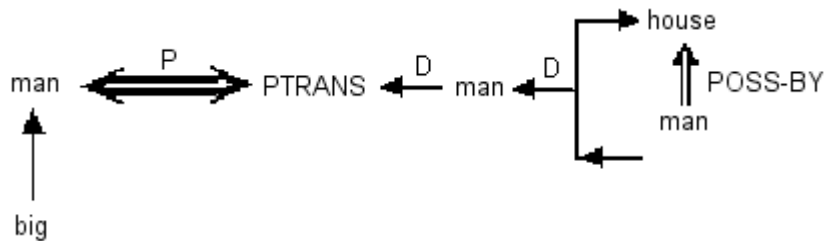
Level 1: The message
The user must first decide whether the message is simple or compound, select an assertion, a question, or an imperative, and decide upon negation. If the message is

58

compound, the nature of the relationship between the two component conceptualizations has to be chosen. This is either logical AND, logical OR, implication, temporal, or spatial. Clicking on any of the uninstantiated conceptualizations will bring you to the next stage.

Level 2: The conceptualization
In this stage it is decided upon what kind of primitive ACT is involved in each conceptualization. There are only four ACTs implemented: PTRANS (Physical TRANSfer), MOVE, GRASP, and ATRANS (Alienable possession TRANSfer), see Figure 14.
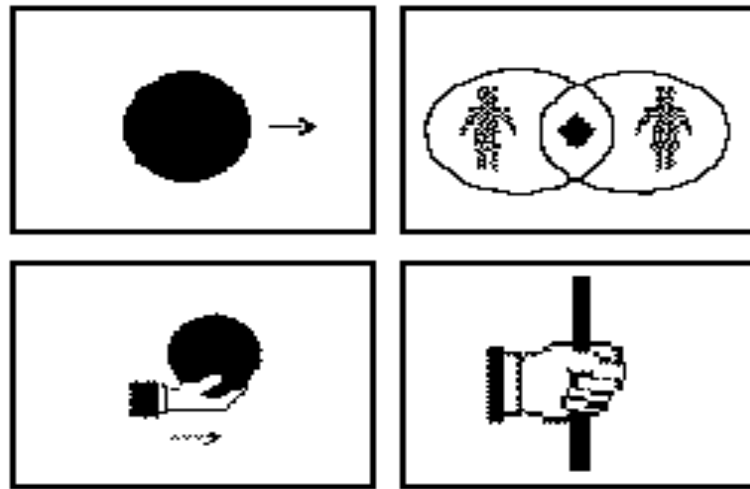
Figure 14: Selecting a primitive ACT (PTRANS, ATRANS, MOVE, GRASP)

Clicking on one of these menu icons using the option key, will result in an animated explanation of the meaning of that ACT. PTRANS, for example, shows an object physically moving from one place to another, whereas MOVE is essentially the same animation but with a person carrying the object. Once the appropriate ACT is selected, the corresponding screen is presented, in this example the PTRANS screen (Figure 15).
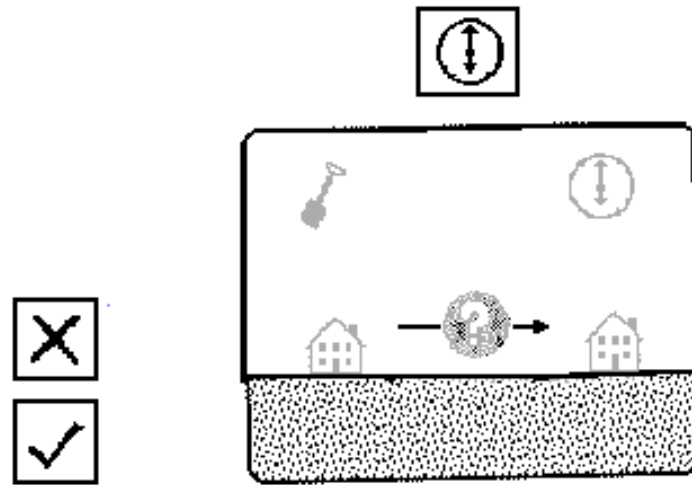
Figure 15: The screen for PTRANS

This screen contains a background for PTRANS with gray icons (uninstantiated) representing the object, origin, destination, and instrument case, as well as tense. A gray house represents the class of places, the gray question mark represents the class of objects, the gray clock represents the class of times, and the gray spade represents the class of instruments. By clicking on any of these icons (except spade and clock), control is passed to stage 3, the picture screen, from which an icon will be returned. Except for the first time, clicking on an icon will result in a menu being presented with two options: a new object (which will take you to the lexicon screen where the head object is selected after which control is passed back to the picture screen), or a known object (which will take you to the picture screen for that icon).

Temporal reference is represented by a clock at the top of the screen representing 'now'. Past, present and future are indicated by setting the clock to 3, 6, or 9 o'clock respectively.

Level 3: The picture
Pictures are made up of a head picture and various modifiers (compare with noun phrase). The type of the head picture is selected from the lexicon before the picture screen is displayed.

Depending on the type of object selected, a set of probable modifier types is also displayed (picture aiders and producers): for typical physical objects these are color, size, owner, location, and name (Figure 16).
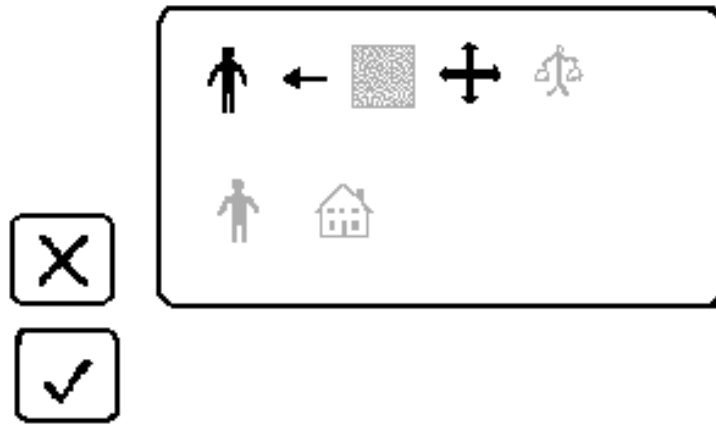
Figure 16: Picture screen for "big man"

Clicking on any of these modifiers will take you to the appropriate page in the lexicon.

Level 4: The lexicon

The lexicon contains icons that denote both picture producers (nouns) and picture aiders (adjectives). The lexicon appears as a series of pages, with each page representing a type, e.g. the type "predefined places" includes icons for room, house, office, shop, city, and countryside.

A completed message is shown in figure 17, and consists of a message screen, a conceptualization screen, and two picture screens. The user may interrogate any of the icons involved to find out more about them.
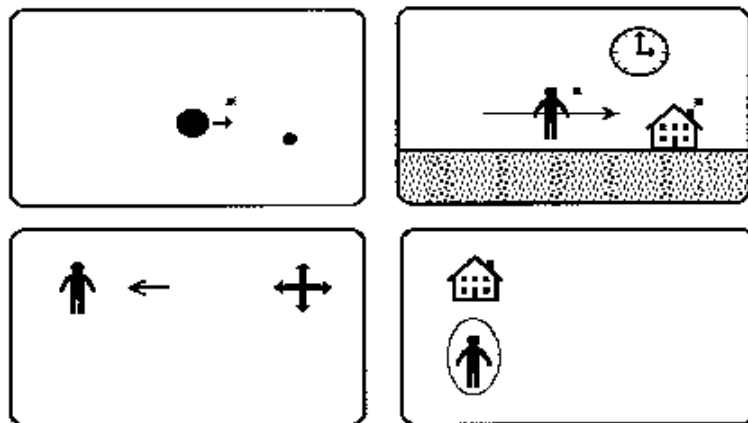
Figure 17: The four screens representing the complete message

There are, however, some issues to be discussed about the system. There are different screens for different ACTs. It is more consistent to have a general screen in which all verbs can be represented. Also, the conceptual dependency definition of instrument requires a recursive sentence modifier, which is undesirable. Furthermore, it uses screen location to indicate case. In figure 15 (the screen for PTRANS), the house on the left indicates source.

There are also some objections to using conceptual dependency as the basis for an iconic communication language. If knowledge is initially presented to the system in a relatively high-level form, such as English, then substantial work must be done to reduce the knowledge into primitive form. We believe that inferences that are made in our heads should not be explicitly represented, like in conceptual dependency theory. If you want to represent the verb "to kill", in conceptual dependency theory you would have to say "cause to die". We believe that it is more desirable to have more verbs than the 18 primitive ACTs. There is no need to decompose verbs this much. Finally, we believe that there would be no incentive to learn the system, because it would take too long to compose messages and take too long to learn.

2.3.5 The Elephant's Memory

The Elephant's Memory was developed by Timothy Ingen Housz[86]. It is a pictorial language consisting of around a hundred and a fifty combinable graphic elements, and is based on a two dimensional grammar. The system is primarily oriented towards children. The Elephant's Memory is designed as an experimental workshop. Figure 18 shows the Elephant's Memory's vocabulary and Figure 19 shows a sample sentence.

---

[86]**Housz, T.I.** "The Elephant's Memory," 1994-1996.
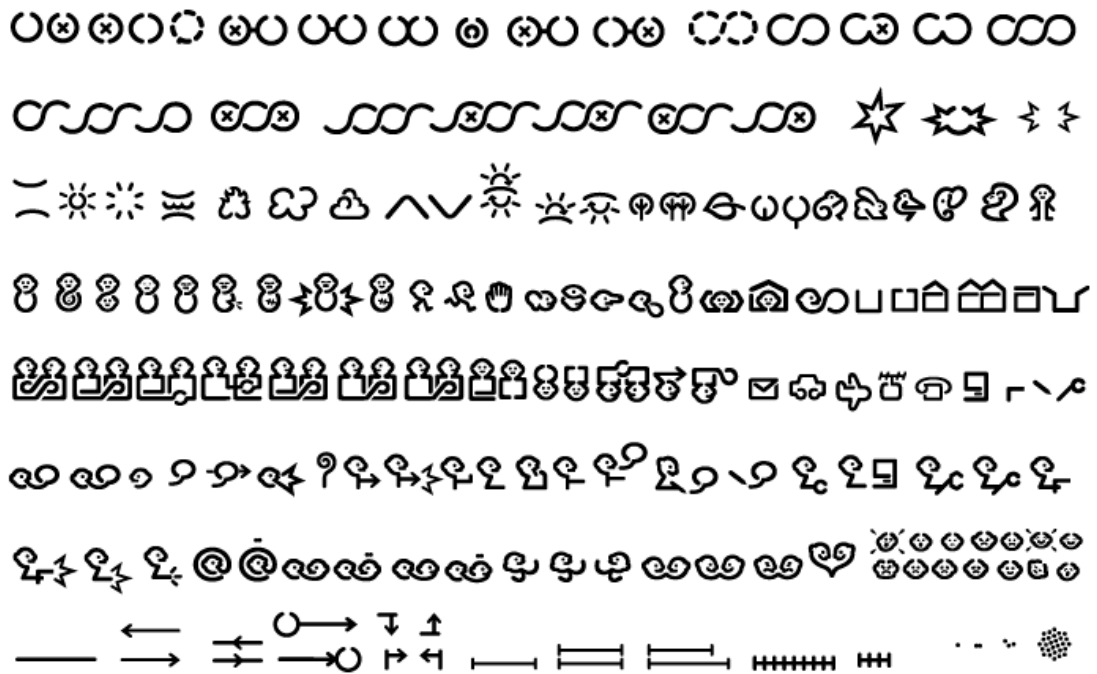
Figure 18: All icons in the Elephant's Memory vocabulary

The Elephant's Memory is more an educational tool than a communication language. Its vocabulary is too limited and its expansion would increase the learning time too much.
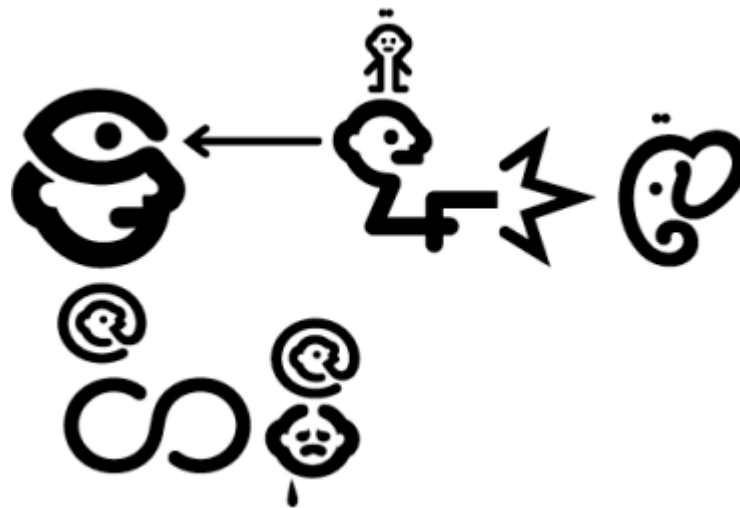
Figure 19: "Seeing elephants shot by men makes me cry."

In the next chapter VIL will be discussed. **VIL** stands for **V**isual **I**nter **L**anguage, and is also a computer-based iconic communication system. It is designed to allow people to communicate with each other when they share no common language. The goal is to make the system language independent so that it can be used universally. So, the system should be able to be used to construct sentences irrespective of the languages known by the users. With most iconic communication systems there is no immediate incentive to learn the language. The languages are either based on too complex theories (like Schank's conceptual dependency theory), or the symbols and icons are not intuitive and self-explanatory (conventional abstract icons should be avoided as much as possible). Furthermore, most systems are too complex to learn. VIL is designed with the emphasis on simplicity and ease of use. It allows for fast learning and flexibility. To help VIL attain this simplicity, many principles will be adopted from research on pidgins and the basic languages such as Basic English. We have designed a grammar for VIL, which reflects this simplicity.

3. Language Design Semantics

This chapter describes the design of the language VIL. To facilitate the reader's understanding of the topics, the review of previous research on a topic and the design decision are put in the same section. This allows the introduction of terms close to their use. Various principles for the language VIL are stated in section 3.1. A lot of these principles have been compiled from the information on pidgins and Basic English discussed in the previous chapters. In order to come to VIL's grammar, we have researched verbs, nouns, and adjectives. The first part of section 3.2 deals with previous work on grammar, the second part studies the organization of the iconicon and also describes previous research into classification. In section 3.2.2.3. noun classification is described and in section 3.2.2.3.1 VIL's organization of nouns is presented. In section 3.2.2.4. adjectives are studied and in section 3.2.2.4.2 VIL's adjective classification is presented. In the last sub-section of 3.2, section 3.2.2.5. verbs are researched and VIL's verb classification is presented in section 3.2.2.5.7. The organization of VIL's verb system was then used in composing VIL's grammar, which is described in section 3.3.

3.1 Principles of VIL

VIL's main principle is that it uses the greatest common denominator of a great number of languages in the design of its grammar. The grammar will then be drastically reduced in complexity. Research on pidgins and Basic English has shown that it is possible to communicate with a drastically reduced grammar and vocabulary.

Principle 3.1: *Greatest Common Denominator (Minimal Grammar & Vocabulary)*

> VIL will adopt a *minimal grammar*. Every grammatical feature which is encountered in most languages, shall be retained in the grammar of VIL, i.e. no grammatical feature shall be retained if it is missing from any major language.

Thus, for instance, the feature of grammatical gender can be dispensed with because it is missing in one of the major languages spoken by a large percentage of the people in the world, i.e. English.[87]

---

[87] **Gode, A.** and **Blair, H. E.** *"Interlingua: a grammar of the international language,"* Storm Publishers, New York, 1955.

In section 3.1.2 the different parts of normal speech and if and how they will be presented in VIL will be discussed. But first let us review some more principles of VIL

3.1.1 Principles

Almost all of the following principles of VIL have been compiled from the information on pidgins and Basic English in chapter 2. Just like Pidgins, VIL primarily has the following two functions:

Principle 3.2: ***Directive and Referential Function***

>VIL is a language to communicate about simple day-to-day events, i.e. facilitate mutual understanding on a limited range of topics (this might be in, for example, a tourist or trading context).

VIL is not intended as a language in which write poetry, neither was it designed as a scientific language. As mentioned, VIL adopts the principle of the greatest common denominator. The next table shows some of the simplifications VIL will adopt.

Simplified linguistic structure in VIL	
Lexicon	Smaller vocabulary, resulting in generic rather than specific icons
Syntax	No linear word order
Morphology	No inflections

We will now give a more detailed explanation of the simplifications mentioned in the table, presented as a number of principles.

Principle 3.3: ***No Linear Word Order***

>Word order differs greatly among the major languages in the world. To make a consistent language and interface, VIL has no linear word order. The icons can be placed in any order since the different cases (e.g. subject, object, indirect object, etc.) will be clearly visually distinguished.

Principle 3.4: ***Number***

>There are no synthetic plurals of the 'man/men' or 'computer/computers' type. Instead nouns are invariable, like 'sheep'. Plurality is normally implicit in the context and can be indicated by a *quantifier* or a *number*, e.g. 'all book' or

'five book'. Number is also missing from the Chinese language, which is spoken by a large percentage of people in the world.

Principle 3.5: *Gender*

Gender distinctions are eliminated in nouns, as is in the English language

Principle 3.6: *Inflection*

There is no concordial agreement between subject and predicate. Both noun and verb are invariable, where the singular is used for nouns and the imperative for verbs. Since the verb form is invariable, distinctions relating to time and continuity of action are either understood from the context or are indicated by adverbials. Adjectives are invariable as well. Inflection is missing in many languages.

Principle 3.7: *Negation*

Negation is done by having an invariant way to "modify" an icon, for example by a diagonal bar through the icon, and will not change any order in a message.

Principle 3.8: *Verbs and tense*

All contrasts of number, person, and tense have been lost in the verb, and every verb has only one form used in inflection. Tense is either implicit in the meaning of the sentence when indicated by *time-demonstratives* (last, this, next) or explicit when indicated by *time-pronouns* (past, present, yesterday, tomorrow, etc.).

Principle 3.9: *Auxiliary verbs*

As a consequence of Principle 3.8, no auxiliary verbs are used in tenses (as in "I *have* eaten").

Principle 3.10: *Article*

VIL does not have a specific article, i.e. there is not a specific icon for 'article'. The default article is "a". The definite article 'the' is used in normal speech to refer to something specific that both the speaker and hearer know about. When a speaker says "did you receive the book?", the hearer knows what book the speaker is referring to. In VIL, definiteness is represented by

the *nominal-demonstratives* "this" and "that" as in "this book". Article too is also missing from the Chinese language.

3.1.2 Parts of Speech

VIL's parts of speech are reflected in its grammar. The grammatical categories that VIL distinguishes are:

- time-when (when does the event take place)
- place-where (where does the event take place)
- instrument (what is used to carry out the action)
- verb
- auxiliary modal: should, must, can
- aktionsarten: start, stop, and continue (indicating continuity of action)
- adverb
- frequency of occurrence (how often does the event take place)
- duration (how long does the event take place)
- intensifier: very, most, least
- noun
- nominal pronoun: I, you, he/she/it, we, you all, they
- nominal demonstrative: this, that, which
- adjective
- quantifier: some, many, all
- number: 0, 1, etc.
- subordinator: before, while, if, because, etc.
- coordinator: and, or

These grammatical categories are addressed in section 3.2 when VIL's grammar is explained. In this section, we will list some principles for the major categories: noun, adjective, adverb, and verb.

Principle 3.11: *Distinguishing Roles/Cases*

In VIL, roles and cases are not determined by word order in a sentence, but rather a visual distinction between nominal constituents denotes each particular semantic roles or cases, like subject, direct object, and indirect object.

Principle 3.12: *Distinguishing the Major Categories*

The major categories noun, adjective, adverb, and verb are further distinguished by each having a different border around the icon. For example,

an icon with a square border indicates that the icon is a noun, and an icon with a square border with rounded edges indicates that the icon is a verb.

This may sound redundant, since all grammatical cases already have a visual element to distinguish between them (see Principle 3.11). However, for verbs, nouns, adjectives, and adverbs it is important, since it makes the next principle possible.

Principle 3.13: *Derivational Morphology: Changing a Category*

In VIL, the meaning of an icon can be changed by changing its category marker, the border of the icon. This means that categories can be changed, for example a noun can now become a verb by adding a verb border to a noun icon. So, an icon can have more than one border and the outermost border determines its category. In the case of a verb derived from a noun, this would mean that the icon would have an outer border specific for verbs, and an inner border specific for nouns. For example, an icon for the noun "food" could have a noun-specific border, and the same icon with a verb-specific outer border could stand for the verb "to eat". Thus the outer border denotes what part of speech is meant (i.e. a verb), and the inner border denotes what part of speech it is derived from (i.e. a noun). This can be compared with natural languages, where adverbs are formed by putting –ly at the end of names of qualities as in "wise-ly". There are several ways a category can change:

1. De-Verbal Nouns and De-Nominal Verbs
Some verbs will be used as nouns, but usually it will be the other way around, because it is easier to represent nouns visually, since nouns are more likely to be concrete objects.

2. De-Adjectival Nouns and De-Nominal Adjectives
Some adjectives will be used as nouns, but, for the same reason as mentioned above, usually it will be the other way around. To keep the example of the icon for the noun "food", this noun can be turned into the adjective "edible" by putting an adjective-specific border around the nominal icon.

3. De-Verbal Adjectives and De-Adjectival Verbs
Adjectives can also be de-verbal as in "the running man", and the same principle of distinction applies. An example of a de-adjectival verb could be "to whiten", derived from the adjective "white".

4. De-Adjectival Adverbs
In VIL an adjective can be used as an adverb too and vice versa. In the sentence "He solved the problem wisely", the adjective "wise" is used as basis

for the adverb "wisely" (in a wise manner).

The following principles all have to do with verbs.

Principle 3.14: *Modality: Imperative*

> Since VIL has no linear word order, modality can have no effect on word order either. Imperative sentences will have the same construction as assertive sentences, but they are distinguished by an exclamation mark ("!").

Principle 3.15: *Modality: Passive*

> There will be no difference between passive and active sentences as in the following example:

"I kicked the ball to Jim."
"The ball was kicked to Jim by me."

These sentences have the same deep structure, and will be represented the same way.

Principle 3.17: *Yes/No Questions*

> Yes/No questions will, like imperatives, have the same construction as assertive sentences, but are distinguished by a question mark ("?").

Question in English	Question in VIL
Do you like soccer?	You like soccer?

Principle 3.18: *Wh-Questions*

> When the question word is the subject, then in VIL a question mark is also put in place of a noun phrase denoting the particular role/case:

Question in English	Question in VIL
What knocks at my window?	*"?"* knocks at my window (subject)
What did you buy?	You buy *"?"* (object)
When does the train come in?	The train comes in (*"?"* in the time-when field)
Where did you buy the sweater?	You buy the sweater (*"?"* in the place-where field)

3.2 Grammar and the Organization of the Iconicon

This section focuses on designing a grammar for VIL, and also describes the organization of the *iconicon* (i.e. lexicon of icons, see section 3.2.1), which deals with the hierarchical organization of three substantive grammatical categories (nouns, adjectives, verbs). First we will give some background in the study of syntax.

In the 60s and 70s, the syntax of natural languages was generally described by specifying a set of rules (phrase structure rules and/or transformations) and a way in which the rules interacted (how, for instance, they were ordered with respect to one another in application) to define derivations. Phrase structure rules expanded the initial symbol S (Sentence) to create an underlying constituent representation for the sentence (deep structure). Lexical items, as well as syntactic categories, were introduced by these rules. Transformations performed a variety of functions. For example, they changed the underlying grammatical relations, as in the case of passives, which were derived from actives; and they created complex sentences out of simple ones, e.g. the sentence "John thinks that Bill will leave" was derived by an embedding transformation that combined "John thinks" and "Bill will leave". Phrase structure rules and transformations were introduced by Chomsky's Syntactic Structures.[88] Grammatical sentences were those for which legal derivations were defined: ungrammatical sentences were those for which no legal derivation was defined. The rules proposed tended in general to be quite specific, in that they were particular to individual constructions in individual languages.

An advantage of a computer-based language with a grammar is that it allows the derivation process to a completed sentence to be guided by the system. This facilitates creating syntactically correct sentences. Users are free to fill in the cases, but the number and type of cases are determined by the system. A meaningless grammatical incorrect sentence like "I eat swim you book" is not possible, since one can only select a single verb in a sentence (unless a subordinator or coordinator is used). Also, other than the principles of derivational morphology, the system allows only correct words to be chosen for each category, e.g. one cannot select a verb for the grammatical category of time-when, but only time related items. For this reason, we may have to put restrictions on derivational morphology. We would probably not allow every verb, noun, or adjective to be used in derivational morphology. Of course, this means that we have to add more knowledge to the system, but this would result in more meaningful new meanings.

---

[88] **McCawley, J.D.** "Lexical insertion in a transformational grammar without deep structure," in CLS 4, 1968 reprinted in McCawley "Grammar and Meaning," 1973.

Semantics was assumed to be derived from the deep structure (i.e. result of the phrase structure rules). However, some researchers believed that the semantics was basic and that the syntactic structure was derived from that. This led to the approach called "Generative Semantics" (McCawley)[89].

More recently, syntactic representations are seen as being constructed by very general mechanisms (such as the principles of *X-bar theory*, which define what is to be a well-formed phrase structure), which are then subjected to conditions on well-formedness, which filter out unwanted structures, i.e. those corresponding to ungrammatical sentences. There is also a tendency to explain particular facts as far as is possible in terms of general principles of grammar rather than in terms of principles unique to the language being investigated.

There are five frameworks that developed from Chomsky's *Syntactic Structures* and *Standard Theory* of grammar: government and binding (GB), generalized phrase structure grammar (GPSG), lexical functional grammar (LFG), relational grammar (RG), and head-driven phrase structure grammar (HPSG). GB has seen a diminution of language-specific rules (transformational or otherwise), and their replacement by general universal principles. GPSG and LFG are competitors to GB and are more oriented toward surface structure than GB. GPSG is a framework for writing fully explicit formal grammars for natural languages (language-particular). HPSG developed from GPSG. RG uses grammatical relations to bring out what is common to clause structures in languages, which differ in word order, phrase-structure configurations, and case-marking patterns. Rather than being defined in terms of other notions, grammatical relations are primitives of the theory and are used to represent clause structure and to formulate linguistic universals as well as language-particular rules and condition. In LFG, the role of the lexicon is central and grammatical functions are seen as syntactically primitive. Many syntactic generalizations (e.g. passive) are formulated as lexical redundancy rules rather than as operations on phrase structure, like transformations. Their use allows monotonic, order-free composition on the sentence level. LFG postulates a system of interdependent representations, which encode different types of linguistic information. The most important representations are C[ONSTITUENT]-STRUCTURE and F[UNCTIONAL]-STRUCTURE. A c-structure encodes linear order and syntactic category (part of speech) in a tree. This representation is used to link syntax and phonology. The f-structure encodes information about grammatical functions in hierarchical ATTRIBUTE-VALUE MATRICES (where most grammatical functions have traditional names like 'subject' and 'object').[90]

---

[89] **McCawley, J.D.** "Lexical insertion in a transformational grammar without deep structure," in CLS 4, 1968 reprinted in McCawley "Grammar and Meaning," 1973.

[90] **International Encyclopedia of Linguistics**, William Bright (ed.), Oxford University Press, 1992.

In the 60s and 70s, the grammar of a given language had as its base a list of phrase structure rules of the general form:[91]

A   XYZ   ⟶

Such rules define as well-formed a substructure like:

$$A$$
$$X\quad Y\quad Z$$

The rule specifies precedence relations (e.g. X precedes Y and Z) and hierarchical relations (e.g. that A immediately dominates X, Y, and Z). Much recent work separates out these two functions, hierarchical structure and linear order being specified by different subsystems. Hierarchical relations are defined by the general principles of phrase structure, known collectively as X-bar theory. Linear order of constituents is defined by separate conditions - linear precedence in GPSG, and in GB principles that have to do with the direction in which case or semantic roles are assigned.

In a universal language like VIL, we are not concerned with linear order, and are only concerned with the hierarchical composition. VIL's grammar will be the simplest possible to govern the use of the chosen vocabulary. For this reason, before we created VIL's grammar, we researched and created the organization of the verb system. The verb is the main part of the sentence and all other parts revolve around it.

One approach closely related to semantic grammars is of particular interest, as is one AI semantic representation. We consider Fillmore's case grammar and Schank's conceptual dependency theory to be of special interest. In case grammar, the verb is, just as in VIL, the main part of the sentence. A sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship. Conceptual dependency theory is of interest because it is used in the iconic system CD-Icon, and because it restricts the verb system to eleven primitive ACTs. These ACTs are used to represent more complicated verbs.

3.2.1.1 Fillmore's Case Grammar

In case grammar, a sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship. Although

---

[91] **McCloskey, J.** *"Syntactic Theory,"* in *"Linguistics: the Cambridge Survey,"* Vol1 Linguistic Theory: Foundations, F.J. Newmeyer, Cambridge University, 1988, pp. 18-59.

there can be compound instances of a single case (through noun phrase conjunction), each case relationship occurs only once in a simple sentence. In the sentences,

John gave the book to Jim.
Jim was given the book by John.

the surface structures are different, but the deep structures are the same. John is still the agent and the book still the object:
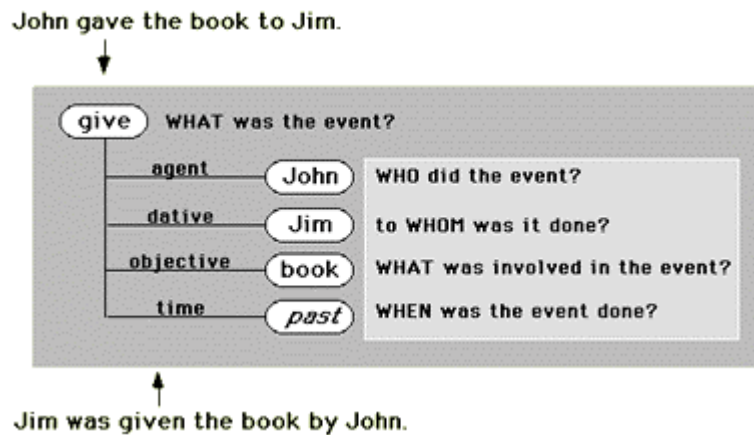
In case grammar the verbal elements of the sentence are the major source of the structure: the main verb in the proposition is the focus around which the other phrases, or cases, revolve, and the auxiliary verbs contain much of the information about modality. According to Fillmore, a sentence is made up of a *proposition*, separated from the *modality* constituent. To put it in rule form:

$$\text{Sentence} \longrightarrow \text{Modality} + \text{Proposition}$$

where the proposition is the verb and the various cases related to it (see Figure 20).

S

M   P

Negation   Verb   C1 C2 .... Cn
Tense
Aspect
.
.
.

Figure 20: tree for case grammar
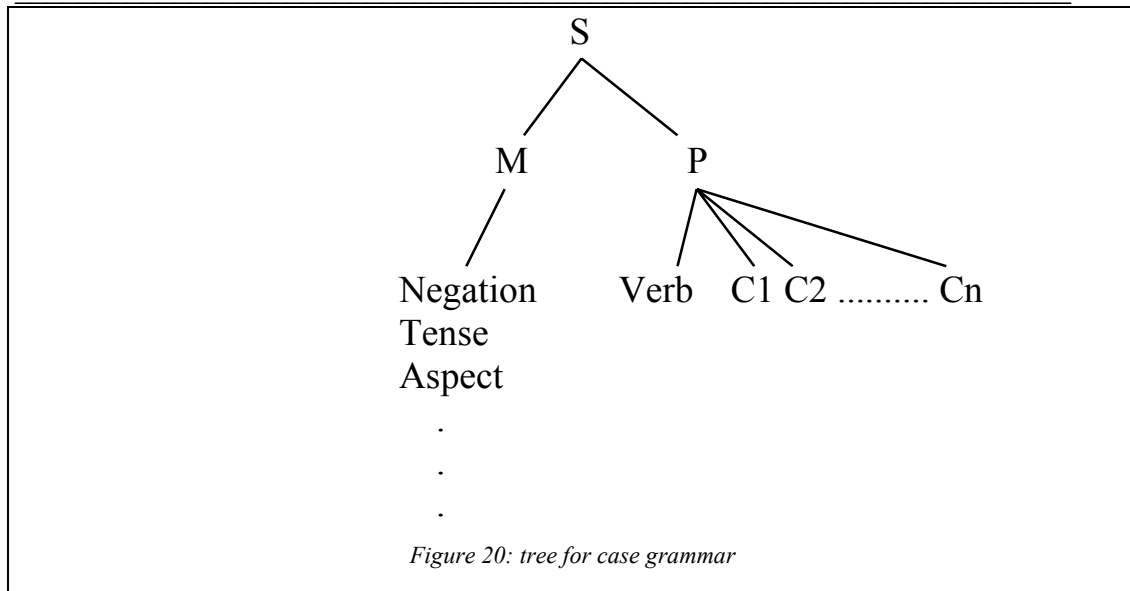
The case notions make up a set of universal concepts, which identify certain types of judgments human beings are capable of making about the events that are going on around them, judgments about such matters as who did it, who it happened to, and what got changed. Some cases that are needed are:

- **Agentive (A),** the case of actor of the action identified by the verb.
- **Instrumental (I),** the case of the object or force involved in the action state identified by the verb.
- **Dative (D),** the recipient of the state or action identified by the verb.
- **Locative (L),** the case that identifies the location or spatial orientation of the state or action identified by the verb.
- **Objective (O),** the thing being acted upon.

Case frames are the mechanisms for identifying the specific cases allowed for any particular verb. The case frame for each verb indicates the relationships which are required in any sentence in which the verb appears and those which are optional. So the case frame for the verb *kick* might be:

KICK [-- O (A) (I)]

The verb *kick* occurs in sentences with a noun phrase in the objective case, and optionally noun phrases in agentive and instrumental cases (like 'The boy kicked the ball with his foot'). The relationship between the verb and the noun phrase may be expressed by a preposition. The rules for English prepositions may look like this: the

A preposition is BY; the I preposition is BY if there is no A, otherwise it is WITH; the D preposition is typically TO; the L and T (for time) prepositions are either semantically non-empty (in which case they are introduced as optional choices from the lexicon), or they are selected by the particular associated noun (on the street, in the corner, on Monday, in the morning).

Specific verbs may have associated with them certain requirements for preposition choice that are exceptions to the above generalizations.

Verbs can be transitive or intransitive.

The boy ran.   (intransitive)
The boy kicked the ball.   (transitive)

Only transitive verbs have direct objects (like 'the ball'). Some transitive verbs have two objects (like 'give' and 'tell').

Ilan gave Mary a toy.

(Ilan-subject, gave-verb, Mary-indirect object, a book-direct object).

3.2.1.1.1 The definition of case grammar

Let us take a look at the rewrite rules used in case grammar:

S $\longrightarrow$ M + P
M $\longrightarrow$ tense, aspect,.... (the modes)
P $\longrightarrow$ V + C1....Cn
V $\longrightarrow$ kick, give, run,..... (the verbs in the vocabulary)
Ci $\longrightarrow$ (K) + NP

A sentence S is made up of the modality M (see next section) and proposition P. Proposition P is a verb V and some cases, where a case Ci is a Kasus K and a noun phrase NP. The Kasus K (which is optional, indicated by brackets) is the preposition, which introduces the noun phrase and defines its relationship with the verb (by, with,..). A noun phrase for the English language could be defined as:

NP $\longrightarrow$ (DET) + (ADJ|N)* + N + (S|NP)

This rule for NP allows an optional determiner DET, zero or more adjectives ADJ

and/or noun modifiers N for the primary noun N in the phrase, and an optional postpositional modifier (S|NP). Examples of noun modifiers are fire truck, telephone line, and office door handle. An example of a postpositional modifier is

> the boy who smokes cigarettes
> DET N ....S.................

3.2.1.1.2 Modality

Simmons distinguishes the following modalities:

Tense - present, past, future
Aspect - perfect, imperfect
Form - simple, emphatic, progressive
Mood - declarative, interrogative, imperative
Essence - positive, negative, indeterminate
Modal - may, can, must
Manner - adverbial
Time - adverbial

The last two modes, manner and time, are indicated by the adverbial parts of a sentence.

> He caught the train at the last minute. (time)
> The boy ate his food quickly. (manner)

The modes all provide additional information about the verb phrase in a sentence. Since the verbal elements are the focus of the sentence in case grammar, determining the modality establishes a large part of the overall meaning.

3.2.1.1.3 Case frames

The particular cases found in any sentence depend on the verb in that sentence. For example, the verb 'put' requires the objective and locative cases, and optionally has an agent:

Ilan put the fish in the fridge.
The fish was put in the fridge by Ilan.
The fish was put in the fridge.

But not,

Ilan put the fish. (no locative)
The fish was put. (no locative)
Ilan put in the fridge. (no objective)

Since VIL is designed much like case grammar, when the verb 'put' is chosen in our system, the system should know what cases it takes and present the user these cases.

Case frames are used to specify which cases are required or allowed with each verb. For each case there are three possibilities for its relation to a verb: required, optional, not allowed. The case frame for the verb *push* would look something like:

Verb: **push**
Case frame:
Objective - required
Locative - optional
Agent - required
Instrument - optional
Dative - not allowed

VIL works in much the same way. When a particular verb is chosen, the system presents the user with the required and optional cases. The cases that are not allowed are simply not shown. When a sentence is completed, the optional cases, which were not filled in, are removed.

3.2.1.2 Schank's Conceptual Dependency Theory

Conceptual dependency[92] theory is similar to case grammar. The representation of the meaning of a sentence revolves around the action of the sentence and contains some number of cases to relate the other parts of the sentence to the action. There are, however, some significant differences between the two. The action of the sentence is not represented by the verb (as it is in case grammar), but by the interrelationship of a set of primitive acts, each of which is a concept involved in the meaning of the verb. A major axiom of conceptual dependency is that any two sentences with the same meaning will have the same internal representation. Consider the next sentences,

Ilan gave the boy a book.
The boy received a book from Ilan.

---

[92] **Schank, R.** *"Computer Models of Thought and Language,"* W. H. Freeman, San Francisco, 1973.

These sentences have essentially the same meaning. In conceptual dependency theory, the difference between them would only be apparent in the internal representation of the person doing the action. In the first sentence, the representation of the event would be:

> EVENT1
> ACTOR: Ilan
> Action: ATRANS
> OBJECT: a book
> DIRECTION: FROM: Ilan
>   TO: the boy

ATRANS is a primitive act in conceptual dependency theory, which specifies the transfer of possession of an object. The second sentence would be:

> EVENT2
> ACTOR: the boy
> Action: ATRANS
> OBJECT: a book
> DIRECTION: FROM: Ilan
>   TO: the boy

So event1 and event2 show the basic meaning of the two sentences while maintaining the difference in emphasis provided by the different actors.

Not all sentences represent actions. In these sentences a situation, rather than an event, is described (there is no action). Consider the next sentences,

> Paul is in Amsterdam.
> The alligator is 5 yards long.

These sentences would be described as:

> SITUATION1
> OBJECT: Paul
> STATE: location
> VALUE: Amsterdam
>
> SITUATION2
> OBJECT: The alligator
> STATE: length
> VALUE: 5 yards

In each situation, there is an object, which is in a particular state. The value of the state provides specific information. States can contain physical attributes (like color, size, weight), relationships between objects (like ownership, control, possession), and a number of descriptive attributes, which apply to animate beings (like anger, health, consciousness).

The information in a sentence is represented by an internal structure known as conceptualization. Within these structures there are various roles to be played, such as actor, action, object, direction, instrument, and state. There are a number of rules, which are selected to match a particular ACT involved. These rules are called conceptual syntax rules, which will not be described here. The three roles, object, direction, and instrument, work much like the cases in case grammar and are sometimes referred to as conceptual cases. [93]

There are a small number of primitive acts, which in combination are used to represent the meaning of all actions.

PRIMITIVE ACTs

ATRANS	Transfer of an abstract relationship (give)
PTRANS	Transfer of the physical location of an object (go)
PROPEL	Application of physical force to an object (push)
MOVE	Movement of a body part by its owner (kick)
GRASP	Grasping of an object by an actor (clutch)
INGEST	Ingestion of an object by an animal (eat)
EXPEL	Expulsion of something from the body of an animal (cry)
MTRANS	Transfer of mental information (tell)
MBUILD	Building new information out of old (decide)
SPEAK	Production of sounds (say)
ATTEND	Focusing of a sense organ toward a stimulus (listen)

There is also a dummy DO that can stand in as an unknown. Furthermore, there are also a number of semantic constraints. For example, various ACTs (like MOVE) require an animate agent. Consider if the lexical item input is the verb "throw", then the primitive act for that word will be PROPEL. Knowing the primitive act provides the overall structure for the conceptualization, since the various parts of the conceptual structure are determined by the specific act. Thus for PROPEL, the parts required are actor, object, and direction. Optional are time, location, and instrumental.

---

[93] **Schank, R.** *"Computer Models of Thought and Language,"* W. H. Freeman, San Francisco, 1973.

The major advantage of representing knowledge in terms of conceptual dependency primitives, is that fewer inference rules are needed. Rules need only be represented once for each primitive ACT rather than once for every word that describes that ACT. For example, all of the following verbs involve a transfer of ownership of an object:

Give, Take, Steal, Donate

If any of them occurs, then inferences about who now has the object and who once had the object may be important. In a conceptual dependency representation, those possible inferences can be stated once and associated with the primitive ACT PTRANS. As we will see, VIL adopts this way of representing knowledge to some extent, resulting in fewer inference rules.

A disadvantage of using a universal language based more or less directly on conceptual dependency would be that it requires that all knowledge be decomposed into fairly low-level components. For one thing, simple high-level facts may require a lot of storage when broken down into primitives. Much of that storage is really wasted since the low-level rendition of a particular high-level concept will appear many times, once for each time the high-level concept is referenced. For example, suppose that actions are being represented as combinations of a small set of primitive ACTs. Then the fact that John punched Mary might be represented as in figure 21(a). The representation says that there was physical contact between John's fist and Mary. The contact was caused by John propelling his fist towards Mary, and in order to do that John first went to where Mary was. But suppose we also know that Mary punched John. Then we must also store the structure shown in figure 21(b). If, however, punching were represented simply as punching, then most of the detail of both structures could be omitted from the structures themselves. It could instead be stored just once in a common representation of the concept of punching. Furthermore, storage is not as important as cognitive effort (and time) to interpret it, since we are used to the 'size of words' in natural language.
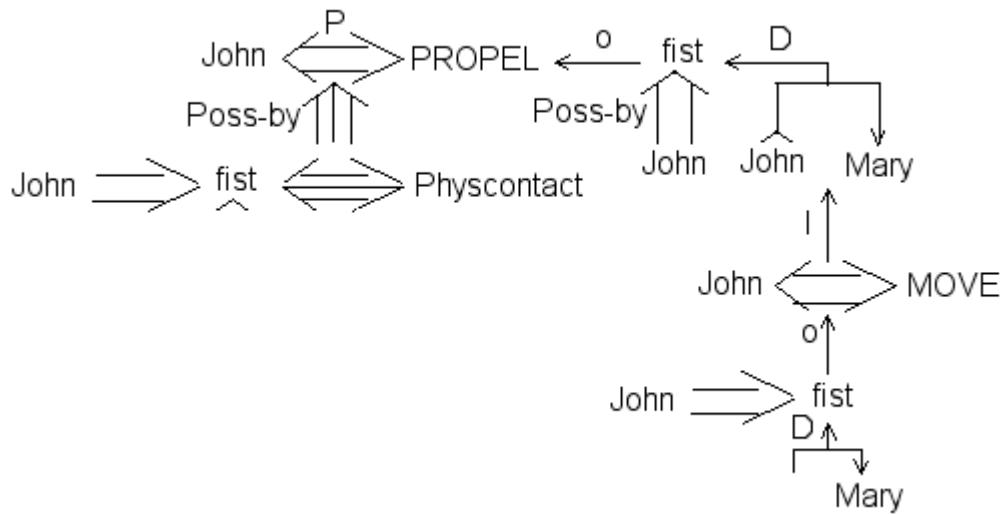
Figure 21 (a) Redundant representations: "John punched Mary."
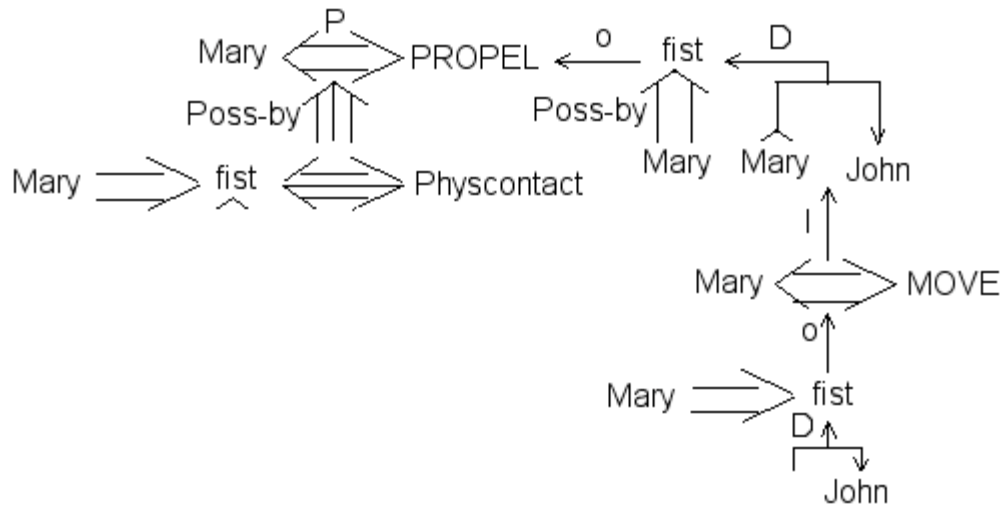
Figure 21 (b) Redundant representations: "Mary punched John."

A second, related problem is that if knowledge is initially presented to the system in a relatively high-level form, such as English, then substantial work must be done to reduce the knowledge into primitive form. Yet, for many purposes, this detailed primitive representation may be unnecessary. Both in understanding language and in interpreting the world that we see, many things appear that later turn out to be irrelevant. For the sake of efficiency, it may be desirable to store these things at a

very high level and then to analyze in detail only those inputs that appear to be important.

Conceptual dependency is a theory of representing fairly simple actions. To express, for example, "John bet Sam fifty dollars that the Mets would win the World Series" takes about two pages of conceptual dependency forms. This does not seem reasonable. As we will see, VIL will not store all verbs that lead to a sentence with the same internal representation, as in the "give" and "receive" sample above. VIL will, however, store its verbs at a higher level, meaning that decomposition will not go as deep as conceptual dependency.

3.2.2 Organization of the VIL Iconicon

In theory on grammar, the word lexicon is used to refer to vocabulary. In this dissertation we adopt the word *Iconicon* to denote a lexicon of icons. In this section we discuss the organization of three substantive categories: nouns, adjectives, and verbs. These organizations will eventually determine VIL's grammar. First we will give a principle about the depth of hierarchies and explain some of the terminology used.

3.2.2.1 Organizing Items within Categories

Studies by Liebelt et al.[94] have shown the importance of a meaningful organization of items. Semantically meaningful categories - such as food, animals, minerals, and cities - led to shorter response times than did random or alphabetic organization. A tree structure is a convenient way of organizing a large set of items. One can partition the items into (mutually) exclusive groups, although VIL actually uses lattice (more than one way to get to an item, i.e. cross-classification). If each level has eight items, then a tree with four levels has the capacity to lead the user to the correct item out of a collection of 4096 items. If there are a lot of items to classify, then there is a depth-breadth trade-off. Should one go deeper or wider? Kiger[95] has shown that breadth is preferred over depth. In his test, in which he grouped 64 items, the deep narrow tree (2x6) produced the slowest, least accurate, and least preferred version, whereas the 8x2 tree was among those rated highest for speed, accuracy, and preference. Wallace

---

[94] **Liebelt, L., McDonald, J.,** and **Stone, J.** *"Searching for items in menus: the effects of organization and type of target,"* Proc. Human Factors Society, Twenty-Seventh Annual Meeting, 1983, pp. 834-837.

[95] **Kiger, J.** *"The depth/breath trade-off in the design of menu-driven user interfaces,"* International Journal of Man-Machine Studies 20, 1984, pp. 201-213.

et al.[96] confirmed that broader, shallower trees (4x3 vs. 2x6) produced superior performance, and showed that, when users were stressed to work quickly, they made 96 percent more errors and took 16 percent longer when they used larger/deeper trees. Furthermore, Rosch has found that natural categories are 5 layers deep.[97]

Principle 3.19: *Depth of Hierarchies*

> To facilitate speed of access, each VIL tree or lattice will not be deeper than 5 levels.

One should also take into account that it is important to "reduce short-term memory load"[98]. Humans can remember *seven plus or minus two* chunks of information. This would suggest that having more than nine items at one level in the tree would negatively influence the previous test results.

Principle 3.20: *Width of Hierarchies*

> Taking short-term memory into account, each non-terminal level of a hierarchy in VIL has between 7 and 9 objects in it. The terminal levels with the actual objects to choose can have more than 9 items, since the terminals are often less complex than non-terminals.

A tree that has five levels and has eight objects per level, can already hold 32768 items. The size of the vocabulary of VIL is much smaller than this. However, it does show that there is plenty of room to extend the language to include more words.

3.2.2.2 Terminology

Let us first explain some of the terminology that will be used when talking about hierarchies.

- Hyponymy - also called the ISA relation. Used as: *An x is a (kind of) y*. For example, *maple* is a hyponym of *tree*, and *tree* is a hyponym of *plant*.
- Meronymy - also called the HASA relation or part/whole relation. Used as: *A y has an x (as a part)* or *An x is a part of y*.

---

[96] **Wallace, D., Anderson, N., and Shneiderman, B.** *"Time stress effects on two menu selection systems,"* Proc. Human Factors Society, Thirty-First Annual Meeting, 1987, pp. 727-731.

[97] **D'Andrade, R.** *"The development of Cognitive Anthropology,"* chapter 5: Folk Taxonomies, Cambridge University Press, ISBN: 0-521-45976-1, 1995.

[98] **Wallace, D., Anderson, N., and Shneiderman, B.** *"Time stress effects on two menu selection systems,"* Proc. Human Factors Society, Thirty-First Annual Meeting, 1987, pp. 727-731.

- Taxonomy - A hierarchy where knowledge is organized by kinds (ISA relation).
- Partonomy - A hierarchy where knowledge is organized by parts (HASA relation).

3.2.2.3 Nouns

There has been considerable speculation about the semantic organization of nominal concepts. It is generally agreed that they are organized hierarchically into levels, from generic to specific.[99]

According to Miller and Fellbaum[100], the topmost, or most generic level of the hierarchy is almost vacuous semantically: it is possible to put some empty abstraction designated *{entity}* at the top; to make *{object}* and *{idea}* its immediate hyponyms, and so to continue down to more specific meanings, thus pulling all nouns together into a single hierarchical structure. In practice, however, these abstract concepts carry little semantic information. Their alternative is to select a relatively small number of generic concepts and to treat each one as the unique beginner of a separate hierarchy.

Miller and Fellbaum have come up with 26 unique beginners that capture all the English nouns. They have made distinctions between abstract concepts and physical objects. The unique beginners are:

{act, action, activity}	{natural object}
{animal, fauna}	{natural phenomenon}
{artifact}	{person, human being}
{attribute, property}	{plant, flora}
{body, corpus}	{possession, property}
{cognition, ideation}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation}
{feeling, emotion}	{shape}
{food}	{society}
{group, collection}	{state, condition}
{location, place}	{substance}
{motive}	{time}

---

[99] **Smith, E.E.** *"Theories of Semantic Memory,"* in W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes (vol. 5), Hillsdale, NJ: Erlbaum, 1978.

[100] **Miller, G.A.** and **Fellbaum, C.** *"Semantic Networks of English,"* in Levin and Pinker (Eds.), Lexical & Conceptual Semantics, Blackwell Publishers Ltd, ISBN: 1-55786-354-7, 1995.

The problem with their hierarchies is that they are too exhaustive and sometimes go about 10 levels deep. The principle about the depth of the levels states that the hierarchies must not be very deep, because the user must be able to pick nouns reasonably fast. The deeper the levels the more complicated the distinctions, and the harder it will be for the user to choose which branch to follow. In order to pick the nominal concept *car,* the user would have to go the following path: artifact -> conveyance -> vehicle -> wheeled vehicle -> motor vehicle -> car. Whereas in the VIL hierarchy, we will see that the user takes a much shorter path: Transportation -> by Road -> car. Furthermore, having 26 unique beginners is too many. Keeping short-term memory requirements in mind, the ideal amount for a user to choose from is between 7 and 9. This means that, in the above classification, artifact should have a superordinate class, which in turn means that the path to car is going to be even longer.

The hierarchies of nominal concepts have been said to have a level, somewhere in the middle, where most of the distinguishing features are attached. It is referred to as the *base level* of the noun lexicon, and concepts at this level are basic concepts.[101] For lexical concepts at the base level, people can list many distinguishing features. Above the base level, descriptions are brief and general. Below the base level, little is added to the features that distinguish basic concepts. The best examples are to be found in the hierarchical lexicon for concrete objects. Basic level categories are the most general categories having members with similar and recognizable shapes. They are also the most abstract categories for which a single image can be formed for the category. It is the level at which pictures of objects are identified most rapidly. Superordinate categories seem to primarily share functional features - vehicles are for transporting, and tools are for fixing. They do not seem to share perceptual features, in sharp contrast to objects belonging to the same basic level category, which appear to share both perceptual and functional features (especially parts).

The concept of prototypes, which contain the most representative attributes inside the category, may also be useful for us. Rosch et al. found that the more prototypical of a category a member is rated, the more attributes it has in common with other members of a category and the fewer attributes it has in common with members of contrasting categories.[102] Since prototypes contain the most representative attributes inside the category, they can be used visually to represent that category. For example, the category of birds could be represented by a picture of a robin. However, question remains if prototypicality is universal. The prototypical bird for an Eskimo may very

---

[101] **Rosch, E., Mervis, C.B., Gray, W., Johnson, D.** and **Boyes-Braem, P.** *"Basic Objects in Natural Categories,"* Cognitive Psychology, 9, pp. 382-439, 1976.

[102] **Rosch, E., Mervis, C.B., Gray, W., Johnson, D.** and **Boyes-Braem, P.** *"Basic Objects in Natural Categories,"* Cognitive Psychology, 9, pp. 382-439, 1976.

well be a penguin, since it could be the only bird they see. Nominal categories, like birds, are represented in VIL by showing several items from the category. Consequently, the category icon for birds will show several birds.

Although the overall structure of noun hierarchies is generated by the hyponymy relation, details are given by the features that distinguish one concept from another. For example, a canary is a bird that is small, yellow, sings, and flies, so besides entering a canary as a hyponym of bird, the attributes of small size and yellow color could also be included, as well as the activities of singing and flying, creating a partonomy. This way, a canary should inherit from bird the fact that it has a beak and wings with feathers. So, people associate canary with at least three different kinds of information:

1. Parts: *beak, wings*
2. Attributes: *small, yellow*
3. Functions: *sing, fly*

A functional feature of a nominal concept is intended to be a description of something that instances of the concept normally do, or that is normally done with or to them. These nominal concepts can play various semantic roles as arguments of the verbs with which they co-occur in a sentence: instruments (knife-cut), materials (wallpaper-attach; wool-knit), products (hole-dig; picture-paint), containers (box-hold), etc.

The use to which a thing is put is part of a person's conception of that thing. To call something a box, for example, suggests that it should function as a container, which blocks out the thought of using it for other things. This conception may help predict what the verb of a sentence may be, or make it easy for the user to pick a verb. If the user is talking about a canary, he is likely to say something like "that canary sings beautiful" or "that canary flies really fast." Although this is an interesting conception, it is initially not useful for VIL, since in VIL the verb is chosen before any of its cases. However, we could use it in modifying sentences, where we already know the subject. The following 2 principles could speed up the encoding process, but are not implemented in the current system for VIL. They could be possible enhancements for the future.

Principle 3.21: **Using items to predict subsequent items**

> Items previously chosen could be used to reduce the search time for subsequent items. More specifically, items previously chosen could be used to start the search/choosing of subsequent items at locations other than the top of the lattice of the iconicon. For example, whenever a verb is selected, the system could suggest or even predict the possible nouns for a case, or suggest

whether the case is animate or inanimate. This could even result in the system taking the user to the right page in the iconicon, instead of starting at the top on each occasion.

Another example, in noun modifying sentences, the noun in question can be used to predict the verb of its modifying sentence. For example, consider the sentence

"The picture, that I painted last summer, sold for 500 dollars."

In VIL we would choose *to sell* first, and then *the picture*. The phrase *that I painted last summer*, however, is chosen after we picked the noun *picture*, so *picture* can be used to predict *to paint*.

One major extension to the system could be to have features associated with nominal concepts in the taxonomy. For example, some more nouns that belong in Transportation are: 1. places associated with transportation (airport, harbor, train station, gas station, etc.); 2. people associated with transportation (pilot, driver, conductor, stewardess, etc.); 3. parts of transportation media (engine, sail, wheel, cockpit, etc.). These nouns, most of which will be cross-classified, will all be part of the features of the nominal concepts in the categories. So a car will have associated with it nouns like: engine, gas, gas station, driver, engine, and other parts.

Principle 3.22: *Attributes: Adjectives as Features*

> A noun in the hierarchy can have a list of "most common" attributes associated with it. The user can say something about the attributes of a noun without explicitly having to look up an adjective for it. For example, the user may want to say "there goes a large canary". Instead of having to look up the concept of size in the adjective tree and set it to large, he/she can just look into the attribute list for canary and pick the attribute size (which is by default set to small) and change it to large.

3.2.2.3.1 VIL's Noun Classification

VIL's noun classification is composed from information, that was collected from various visual dictionaries and encyclopedias (Visual Desk Reference[103], Visual Dictionary[104], and Visual Encyclopedia[105]). The hierarchies (of each unique beginner)

[103] **MacMillian**, *"The MacMillian Visual Desk Reference,"* MacMillian Publishing Group, ISBN: 0-02-531310-X, 1993.

[104] **Corbeil, J.** *"Visual Dictionary,"* ISBN: 0-8160-1542-9, 1986.

vary in size and are not mutually exclusive. Some cross-reference is required. They will not contain entries for every possible noun. For one thing, there is no need for synonyms (since it is visual). Furthermore, the hierarchies must not be very deep, because the user must be able to pick nouns reasonably fast. The deeper the levels the more complicated the distinctions, and the harder it will be for the user to choose which branch to follow. For example, let us assume that the user wants to choose the noun *"horse"*. He/she does not want to go 7 levels deep: an animal, a vertebrate, a mammal, a herbivore, a perissodactyl, an equid, horse. This will take too long and the decisions he/she has to make are far too hard. A better and shorter way to get to horse would be: an animal, a mammal, horse. This is possible because there is a limited amount of entries listed for each category. These are generally entries that people talk about the most in everyday speech, and is based on data about frequency of usage.

Principle 3.23: **Lattice in Hierarchies or Cross-Classification**

> A lattice allows multiple paths to one entry, thus it supports cross-classification. Certain items in our noun and verb hierarchies will be cross-classified for convenience and speed of encoding.

Furthermore, in our hierarchies, objects and abstract concepts are classified together, instead of having separate unique beginners for concrete and abstract concepts like Fellbaum and Miller. For example, to get at the abstract noun "memory", one has to take the following path: Physical World -> Living World -> Humans -> Psychology -> Mind -> memory (5 levels deep + terminal level). This may not seem straightforward to some people.

Following is the classification of nominal concepts that we propose for VIL. There are two unique beginners:

1. Natural Objects/Concepts
2. Man-Made Objects/Concepts

However, because of the principle about width of hierarchies, we present the user with 7 to 9 items per level as much as we can. Consequently, the user never sees these two high level categories. They only exist on paper to make a clearer grouping. The categories the user sees are the following seven beginners, of which categories 2, 3, 4, 5, 6, and seven fall under man-made objects and category 1 under natural objects:

1. Physical World
2. Beliefs, Customs, and Society

---

[105] **Dorling Kindersley**, *"Visual Encyclopedia,"* ISBN: 1-56458-985-4, 1995.

3. Arts & Entertainment
4. Sports
5. Communication
6. Science & Technology
7. Transportation

We will now discuss each of the hierarchies, and give subclasses (for non-terminal categories) or examples in italics (for terminal categories) of nouns that fall into the category in question.

3.2.2.3.2 The Physical World

The Physical World consists of:

1. Living World
2. Universe & Space: *universe, space, planets, sun, moon, and galaxy*
3. Geography: subclasses countries, rivers, mountains, deserts, and environmental issues like *pollution, acid rain, ozone layer, and rain forest*
4. Geology: *stone, mineral, earthquake, clay, soil, and gems*

Not all subclasses for Geography contain icons. Some nouns, e.g. specific countries are represented graphically. A map is used from which one can select the country in question.

The Living World consists of:

1. Animals
2. Evolution: *dinosaur, fossil*
3. Fruit
4. Humans
5. Micro Organisms   *bacteria, virus, cell*
6. Flora & Fauna: subclasses plants, trees, flowers, herbs, and spices
7. Vegetables
8. Humans
9. Food

The Animals category consists of:

1. Mollusks: *snail, squid, shell*
2. Insects, Aranchids and
   Crustaceans: *crab, fly, grasshopper, lobster, shrimp, spider*

3.	Reptiles & Amphibians:	*alligator, frog, salamander, snake, turtle*
4.	Fish:	*whale, shark, tuna, swordfish*
5.	Birds:	*chicken, duck, ostrich, penguin, rooster, turkey*
6.	Mammals:	*cow, elephant, horse, pig, rabbit, sheep*
7.	Pets:	*cat, dog, bird, rabbit, hamster, horse*

The <u>Humans</u> category consists of:

1.	Body	subclasses interior parts (e.g. *digestion, heart, blood, nerve, brain, muscle, throat*), exterior & skeleton parts (e.g. *arm, leg, joint, teeth, skin, hair, nail*), physical fitness (e.g. *pulse, blood pressure, temperature, shape, sleep, weight, height)*
2.	Food	subclasses cooking liquids, drinking liquids, dishes, dairy products, grain products, spices, and types of meals
3.	People:	subclasses family, occupations, other people
4.	Household	see science -> household
5.	Psychology	subclasses mind (*perception, learning, thought, memory, hypnoses, intelligence, creativity, emotions, personality*) and mental disturbance (*neurosis, eating disorder, phobia, paranoia, schizophrenia, addiction*)
6.	Medicine	subclasses drugs, illnesses, occupations, type of medicine, and first aid

Some nouns are de-verbal, like perceive-perception and think-thought. Depending on the importance/frequency of use of the de-verbal noun, these nouns are represented in one of two ways. The user could find a noun, for instance "thought", in either of two ways:

1. go to the noun classification, elect to make a verb de-verbal, and select the verb "to think" from the verb classification. The system will put a noun-border around the selected verb to make it the noun "thought" (see Principle 3.13: Derivational Morphology).
2. go to the noun classification and search for the noun "thought". This noun is derived from the verb "to think" and is cross-classified (see Principle 3.19: Lattice in Hierarchies or Cross-Classification). The verb and the noun would still have the same representation except that the noun would have an extra border.

3.2.2.3.3 Beliefs, Customs, and Society

The Beliefs, Customs, and Society category consists of:

1. Law & Order: *peace, law, lawyer, court, officer/cop, and jail*
2. Religion: *Jew (Chanukah), Christian (Christmas, Easter, bible), Islam (Muslim, Koran), Hinduism, prayer, god, worship, faith, philosopher, myth*
3. Customs: *wedding, rituals, gesture, charm, clothing*
4. Money: *coins, bank, currency (dollar, pound, yen), credit card*
5. Politics & Royalty: *president, vote, queen, prince, parliament, communism, capitalism*
6. War & Peace: *war, peace, civilization, weapon (gun, knife, canon), revolution, murder, assassin, death*
7. Occupations   *plumber, mechanic, lawyer*
8. Family: *marriage, wedding, father, mother, son*

3.2.2.3.4 Arts & Entertainment

The Arts & Entertainment category consists of:

1. Theater: *theater, stage, cinema, opera, comedy, drama, musical*
2. Music & Dance: *opera, instruments (guitar, piano, etc.), song, anthem, singer, conductor, ballet, dancer*
3. Painting & Sculpting: *sculpture, art, painting, statue, icon, mural*
4. Literature: *novel, poem, dictionary, book, magazine*
5. Cinema & Photography: *photo, camera, film, movie (western, comedy, drama), cinema*
6. Architecture: *temple, castle, house, skyscraper, museum, stadium*
7. Furniture: see Science->Household->Furniture
8. Fashion   see Science->Household->Personal Articles->Clothing

3.2.2.3.5 Sports

The Sports category consists of:

1. Athletics: *track (hurdle, sprint, long-distance, relay), field (javelin, jump, discuss)*
2. Display: *gymnastics, figure skating, parachuting*
3. Target: *darts, golf, bowling, pool, billiards*

4. Court:	*tennis, table tennis, badminton, squash, racquetball*
5. Team:	*baseball, basketball, cricket, football, soccer, field hockey, ice hockey*
6. Racing:	*swimming, sailing, skiing, bobsled, rowing, auto race, cycling, motor race*
7. Fighting & Combat:	*boxing, karate, judo, wrestling, fencing*
8. Olympic Games:	*summer games, winter games, paralympics*

3.2.2.3.6 Communication

The Communication category consists of:

1. Language:	*language, writing, signs, icons, alphabet, letter, word, sentence, Morse, Braille*
2. Mail:	*letter, stamp, package, telegram, airmail, e-mail*
3. Computer:	*software (floppy disk, CD-ROM), hardware (printer, modem, mouse, fax, chip)*
4. Electronic:	*fax, telephone, Morse, satellite, video, typewriter, radio, phonograph, tape recorder, television, compact disc*

3.2.2.3.7 Science & Technology

The Science & Technology category consists of:

1. Physics & Chemistry:	*matter, atoms, energy, gravity, magnet, metal, chemical, light, force, pressure, temperature, sound*
2. Time & Measurement:	subclasses are time, distance, measurements, and general
3. Machines:	*bulldozer, dump truck, fork lift, machine*
4. Household:	subclasses are cleaning, furniture, kitchen, and personal articles
5. Communic. Electronics:	see Communication -> Electronic
6. Computer	see Communication -> Computer
7. Transportation	see Transportation

The Time & Measurement category consists of:

1. Time:	*second, hour, day, week*
2. Distance:	*meter, yard, inch*

| 3. | Measurement: | gradables like weight, size, area, volume, density (could be de-adjectival) |
| 4. | General: | *square, circle, triangle, angle, area* |

The <u>Household</u> category consists of:

1.	Cleaning	*broom, sponge, bucket*
2.	Furniture:	*bed, table, armoire, lamp*
3.	Kitchen:	subclasses electric machines, cooking utensils, pots & pans, and silverware
4.	Personal Articles	subclasses clothing, personal articles (*razor, comb, umbrella, luggage, etc.*), and personal adornment (*jewelry, eyeglasses*)

3.2.2.3.8 Transportation

The <u>Transportation</u> consists of:

1.	by Road:	*car, motorcycle, bus, bicycle*
2.	by Rail:	*train, subway, locomotive*
3.	by Sea:	*boat, submarine, ferry, cruise, sailboat*
4.	by Air:	*plane, jet, helicopter*
5.	through Space:	*rocket, space shuttle*

3.2.2.4 Adjectives

Two kinds of modifiers are usually distinguished. Roughly, those that modify nouns are called adjectives, and those that modify anything else are called adverbs. We have not explored the semantics of adverbs. Since the majority of adverbs are derived from adjectives, we have adopted the same semantic organization as adjectives.

In English, there are two sorts of adjectives, predicative and non-predicative adjectives. A predicative adjective is one that can be used in the predicate of a sentence. For example, *tall* is a predicative adjective in *the girl is tall*. Most adjectives can be used attributively: *the tall girl*. Many adjectives cannot be used predicatively: *the former champion* is acceptable, but not *the champion is former*.

Levi has provided several criteria for distinguishing between predicate and non-predicate adjectives:[106]

---

[106] **Levi, J.N.** *"The Syntax and Semantics of Comples Nominals,"* New York: Academic Press, 1978.

1. Predicate and non-predicate adjectives cannot be conjoined: *the tall and corporate lawyer* is awkward.
2. Non-predicative adjectives are not gradable: *the extremely corporate lawyer* is odd.
3. Non-predicative adjectives cannot be nominalized: the predicative use of *nervous* in *the nervous person* admits such constructions as *the person's nervousness*, but it's non-predicative use in *the nervous disorder* does not.

Miller and Fellbaum[107] state that by all three criteria, non-predicative adjectives resemble nouns that are used as adjectives. For example, in "baseball game" the noun baseball is used as an adjective to modify game, but, like a non-predicative adjective, the nominal adjective does not conjoin (the long and baseball game), is not gradable (the extremely baseball game), and cannot be nominalized (the game's baseballness). Consequently, Miller and Fellbaum consider non-predicative adjectives stylistic variants of modifying nouns, and have assumed that they are entered in lexical memory as a special class of adjectives, and are defined by pointers to the pertinent noun concept. Consequently, in VIL non-predicate adjectives are not classified with the adjectives, but are selected via the noun hierarchy, i.e. they are de-nominal adjectives.

3.2.2.4.1 Antonymy

The semantic organization of predicative adjectives is different from that of nouns. There is no hyponymic relation (ISA) for adjectives: it is not clear what it would mean to say that one adjective *is a kind of* some other adjective. The basic relation among adjectives is antonymy. When a probe word in a word association test is a familiar adjective, the response commonly given is the antonym of the probe. For example, when the probe is *good*, the common response is *bad*; when the probe is *bad*, the common response is *good*. This mutuality of association is acquired as a consequence of these pairs of words being used together frequently in the same phrases and sentences.[108]

The importance of antonymy is understandable when it is recognized that the function of predicative adjectives is to express values of attributes, and that most attributes are

---

[107] **Miller, G.A.** and **Fellbaum, C.** *"Semantic Networks of English,"* in Levin and Pinker (Eds.), Lexical & Conceptual Semantics, Blackwell Publishers Ltd, ISBN: 1-55786-354-7, 1995.

[108] **Justeson, J.** and **Katz, S.** *"Co-occurences of antonymous adjectives and their contexts,"* Yorktown Heights, NY: IBM Research Report, 1989.

bipolar. Antonyms are adjectives that express values at opposite poles of an attribute.[109]

Gradable Adjectives

Two terms are said to be contradictory if the truth of one implies the falsity of the other. Two terms are said to be contrary if only one proposition can be true, but both can be false. For example, *alive* and *dead* are contradictory, and *fat* and *thin* are contrary (one can be of average weight). Gradable adjectives can be defined as those whose value can be multiplied by such adverbs of degree as very, rather, quite, somewhat, pretty, and extremely.[110] Contraries are gradable, contradictories are not.

3.2.2.4.2 VIL's Adjective Classification

The following is the organization of adjectives we propose. Almost all of these adjectives are predicative. The non-predicative adjectives, as stated before, are derived from nominal concepts in the noun taxonomy. First a few principles are given.

Principle 3.24: *Non-Predicative Adjectives*

> Non-predicative adjectives are derived from nouns. In VIL the user selects the noun and uses it as an adjective, distinguished by a visual element as an adjective (see Principle 3.13: Derivational Morphology).

The basic relation among adjectives is antonymy. Since antonyms may not be very clear when presented in isolation, it is better to represent them together for maximum clarity. The following two principles are about antonymy:

Principle 3.25: *Contrary Adjectives*

> Contrary adjectives are gradable. They should be represented together using a continuum. In VIL these are represented by sliders with at either end the two antonyms (see chapter 4 to see how this is represented visually). For example, *fat* and *thin* are contrary, but there are other values between these opposites. In VIL, the value of the slider indicates the actual value (e.g. of average weight).

---

[109] **Miller, G.A.** and **Fellbaum, C.** *"Semantic Networks of English,"* in Levin and Pinker (Eds.), Lexical & Conceptual Semantics, Blackwell Publishers Ltd, ISBN: 1-55786-354-7, 1995.

[110] **Cliff, N.** *"Adverbs as Multipliers,"* Psychological Review, 66, pp. 27-44, 1959.

___

Principle 3.26: ***Contradictory Adjectives***

> Contradictory adjectives are not gradable. They should be represented together as distinct choices. In VIL these are represented, not by sliders, but by showing the two antonyms together and having an arrow pointing to the one that is meant. This is to make the distinction clearer between the two antonyms.

The adjectives are divided into three groups:

A. Lower Level Perceptual
B. Higher Level Evaluative
C. Comparative/Relational

Let us look at each of them in turn.

A. Lower Level Perceptual
The lower level perceptual adjectives are primarily comprised of adjectives that involve our 5 senses: sight, hearing, touch, smell, and taste.

1. Visual: *round/flat, straight/bent, dark/light, empty/full*
2. Color: *black, white, red, yellow, green, blue, brown, purple, pink, orange, gray*
3. Aural: *loud/soft*
4. Tactile: sticky, round/flat, straight/bent, hard/soft, light/heavy, sharp/smooth, stiff/elastic, loose/tight, warm/cold, wet/dry
5. Olphactory: *smelly/fragrant*
6. Taste: *acid, bitter, sweet, sour*
7. Measurement: subclasses size & dimension, distance, temperature, volume, weight

The category of Measurement consists of the following categories:
1. Size/Dimension: *deep/shallow, high/low, large/small, long/short, thick/thin, minimum/maximum*
2. Distance: *meter, yard, inch*
3. Temperature: *Fahrenheit, Celsius*
4. Volume: *centiliter, deciliter, liter, gallon, fluid ounce, pint, quart, cup, spoon, teaspoon*
5. Weight: *gram, kilogram, ounce, lbs, hand full*

Colors

The eleven colors mentioned above are a result of a cross-cultural study of color terminology by Berlin and Kay.[111] They found that the number of basic color terms varied - some languages had as few as two basic level color terms, while others had as many as eleven. The basic color terms appeared in a particular order. If a language had just two basic color terms, these were always *black* and *white*. If a language had three basic color terms, they were always *black*, *white*, and *red*. If a language had four color terms, they were *black*, *white*, *red*, and *green* or *black*, *white*, *red*, and *yellow*. The following table shows the seven ordered stages that Berlin and Kay found for focals:

Stage I	black, white
Stage II	black, white, red
Stage IIIa	black, white, red, yellow
Stage IIIb	black, white, red, green
Stage IV	black, white, red, yellow, green
Stage V	black, white, red, yellow, green, blue
Stage VI	black, white, red, yellow, green, blue, brown
Stage VII	black, white, red, yellow, green, blue, brown, purple, pink, orange, gray

This kind of ordering is called an "evolutionary sequence" because it predicts the sequence in which languages will develop new basic level terms. Thus if a language has a stage IV color term system, one can predict that when a new basic level color term enters the system, the new term will be blue.[112]

The international SI-system

Noun modifying properties pertaining to weight, volume and distance are harder to visualize, since different cultures may use different systems of measure. Even if we decide upon usage of one particular system, how does one represent, for instance, a "kilogram". In ancient times elements of weight, volume, and distance were expressed in terms of real-life objects. Examples of currently still used measurements are a foot, which is the average height of a grown man's foot, measures around 30 centimeter. An inch is the length of a thumb, around 2.5 cm.

Most of these real-life measurements are not used anymore. We have chosen a system that is widely used and agreed upon as an international standard. It is called the SI-system[113]. The SI system was proposed over 300 years ago. Around 1875 about 18

---

[111] **Berlin, B.** and **Kay, P.** *"Basic Color Terms,"* 1960s.

[112] **D'Andrade, R.** *"The development of Cognitive Anthropology,"* chapter 5: Folk Taxonomies, Cambridge University Press, ISBN: 0-521-45976-1, 1995.

[113] **Elsevier** *"Grote Winkler Prins Encyclopedia,"* Elsevier Publishers, Part 8, pp. 143-145, Amsterdam, 1984.

major countries had agreed to use the system as their standard. In 1971 the European Community decided to adopt the SI system as their system for measurement. The SI units are international symbols, and even though they are comprised of letters, e.g. "kg", they are symbols. VIL uses these symbols as part of its icons for measurement. The following table gives some symbols from the SI system.

Type of measurement	Unit	SI Symbols	Derived from
Length	meter	m, dm, cm	
Weight	kilogram	kg, g	
Volume	liter	L, dl, cl	1 L = 1 dm³
Temperature	degree Celsius	ºC	

There are of course still countries using a different system. VIL also uses some of these symbols to achieve a wider coverage.

Type of measurement	Unit	Symbols
Length	foot, yard, inch	
Weight	ounce, pound	oz, lbs
Volume	gallon, pint, quart, fluid ounce, cup	gal, pt, qrt, floz
Temperature	degree Fahrenheit	ºF

One good enhancement for VIL would be to have a built in conversion tool, which converts one unit of measurement to another, e.g. 1 gallon = 3.94 liter. This will be built in the next release of VIL.

B. Higher Level Evaluative
The higher level evaluative adjectives are adjectives that involve our ability to judge and evaluate situations and events.

Group 1:   antonyms: *right or correct/wrong, true/false, good or great/bad*
Group 2:   *important, serious, regular, normal, common, fixed, strange, special, possible, probable, necessary, able, ready*
             antonyms: *complex/simple, general/specific, private/public*
Group 3:   antonyms: *beautiful/ugly, cheap/expensive, clean/dirty, safe/unsafe*
Group 4:   antonyms: *male/female, married/divorced, new/young/old, poor/rich*
Group 5:   *tired, awake*
             antonyms: *happy/sad or angry, fat/thin, healthy/ill, drunk/sober, death/alive*
Group 6:   *violent, lazy, selfish, greedy, jealous, afraid, honest, nervous*
             antonyms: *kind/cruel, wise/stupid or foolish*

Group 7:   *physical, automatic, conscious*
Group 8.   *free, complete, full*

The following groups can be related to grammatical cases used in VIL like manner, time-when, and frequency of occurrence. Most are de-adverbial:

1. Manner: *sudden, quick or fast/slow*
2. Time: *past, present, future, early/late*
3. Ordinal: *first, second, ..., last, one, two, three, etc.*
4. Freq of Occurrence: *frequent, seldom/often*

C. Comparative/Relational
The Comparative/Relational adjectives are adjectives that involve our ability to compare and relate things.

There is only one group and it consists of adjectives like: *parallel, separate,* and *like*. It contains the antonyms: *same/different* and *equal/opposite*.

3.2.2.5 The Classification of the Verb System

This section focuses on finding some general principles by which to categorize verbs. Verbs will be classified according to their shared meaning. The resulting hierarchical ordering of verbs will serve as a basis for choosing verbs in VIL. Since VIL is a communication program by which people that speak different languages can communicate, the classification of verbs must hold across languages. The meaning of verbs is the same across languages (e.g. "to give" means "a transfer of possession"), and therefore a classification according to shared meaning like we propose should be universal. Since our system is an iconic system, we are not only concerned with how people choose and think of verbs, but also how to represent them (iconically). How to represent the verbs will be discussed in the next chapter.

Another issue of concern is that of primitive elements versus indexing. Do people choose verbs by thinking of their primitive elements (e.g. "to run" is "to move fast in a direction") or do people choose verbs by indexing (e.g. "all verbs that have to do with ice cream")? In the latter case, the user would be focusing on the *topic* of the sentence, namely "ice cream". In the former case, the user would be focusing on the *goal* or on the *result* of the activity, namely "to get somewhere by moving fast". A third way of choosing verbs might be a combination of the two. For example, once we have determined that the verb is one of motion, then one could choose from verbs of motion over *land* vs. verbs of motion through *air* (e.g. fly) or *water* (e.g. row). Psychological experiments will have to determine which is the proper way. We have

100

adopted the notion that people think of the result that the sentence will have. Consider the sentence,

I give Mary the Book.

We assume that people focus on the goal, which is a transfer of the book from me to Mary, i.e. a transfer of possession. People tend not to think of "all verbs that have to do with books".

Principle 3.27: ***Decomposition of Verbs***

> Verbs are decomposed into primitive elements. This means that more specific verbs will have to be described by their more general version modified by some adverbs of manner or 'aktionsarten' (e.g. *argue angrily* for *quarrel*)[114].

Now that we have decided on what principle to use for classifying verbs, namely decomposition into their primitive elements, another issue of concern becomes that of to what extent can or should verbs be decomposed into their primitive elements. For a universal language, we should ideally take all languages into account. For example, consider an Eskimo language that has ten different verbs for throwing snow. It is more than likely that other languages will only have one or not even one, but just a verb for throwing. In this case only throwing will be categorized. As with other issues, we will be guided by the principle of the '*greatest common denominator*'. This principle states that we represent that which most languages share (see Principle 3.1: Greatest Common Denominator or Minimal Grammar and Vocabulary). For this case, to represent different ways of throwing anything, one will have to use modifiers, like adverbs (for manner) or '*aktionsarten*' (begin to throw, not yet throwing, in the middle of throwing, stop with throwing).[115] The ten verbs of throwing snow need to be decomposed, as they are too specific. On the other hand, decomposing verbs too much might quite possibly confuse the user in choosing which category the verb belongs to. Consider the following lengthy path to the verb "to throw": choose verbs of "transfer", then choose verbs of "transferring an object", then choose verbs of "physically transferring an object", and then choose verbs of "physically transferring an object by applying force to it", of which "to throw" is a verb. Also, a verb like "to swim" should be classified instead of having to paraphrase it as "to move through water by motion of one's hands".

---

[114] 'Aktionsarten', like start, stop, and continue, are a means of paraphrasing verbs. They reduce the number of verbs that need to be classified.

[115] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

Some of the criteria for the classification are:
- adequate coverage of actual verbs, and the groups of verbs should be learnable and understandable, and also have consistent roles for arguments.
- limited decomposition.
- If one class of verbs is equal to another class of verbs by using some modifier, then collapse these two classes into one, and use the modifiers to represent those verbs.
- Verbs that are derived from *nouns* are cross-classified in the verb hierarchy. The icons representing these verbs are the same as the icons for their nominal version. These include verbs of motion that involve transportation vehicles (e.g. "to bike", "to skate"), and verbs that are means of communicating (e.g. "to radio", "to fax", "to modem", "to telephone"). See Principle 3.13: Derivational Morphology.
- small number of primitive (visual) elements.

We have looked at several classification schemes for English verbs to determine what classes existed and what the criteria were for splitting up verbs into these classes. We will describe some of them and discuss how they are or are not useful for our classification. We researched the following work: the classification of speech acts by Ballmer and Brennenstuhl; English verb classes and their alternations by Levin (Levin has quite an extensive set of verb classes that will turn out to be very useful for us); Schank's primitive acts; the verb system of Basic English; Matsukawa and Yokota's approach on developing a concept dictionary; and non-spatial semantic fields by Jackendoff. After presenting some relevant aspects of this research, we present our proposed hierarchy.

In order to classify verbs, we have to find distinguishable elements for separating verbs and putting them in different groups. For example, some distinguishable elements could be verbs of contact, verbs of state change, or even verbs that are opposites of each other. We have found three kinds of discriminatory features:

- ***Shared meaning***

   Shared meaning is the basis for the hierarchy, i.e. grouping/sorting takes place on basis of shared meaning. They can be seen as hierarchical browsing features, since they are sorting features, used to go through the tree.

- ***Case predictive features***

   These feature serve as a basis for suggestion. In other words, having chosen the verb, the system could make suggestions about its arguments. It could not only predict the presence or absence of specific cases, but also the nature of these

cases. For example, a verb like "to give" that expresses a transfer of possession, usually takes three arguments, a subject, a direct object, and an indirect object, of which the subject and indirect object are typically animate and the direct object typically inanimate.

- ***Primitive combining features (visual elements)***

These are features that can be used in composing icons. They will be part of icons, e.g. a visual entry in an icon to distinguish/characterize verbs of motion.

3.2.2.5.1 Ballmer and Brennenstuhl's Speech Act Classification

In this approach Ballmer and Brennenstuhl[116] try to classify all speech acts. The basic idea was to group verbs that are similar in meaning. This research was in the first place meant to be a piece of linguistic work. It could also be used as a dictionary for synonyms. They succeeded in the structuring of part of the lexicon into semantic categories. The classification they came up with consisted of 8 model groups, 24 models and typifications, 600 categories, and 4800 speech act verbs. The next few sections describe how they selected verbs, formed categories, and formed models. At the end, the possibility of applying their approach to the entire verb system will be discussed, and also whether some of it is useful for our purpose.

3.2.2.5.1.1 Selecting Verbs

The first step in the classification task is to pick out the speech act verbs from the entire verb thesaurus. For material they used Oxford English Dictionary, Webster's, and the classification of German speech act verbs. The main rule they use to find whether a verb is a speech act, is to insert a VP containing the verb in question in the following sentence frame:

Someone VP-past "..."

If the frame makes sense when filled, then the verb is a speech act verb. The VP can be understood to specify the content or means of expression of the linguistic activity "....". Consider the next examples:

John declared "I am hungry".

---

[116] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

The criminal blackmailed the director "If not, I'll kill you".
The teacher disapproved of his pupil "You will never succeed".
* The car driver transported the noodles "Here they are".
* David kicked George "You asshole".

There are also some borderline cases that have to do with verbs of thinking:

Charley calculated "Four and three is nine".
John pondered "If they were to get a divorce, things would not get much better".

One can say that the speech act is occurring within the persons themselves, without explicitly saying something.

3.2.2.5.1.2 Forming Categories

After determining which verbs fall under speech acts, categories can be formed. There are two phases in forming speech act categories. The first is a heuristic phase in which the meaning content of the verbs is taken as a guideline to get groupings of verbs in areas around semantic centers. Some semantic centers could be: Expressing Emotion, Influencing Others (Enactions), Verbal Struggle, Normative Behavior (Nouns), Expressing Values, or Complex Discourse Functions. As an example, Ballmer and Brennenstuhl take *verbs of verbal struggle*. The heuristic method yielded a list of verbs related in their major meaning component to *verbal struggle*. Let us assume they had 90 of those verbs (e.g. answer, claim, decline, reject, settle, withdraw, ...). Now smaller groups of verbs that are semantically more homogeneous need to be set up. This is the second phase, in which they make use of the semantic relation meaning similarity (or meaning adjacency), but also of the relations prototypicality and group-homogeneity.

For example, the next group of verbs denotes the activity of being "in the middle of a verbal struggle": argue, bargain, bicker, contend, contest, haggle, quarrel, struggle, wrangle. These verbs are all similar in meaning. In the process of grouping, the groups are given a name (predicate), which are nouns that cover the meaning of all the verbs occurring in the group. An IS-A relation holds between the verbs and this name. The predicates they found appropriate for verbs of struggle are the following: [117]

---

[117] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

Starting Situation,
Making Claims,
Dissent,
Attack,
Tactical Phase: Defense Attempt,
                Successful Repulsion,
                Reply to Defense,
                Involvement,
Making a Coalition,
Terminal Phase: Retreat,
               Victory,
               Defeat,
Willingness to Cooperate.

A verb group together with its title predicate is called a *category*. For example, the resulting category from phases one and two for "Making Claims" is:

Making Claims: request, require, trap, venture, ask, chat someone up, claim, exact.

3.2.2.5.1.3 Forming Models

After forming categories, models can be formed. They claim that there is a natural order in the models. For the struggle model this is a temporally one (Starting Situation, then Making Claims, ..., and lastly the Terminal Phase). The ordering is not always temporal. Consider the verbs of enaction. Categories are: [118]

Volitions,
Putting into Focus,
Enactions: Call for Help,
              Commissioning,
              Pointing to a Norm or Danger,
              Demanding,
              Commanding,
              Influencing,
              Demanding Effectively,
Reactions to Enactions.

---

[118] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

The four superordinate categories are temporally ordered, but the subordinate categories are not. They are rather ordered according to the rising degree of effectivity of encation. The first categories are weaker attempts to get the hearer to do something, whereas the later categories comprise more and more powerful actions.

3.2.2.5.1.4 The Entire Verb Thesaurus

They classify the entire verb thesaurus as follows (translated from German):

> Models Concerning States of Affairs,
> Models Concerning Processes,
> Models Concerning the Existence of Individuals and Objects,
> Models Presupposing the Existence of Individuals and Objects,
> Movements of Objects in Space,
> Experiences (influence of the environment on individuals and objects),
> Effecting (influence of individuals and objects on the environment),
> Controlled Interference (directed towards oneself, the environment, and others; Actions),
> Controlled Manipulation of Objects and Individuals (Grasping, Transport, and Work at),
> Controlled Production and Destruction of Objects and Environment,
> Controlled Property Transactions (Property, Handing Over),
> Speech Activities (Expression, Appeal, Interaction, and Discourse).

3.2.2.5.1.5 Paraphrasing

Ballmer and Brennenstuhl proposed that the most general verb occurring in a category could be taken as the category name, and could stand for the meaning content in this category. Specific verbs within the category could be formed by paraphrasing the category name. The paraphrases of the verbs in the category are obtained by modifying the category name by adverbs and adverbial, especially prepositional, noun phrases:[119]

Argue	old category name: Involvement
Argue	argue
Bargain	argue about goods with the intention of buying or selling them
Bicker	argue angrily
Content	argue with the intention of reaching an aim

---

[119] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

Haggle	argue about goods with the intention of buying or selling them
Quarrel	argue angrily
Struggle	argue in order to get something
Wrangle	argue grimly

The category name itself could be paraphrased like: [120]

argue	=	*argue*
	=	*fight* verbally
	=	*interact* competitively in order to enlarge one's domain of control by verbal means

The category belongs to the model of Verbal Struggle (= arguing), the more basic model of (nonverbal) Fighting, or to the model group of Interaction. The place of categories of the Struggle Model could be determined by "aktionsarten" (*not yet, prepare, ...*) and paraphrased like:

not yet argue
prepare to argue
begin to argue
intensify arguing
argue
reduce arguing
cease to argue
no longer argue

This reduction to elementary verbs and paraphrasing from there has been done for the entire German verb thesaurus and yielded 11 candidates for elementary verbs: [121]

gelten	hold
ablaufen	proceed
geschehen	happen
existieren	exist
(als Teil) enthalten	entail
verursachen	cause
wahrnehmen	perceive
wollen	want
versuchen	attempt

---

[120] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

[121] **Ballmer, Th.** and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

| beruhren | touch |
| benutzen | use |

Speech acts are not included, because they say that they are often linguistically specialized versions of normal, nonlinguistic verbs or can be related to such verbs.

The notions of 'aktionsarten' are a good means of paraphrasing verbs. It greatly reduces the number of verbs that need to be classified. All the verbs that can be obtained by paraphrasing a more general verb, need not be classified. They can just be chosen by choosing the most general verb (e.g. the verb that stands for the meaning content of verbs in that class, like "argue"), and modifying it.

Although these notions are attractive, the eleven primitive/elementary verbs they found are not adequate for our purpose. Some of the eleven elementary verbs do however capture some of the abstract meaning we are looking for. For example, "to exist" is a big category in all other approaches. The notion of paraphrasing (aktionsarten) will come in handy, but will be used at the lowest levels in the hierarchy to make subtle differences between verbs. For example, the ten verbs for throwing snow might be distinguished by aktionsarten. However, since in VIL we want adequate coverage of verbs, some paraphrasing and subtleties in meaning are present in the icons. So, for these we do have separate icons. Consider, for example, the verb "to give", which is a transfer of possession. Other verbs, which are a transfer of possession could be paraphrased in terms of "to give":

Give	exchange of possession
Receive	give done by other party
Sell	give and receive money for it
Buy	give done by other party in exchange for money
Exchange	give done by both parties
Return	give something back

In VIL, these differences are displayed in the icons.

3.2.2.5.2 Levin's English Verb Classes and their Alternations

The work by Beth Levin[122] consists of two parts. First she distinguishes verbs by their alternations in the expression of arguments, and describes several *diathesis alternations* that can be used to distinguish verbs. In the second part, she gives a

---

[122] **Levin, B.** *"English Verb Classes and Alternations: A Preliminary Investigation,"* The University of Chicago Press, 1993.

complete classification of the English verb system, which turns out to be of much help to our purpose.

3.2.2.5.2.1 Diathesis Alternations

The work is guided by the assumption that the behavior of a verb, particularly with respect to the expression and interpretation of its arguments, is to a large extent determined by its meaning.

Native speakers can make extremely subtle judgments concerning the occurrence of verbs with a range of possible combinations of arguments and adjuncts in various syntactic expressions. Speakers of English know which diathesis alternations verbs may participate in. In this way verbs could be distinguished by their diathesis alternations. Diathesis alternations are alternations in the expressions of arguments, sometimes accompanied by changes of meaning. She distinguishes three major groups of alternations, transitivity alternations, alternate expressions of arguments (mostly within the VP) that do not affect transitivity, and alternations that arise when verbs permit "oblique" subjects. Further, she has five other small groups of alternations.

Various aspects of the syntactic behavior of verbs are tied to their meaning. Verbs that fall into classes according to shared behavior would be expected to show shared meaning components. The verbs *break, cut, hit*, and *touch* are transitive, taking two arguments expressed as subject and object, but they have little else in common.

	touch	hit	cut	break
Conative	no	yes	yes	no
Body-part Possessor Ascension	yes	yes	yes	no
Middle	no	no	yes	yes

Not only can four verb classes be recognized that are defined by shared behavior of their members with respect to the above diathesis alternations, but their members share certain aspects of meaning as well. Thus their members have common syntactic as well as semantic properties.

Conative alternation: involves motion and contact components (I hit the door, I hit at the door).
Body-part possessor ascension alternation: involves contact (I touched Bill's shoulder, I touched Bill on the shoulder). For *break*, the notion of contact is not inherent to its meaning.
Middle alternation: causing a change of state (I broke the vase, Vases break easily). *Cut* and *break* refer to the result of the action, whereas *hit* and *touch* refer to the

action itself. *Cut*: by contact through motion, *break*: pure change of state. There are also the **Causative/Inchoative** alternations, which only apply to verbs that denote a pure change of state (The window broke vs. *The string cut). So, *touch* is a pure verb of contact, *hit* involves contact by motion, *cut* is a verb causing change of state by moving something in contact with the entity that changes, and *break* is a pure verb of change of state.

Middle and conative alternations fall under transitivity alternations (group 1 above), and body-part possessor ascension falls under alternations involving arguments within the VP (group 2 above). Levin states that notions of motion, contact, change of state, and causation must be taken into account in selecting a lexical representation of verb meaning, and that studies of diathesis alternations show that verbs fall into classes on the basis of shared components of meaning. The class members have in common a range of properties, including the possible expression and interpretation of their arguments, as well as the existence of certain morphologically related forms. The lexical knowledge of a speaker of a language must include knowledge of the meaning of individual verbs, the meaning components that determine the syntactic behavior of verbs, and the general principles that determine behavior from verb meaning.

"If the distinctive behavior of verb classes with respect to diathesis alternations arises from their meaning, any class of verbs whose members pattern together with respect to diathesis alternations should consist of members that share at least some aspects of meaning. Once such a class is identified, its members can be examined to isolate the meaning components they have in common. Distinctions induced by diathesis alternations help to provide insights into verb meaning, bringing out unexpected similarities and differences between verbs." [123]

Consider verbs of Motion. Verbs of Motion are cited as a large and important class within the English verb inventory. Yet a study of the syntactic behavior of these verbs inherently shows that this class is not homogeneous. It includes at least a subclass of verbs of inherently directed motion (*arrive, come, go*), and a subclass of verbs of manner of motion (*jump, run, trot, skip*). Levin states that, by providing independent criteria for isolating, narrow classes of verbs known to share certain aspects of meaning, the study of diathesis alternations can lead to the identification of the linguistically relevant meaning components, which determine a verb's behavior.

---

[123] **Levin, B.** *"English Verb Classes and Alternations: A Preliminary Investigation,"* The University of Chicago Press, 1993.

3.2.2.5.2.2 Classification distinctions

Verb classes arise, because a set of verbs with one or more shared meaning components show similar behavior. Some meaning components cut across the classes, attested by the existence of properties common to several verb classes. For instance, the meaning components contact and motion are common to the *hit* and *cut* verbs (Conative alternation). But contact alone would also have picked out the *touch* verbs, as well as *hit* and *cut* verbs. Thus, since most verbs are characterized by several meaning components, there is potential for *cross-classification.*

For Levin, the important theoretical construct is the notion of meaning component, not the notion of verb classes. The verb classes were chosen because their members participated in diathesis alternations or showed behavior that was closely related to that of other verbs found in particular alternations. Verbs belonging to the same class are syntactic "synonyms." They should be substitutable in the same set of syntactic frames, though not necessarily in exactly the same contexts. A representative verb is used to exemplify the characteristic properties of the class (remember from the speech act classification by Ballmer and Brennenstuhl, that the most general verb can be used to represent the category), including argument-taking properties, and behavior with respect to diathesis alternations. Some verbs have several meanings, and are be included in several classes. Often when a verb has several meanings, one of its meanings is basic and the others are systematically related to it, i.e. that are instances of extended meanings. In VIL this can not happen, because we are only representing meanings of verbs, and not verbs explicitly (too language specific).

Although Levin classifies verbs according to shared meaning, and then their shared behavior is looked at. Diathesis alternations can be used to predict or suggest the arguments/cases and their nature, that may or need to be filled in when a specific verb has been chosen. Most of the diathesis alternations are universally and apply to the meaning of a verb, but some are English specific and irrelevant for our purpose.

The verb classes are very useful. Although she narrowed them down to 43, some are extremely specific, such as Weather Verbs or Weekend Verbs. Furthermore, she distinguishes verb classes consisting of verbs that are zero-related to nouns, like Verbs of Instrument of Communication (cable, wire, telephone), and Verbs of Motion Using a Vehicle (all these are de-nominal verbs). Also, as Levin stated, there is little hierarchy above these 43 classes, i.e. no superclasses. If we can find some hierarchy among them, we have a potentially good classification for verbs. The next table shows all of Levin's verb classes with their potential superclasses.

Class, Verbs of:	Super Class
Putting	Physical transfer (PTRANS)
Removing	Physical transfer
Sending & Carrying	Physical transfer
Exerting Force	Physical transfer
Throwing	Physical transfer (PROPEL)
Change of Possession	Possession (ATRANS)
Hold and Keep Verbs	Possession
Concealment	Location
Contact by Impact	Location
Poke Verbs	Voluntary Activity
Contact: Touch Verbs	Voluntary Activity
Cutting	Destroy (Change of Material Integrity)
Combining & Attaching	Create
Separating and Disassembling	Destroy
Coloring	Create
Killing	Destroy
Emission	Involuntary Activity
Change of State	EXIST (Also, Changing Material Integrity, e.g. break)
Aspectual Verbs	EXIST
Measure Verbs	Identificational
Weather Verbs	Identificational
Image Creation Verbs	Create
Creation and Transformation	Create
Engender Verbs	Create
Calve Verbs	Create
Destroy Verbs	Destroy
Existence	Exist
Verbs of Appearance, Disappearance, and Occurrence	Create, Destroy, Exist
Lodge Verbs	Location
Learn Verbs	Communication (MTRANS)
Teach Verbs	Communication (MTRANS)
Predicative Complements	Identificational
Perception	ATTEND
Psych-Verbs	Psychological State of Mind
Desire	Psychological State of Mind
Judgment	Communication
Assessment	MBUILD
Searching	ATTEND
Social Interaction	Communication or Voluntary Act.
Communication	Communication (SPEAK)
Sounds Made by Animals	Involuntary Activity
Ingesting	INGEST
Verbs Involving the Body	Verbs Inv. Body Part
Grooming and Bodily Care	Verbs Inv. Body Part
Body-Internal Motion	Involuntary Activity

Assuming a Position	MOVE w.o. Changing Location

Table 3.2: Levin's verb classes classified with potential super classes.

3.2.2.5.3 Schank's Primitive ACTs

Schank[124] distinguishes eleven primitive ACTs, which are abstract nodes that describe the meaning of verbs. The set of primitive actions should be universal; independent of any language, conforming more to what people do rather than the language in which they describe their actions. Consider again the examples of the Eskimo and his ten verbs of the way to throw snow, where we only have one. The basic ACT of throwing snow is the same, and all these ten versions are based on this one basic ACT, which is the same one all throwing verbs use (PROPEL).

In Schank's conceptual dependency theory, all verbs are broken down into their basic conceptual elements, often in terms of one or more of these eleven ACTs, and an ACT only refers to what the actor has actually done and is treated apart from the possible consequences of this action. His notions of conceptual actions correspond to the way he imagines people conceive of their own actions. Any action that he deposits must be an actual action that can be performed on some object by an actor. Nothing else qualifies as an action. Consequently, all actors are animate. If an inanimate is stated as being the actor in a sentence, it is still not regarded in that way in a Conceptual Dependency representation.

Conceptual Dependency theory states that often the syntactic structure of a language is misleading as to the underlying concept of utterances expressed in that language. Conceptual structures have as their core underlying primitive actions. It is often the case that verbs, in a language, express those actions. However, this is not always the case. Many verbs in English often express final states of objects and leave the action involved as an inference. As an example, consider the following verbs that specifically reference states conceptually and only infer actions: hurt, like, hate, upset, prevent. Using a Conceptual Dependency framework for meaning representation, Schank states that the total number of ACTs necessary to adequately represent any natural language sentence is only eleven.

3.2.2.5.3.1 Eleven Primitive ACTs

A physical ACT is something a PP (picture producer) can do to another PP. There are five ACTs that describe physical actions people can perform:

---

[124] **Schank, R.** *"Conceptual Information Processing,"* North-Holland, New York, 1975.

PROPEL   MOVE   INGEST   EXPEL   GRASP

Associated with each ACT is conceptual semantics: what kinds of objects can be used with any given ACT, and a set of conceptual cases (Objective, Directive, Recipient, Instrumental). Each ACT requires a specific number of cases (either two or three). The Instrument case is often predicted (for example in "go", I: moving one's feet, or in "buy" or "sell", I: money).

PROPEL: "to apply force to". Takes Objective (physical object) and Directive case.

MOVE   :   "to move a body part". The only possible objects are body parts. Also Directive case.

INGEST: "to take inside the body". Takes Objective (ingestible object) and Directive case.

EXPEL :   "to force something out of animate object". Takes Objective (previously ingested object) and Directive case.

GRASP :   "to physically grasp an object". Takes Objective (grasped object) and Directive case.

Often in speaking people tend to focus more on the result of an action than on the action itself. Since Schank is trying to describe a model of human thought as opposed to some notion of absolute truth, he distinguishes two ACTs to account for this focus. They have no real world correlates except for the state changes they cause, but nonetheless are real in the sense that they help account for what people talk about. They are:

PTRANS :   "to change the location of something". Physical Object, Directive and Instrumental case.

ATRANS: "to change some abstract relationship with respect to an object". The object is a combination of a physical object and an abstract relationship that that physical object has with some animate object. The animate object is indicated in the Recipient case.

Consider the verbs "give" and "take" in "I took a book from Mary" and "Mary gave me a book". They are similar, and Schank treats the ACT underlying them as the same, namely ATRANS. However "to take an aspirin" would be considered to be INGEST.

There are two ACTs that are mostly used as the Instruments of some other ACT, usually MTRANS. For example, to "see" or "read" is MTRANS of information by ATTENDing one's eye in the direction of the information. The two ACTs are:

SPEAK : "to produce a sound". Object is type of sound. Also present is the Directive case.

ATTEND: "to direct a sense organ or focus organ towards a particular stimulus". Takes Objective (sense organ) and Directive case.

There are two mental ACTs, i.e. operations a PP can perform on an idea. For this Schank first divides memory up into three pieces that we perceive that humans talk about: The Conceptual Processor (CP), the Intermediate Memory (IM), and the Long Term Memory (LTM). CP is where all of the conscious thought takes place (it admits only one item at a time). IM is a holding place for all items that are currently being used ("have on one's mind", "assume for now"). LTM contains all info that is known by a person. Schank has no piece for Short Term Memory (STM), because he says people do not talk about it.

MTRANS: "to transfer information". Objects are always conceptualizations. Recipient is LTM. It is used in nearly all mental verbs, i.e. whenever info is moved about within one person or between people.

MBUILD : "to create or combine thoughts". Objects are conceptualizations and new concepts are created from them. Recipient is CP, i.e. whenever a new thought is generated, it is thought of.

For example, in "John believes what Mary told him", "John believes" is MTRANS is and "Mary told him" MLOC(LTM(John)). In "John decided to leave the house", "decided" is MBUILD and Recipient is LTM(John) to CP(John).

3.2.2.5.3.2 States

There are also states: "I am happy", "I am shocked". In the sentence "John hurt Jim", the physical state of Jim decreases because of something John did. Schank uses a dummy ACT for this: DO. States are represented as scales. An ACT is something that is done by an actor to an object. All verbs that leave out the actual ACT that was done must be treated as DOs with causals connected to a change of state. States of objects are described by scales with numerical values. Not that humans represent states in this way, but they can detect differences between adjectives that these scales suggest. So "angry" is just a little less that "furious". The scales are:

HEALTH	-10 to 10	dead to perfect health
FEAR	-10 to 0	terrified to calm
ANGER	-10 to 0	furious to calm
MENTAL STATE	-10 to 10	catatonic to ecstatic

PHYSICAL STATE	-10 to 10	dead to okay
CONSCIOUSNESS	0 to 10	unconscious to higher drug consciousness (>10)
HUNGER	-10 to 10	starving to satiated
DISGUST	-10 to 10	nauseated to bothered
SURPRISE	0 to 10	surprised to astounded

He points out that this list is not complete, e.g. certain drives (like sexual) might be useful. Some words are combinations of scales, e.g. shocked = SURPRISE (6) and DISGUST (-5). Other states exist that are not scales, but are more absolute measures (LENGTH, COLOR, LIGHT INTENSITY, MASS, and SPEED) or are relationships between objects:

CONTROL
PART (inalienable possession)
POSS (possession)
OWNERSHIP
CONTAIN
PROXIMITY
LOCATION
PHYSICAL CONTACT (the last two are special forms of proximity)

For our purpose, eleven ACT's are not enough, since they only refer to actions. There are also events and states that need to be considered. Schank has one dummy ACT, called DO, to deal with these, which is too specific for our purpose. It would be too specific to let all these state changes fall under one dummy ACT or even a class. Also, we do not want verbs to be represented by a combination of primitive ACTs. The problem for our purpose is that almost all verbs that we want to express consist of more than one primitive ACT. Consider the sentence "John strangled Mary". We could not choose an ACT for "strangle", but have to pick the ACTs GRASP (her neck) and INGEST (the air she cannot anymore). So it is obvious that the primitive ACTs do not stand for categories that are used to divide up all possible verbs, and underneath have subclasses of verbs that are all GRASP verbs for example. No, the primitive ACTs are rather the most primitive elements that can be used in combination with other primitive ACTs and cases (e.g. Instruments) to form the verbs we want to express. Not a desirable state of affairs for our purpose. However, many of the eleven ACTs do have a place in our classification, but instead they are rather used to group verbs, i.e. as shared meaning components/categories. Also, there will be a big part for states.

3.2.2.5.4 The Verb System in Basic English

As stated in chapter one, Basic English is a restricted version of the English language. It is designed to be minimal, but enough to communicate with. It was supposed to be a universal language that would be easy to learn for foreigners. It consists of only 850 basic words. The verb system of Basic English consists of only eighteen verbs:

Come (*)	Get (*)	Give (*)
Go (*)	Keep (*)	Let (*)
Make (*)	Put (*)	Seem
Take (*)	Be	Do (*)
Have	Say	See
Send	May	Will

Ten of these, marked by a star, are put under the name of "operators", and are used in many ways to reduce the length of the basic wordlist. So it is no longer "planting seeds", but "to put seeds in the earth". There are also many words that can be used in verb-like functions (living, waiting, married). There is no verb for "live" and "wait", and the progressive word is put under words that describe "qualities". For other verbs, its noun is taken plus one of the ten operators: care, change, play, etc. These words are put under "things".

The use of phrases instead of verbs greatly reduces the number of separate words, which a foreigner must learn. These phrases are often extremely idiomatic, and have to be learned as though they were compound words. Although it sounds appealing that nouns can be used in a verb-like manner (adding -ed, or -ing), you are basically converting them into verbs.

For our system, we are trying to classify the meaning of verbs, so we need to classify more than those eighteen verbs. We would also need to classify a lot of these nouns that are converted into verbs. Furthermore, Basic English uses a lot of verb-plus-preposition-adverb sentences (up, off, out, etc.). For example, "to put up", and "to put off" are different in meaning and cannot be classified under "to put". They are essentially different verbs. The eighteen verbs are too little to communicate with iconically. One would for example have to pick "to put" to express "to eat" (because it would be represented as "to put food into one's mouth"), but also to express "to kill" (because it would be represented as "to put to death"). Too many verbs with different meaning would fall under the same elementary verb. They are not abstract meanings of verbs (like "a transfer of possession"), but are the most basic verbs that can be used to describe other verbs with.

3.2.2.5.5 Matsukawa and Yokota's Development of a Concept Dictionary

The goal of Matsukawa and Yokota's work[125] was to build a neutral dictionary for semantic processing of natural language that can be utilized for various application systems. It consists of concept descriptions and a concept classification. The descriptions and classification have been obtained based on the results of an analysis of sentences in the EDR corpus (an on-line corpus consisting of 1 million practical example sentences that are analyzed lexically, syntactically and semantically) and outputs from various natural language processing systems. For now we are only interested in the non-object part of their classification.

3.2.2.5.5.1 Non-Object Categories

The classification of verbal (non-object) concepts was not made by using the hierarchy as a discrimination tree, but by using semantic association from the meanings of the concepts and examples of deep case patterns of the concepts. There are three levels. The highest is "coarse semantic clusters", the second is "fine semantic clusters". They have both been divided based on semantic association. The third level has been divided based on the deep case pattern shared by concepts, the non-object-categories. They have found 14 coarse semantic clusters, 253 fine semantic clusters, and 984 non-object-categories. The 14 coarse semantic clusters are displayed in the following table. Note that, since the classification is for non-objects, not all classes (semantic clusters) are about verbs. Some are about relations that hold between classes and objects as well.

Class	Description	Equivalents
Spatial Relations	relations among physical objects meaning states and changes in space	GRASP, PTRANS, PROPEL
Spatial Attributes	Attr. of physical objects meaning spatial measures in space	State
Social Relations	social relations among persons	State
Class Relations	inclusive relations and comparative relations among objects	State
Possession	relations among possessors and possessions	A(TRANS)
Information	relations among information and information processors	SPEAK, ATTEND, MENTAL
Estimation	relations and attr. of objects meaning states and changes of their estimations	State

---

[125] **Matsukawa, T.** and **Yokota, E.** in Pustejovsky, J. and Bergler, S. (Eds.) *"Lexical Semantics and Knowledge Representation,"* Proc. of the First SIGLEX Workshop, Berkeley, CA, June 17, 1991, Springer-Verlag, Chapter 21 "Development of the Concept Dictionary -- Implementation of Lexical Knowledge".

Possibility	relations and attr. of events meaning states and changes of possibilities	State (scale)
Function	relations and attr. of objects meaning states and changes of their functions	State
Progress	relations and attr. of events meaning degrees of actualization	
Time	relations and attr. of events meaning temporal order or distance	
Quantity	relations and attr. of objects meaning quantity or degree	Scale
Other Attributes	attr. other than those above	
Existence	relations meaning appearance, continuance or disappearance of existence	EXIST

Table 3.3: Matsukawa and Yokota's non-object categories.

Some examples of fine semantic clusters for "Spatial Relations" are:

> For_a_Physical_Object_itself_to_Change_in_Space
> To_Touch_a_Place_or_Physical_Objects_in_Space
> To_Move_Some_Distance_to_a_Direction_in_Space
> For_an_Angle_to_Decrease_in_Space
> For_an_Angle_to_Increase_in_Space

Some examples of non-object-categories under
To_Touch_a_Place_or_Physical_Objects_in_Space are:

> (For_an_Animal)_To_Touch_(A_Physical_Object)_Intentionally
> (For_an_Animal)_To_Touch_(A_Physical_Object)_With_a_Part_of_the_Bod
> y

The deep case pattern for the latter would be:

> [agent: Animals, object: Physical_Objects, implement: Parts_of_Animals]
> c#push(agent: "a person", object: "a button", implement: with "a finger")
> c#kick(agent: "a person", object: on "a can", implement: with "a foot")

3.2.2.5.6 Jackendoff's Non-Spatial Semantic Fields and the Thematic Relations Hypothesis

Jackendoff's Thematic Relations Hypothesis[126] states that the semantics of motion and location provide the key to a wide range of further semantic fields. In any semantic field of events and states, the principal event-, state-, path-, and place-functions are a subset of those used for the analysis of spatial location and motion. The fields differ in only three possible ways:

1) what sorts of entities may appear as theme;
2) what sorts of entities may appear as reference objects;
3) what kind of relation assumes the role played by location in the field of spatial
   expression.

The claim behind this is that the mind does not manufacture abstract concepts out of thin air. It adapts machinery that is already available. So, the claim is basically that the mind adapts what it knows about physical space, and uses it in other semantic fields. He distinguishes the following semantic fields, whose verbs appear in patterns parallel to those of spatial verbs:

Temporal field:
The game is at nine o'clock. (BE)
They moved their vacation from June to July. (GO)
Despite the weather, we kept the game at nine o'clock.
(STAY)

Possession field:
She lost her scarf. (GO...FROM)
She received a new scarf. (GO....TO)
She gave the scarf to Mary. (GO....FROM.....TO)
She kept the scarf. (STAY....AT)

Identificational field:
John is a basketball player. (BE....AT)
Larissa became a mother. (GO....TO)
The ice melted. (GO....FROM.....TO)
The balloon became small. (GO....TO)
The balloon became smaller (GO....TOWARD)

Circumstantial field:
Jim kept watching television. (STAY....AT)
Eddy started watching television.   (GO....TO)
Larissa is watching television.   (BE....AT)
I forced her to watch television. (GO....TO)
I urged her to watch television. (GO....TOWARD)

---

[126] **Jackendoff, R.** *"Semantics and Cognition",* Chapter 10 "Nonspatial Semantic Fields and the Thematic Relations Hypothesis," The MIT Press, 1983.

Existential field: I live. (BE....AT)
         Cliff died of a heart disease. (GO....FROM....TO)

Jackendoff did not analyze all semantic fields. One big field he did not do is the communication or mental field (e.g. verbs like "to say", "to see"). It turns out that these can also be paralleled to the spatial field. For example, remember MTRANS is a "transfer" and consequently fits the hypothesis. For sure, for any category that involves a transfer (ATRANS, PTRANS, MTRANS) the hypothesis holds. For other categories, like "ATTEND" and "SPEAK", one could also imagine a flow of information from or to the source. So, if we were to extend the semantic fields by this, we would end up with table 3.4, which captures most of the semantic abstractions we want to represent:

Semantic Field of Events and States	Equivalents in Other Approaches
Physical (i.e. Spatial Location and Motion)	PTRANS, PROPEL, MOVE
Temporal	Time (Matsukawa)
Alienable Possession	A(TRANS)
Identificational	DO (state), Equational
Circumstantial	Progress (Mats.), Process (Ballmer), is a modifier
Existential	EXIST (Ballmer, Matsukawa, Levin)
Mental	M-, Communication

Table 3.4: Non-spatial semantic fields

The basic primitives used here are GO, BE, STAY for verbs, and the spatial AT, FROM....TO, FROM, TO, and TOWARD. The difference between TO and TOWARD is that the first indicates that a path has been traversed and the goal has been reached, whereas the latter only says that movement is towards a goal, and not whether or not the goal has been reached. Since the spatial field is the most basic one, people tend to extend it to the other fields. It is also strongly supported by non-linguistic cognition; it is the common ground for the essential faculties of vision, touch, and action. From an evolutionary perspective, spatial organization had to exist long before language. Because the spatial field has an easily understood visual representation, they could be used primitive combining features (visual elements) in a visual system. They could be part of an icon, and be represented like:

     AT   ---
     TO   --->|
     TOWARD   - - ->
     FROM |--->
     FROM...TO |--->|

Then, for example, for alienable possession, |---> in an icon might mean "to lose something", |--->| might mean "to give to", and --->| might mean "to receive".

3.2.2.5.7 VIL's Verb Classification

Most researched work uses shared meaning and at some point case predictive features to classify verbs. After reviewing all the previous approaches, we have identified some semantic fields that are common to all approaches. They are Physical (i.e. Spatial Location and Motion), Communication/Mental, Alienable Possession, Existential, and Identificational. These are all verbs of state or change of state. Here the focus is on the result of the sentence. The other primary grouping we propose is Activity, where we distinguish Voluntary and Involuntary Activities, but also Activities by Non-Humans (brooks babbling, sun shining, computer humming, bees humming). These are all activities where the focus is not on the result or goal, but on the activity. In our organization, presented below, one can focus on the result of an activity (a state or a state change) or the activity itself.

In the organization, opposites of classes, like Verbs of Putting vs. Verbs of Removing, will be subclasses of the same class. The spatial primitive combining features will not be part of the hierarchy, but they can be part of the icons. They may also be used to present the user with a sentence frame (cases) once he/she has chosen the verb. This means that if the user chooses "to give to", then not only can |--->| be part of the icon, but also the frame presented will have a place for a source and a goal (and also an object to be transferred). The verbs are not only grouped according to their shared meaning, but also according to their case predictive features.

State & State Changes (focus is on the result of the activity):
> Mental
> Alienable Possession
> Physical Transfer and Location
> Existence
> Identificational

Activities (focus is on the activity itself):
> Involuntary Activity
> Voluntary Activity or Verbs Involving the Body w/o Changing Location

In verbs of state and state change the focus is on the result of the activity, and in verbs of activity the focus is on the activity itself. These two top-level categories will not be presented to the user since they are too specific. The categories are subdivided as follows:

1) <u>Mental</u>:
        Communication Verbs (SPEAK, MTRANS)
        Verbs of Building a Mental Representation (MBUILD)
        Verbs of Psychological State of Mind (Psych-Verbs, Desire Verbs)
        ATTEND Verbs (Perception, Search)

2) <u>Alienable Possession</u>:
        ATRANS Verbs
        Hold and Keep Verbs

3) <u>Physical Location & Motion</u>:
        Verbs involving a Physical Transfer (PTRANS)
        PROPEL verbs
        Verbs of Sending and Carrying
        Verbs of Removing and Erasing
        Put and Pour Verbs
        Lodge Verbs
        Verbs of Exerting a Force
        Verbs of Contact by Impact
        Conceal Verbs

4) <u>Existence</u>:
        Create Verbs
                Transform Verbs
                Reveal Verbs
                Record Verbs
                Present Verbs
                Verbs of Creation
                Connect Verbs
                Calve Verbs
        EXIST Verbs
                Existence Verbs
                Aspectual Verbs
        Destroy Verbs
                Kill Verbs
                Disconnect Verbs
                Disappear Verbs
                Destruction/Break Verbs
                Cut Verbs

5) <u>Identificational</u>
     Verbs of Predicative Complement (e.g. Appoint Verbs)
     Weather Verbs
     Measure Verbs
     Identity Verbs (e.g. be, seem, identify, happen)

6) <u>Involuntary Activities</u>:
     By Non-Humans (like humming of a bee)
     By Humans (like body internal motion)
     Sound Emission Verbs
     Light Emission Verbs
     Modes of Being Verbs (e.g. flow, bloom, corrode)

7) <u>Voluntary Activities or Verbs Involving the Body w/o Changing Location</u>:
     Verbs Involving a Body Part: MOVE w/o Changing Location Verbs
     Touch and Poke Verbs
     INGEST Verbs
     EXPEL Verbs
     Sports Verbs
     Hold and Keep Verbs
     Verbs of Exerting a Force

As can be seen in table 3.2 in section 3.2.2.5.4.3, a lot of verb classes that Levin proposed have been used as subclasses in this hierarchy. Further, notice that five of the seven semantic fields from table 3.4 in section 3.2.2.5.8.1 are represented in our hierarchy. This also means that most of Levin's classes can be mapped to the semantic fields of Jackendoff.

For now, only the categories of the hierarchy are presented. In the next chapter all verbs (the terminals) used in our domain of cooking, eating and drinking will be discussed. Also, icons for various categories and terminals of the in this chapter mentioned classifications are shown. We are now ready to give VIL's grammar.

3.3 VIL's Grammar

VIL's grammar is based on the subcategorization of verbs presented in section 3.2.2.5. As in case grammar, a sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship. Although there can be compound instances of a single case (through noun phrase conjunction), each case relationship occurs only once in a simple

sentence. The verbal elements of the sentence are the major source of the structure: the main verb in the proposition is the focus around which the other phrases, or cases, revolve.

Before we give the grammar, let us give some notation conventions that will be used throughout the grammar.

*: 0 or more instances
+: 1 or more instances
?: 0 or 1 instance
(): group subexpression
|: alternation (or)
{}: has terminals as its members
xx_NP: is an NP, but xx says more about the nature/case of the NP (e.g. agent_NP, loc_NP).

Top Level Sentence:

A sentence consists of a propositional kernel, a time when the event took place (0 or 1 instance), and a location where it took place (0 or 1 instance). A sentence can also be a compound sentence.

S $\rightarrow$ **(({declarative} | {command} | {question}) TIME-WHEN? PLACE-WHERE? PROP_KERNEL) | COMPOUND/PLEX_S**

Even though the rules in the grammar are expressed in an order, VIL is order independent, since not all languages have the same order (e.g. SVO, i.e. Subject Verb Object, or SOV). See Principle 3.3: No Linear Word Order.

A compound or complex sentence are two or more sentences connected together either by a coordinator ("and", "or") or by a subordinator ("although", "because"):

- **COMPOUND/PLEX_S** $\rightarrow$ **(({subordinator} S) S) | (S {coordinator} S)**

Declarative, command, and question determine the communicative intend of the message. A message can be a regular declarative/assertive sentence, a command, or a question:

- declarative $\rightarrow$ "."
- command $\rightarrow$ "!"
- question $\rightarrow$ "?"

- **TIME-WHEN** $\rightarrow$ {temporalizer}**? (time_when_S** | {time_pronoun} | **(**{time_demonstrative}{time_unit}**) | TIME_EXPR)**

1. If a temporalizer is not expressed, its meaning is implicit in and depends on the sentence. Something might happen *on* Friday or *at* four o'clock or *in* the summer. In all cases it is not necessary to specify a temporalizer. The user infers the temporalizer for himself/herself correctly.
2. **time_when_S**$\rightarrow$ **S** (e.g. something happened "while I was eating")
3. **TIME_EXPR** $\rightarrow$ {year-spec}? {season}? {month-name}? {date-of-month}? {day-of-week}? {time-of-day}? {holiday}?

- **PLACE-WHERE** $\rightarrow$ {localizer}**? (place_where_S** | {loc-pronoun} | **MAP** | **loc_NP)**

1. If a localizer is not expressed, it has a meaning that is expressed in English as "in", e.g. "in France" or "in the kitchen".
2. **MAP** means that the user chooses a location from a world map.
3. **place_where_S** $\rightarrow$ **S**
4. **loc_NP** $\rightarrow$ **NP** (e.g. under the table)

Noun Phrases:
A noun phrase can consist of a single noun phrase or multiple noun phrases connected by connectives.

- **NP** $\rightarrow$ **NP (**{connective} **NP+)?**
- **NP** $\rightarrow$ {nom-pronoun}
- **NP** $\rightarrow$ {nom-demonstrative}**? QUANTITY? ADJ-UNIT*** {noun} **owner_NP? COMPARATIVE_NP ? S?**

1. **ADJ-UNIT** $\rightarrow$ **(**{intensifier} | **COMPARATIVE_NP)?** {adj/stativeV}
2. **COMPARATIVE_NP** $\rightarrow$ **(**"more" | "less"**) than_NP**
3. **QUANTITY** $\rightarrow$ {quantifier} | {number}
4. **than_NP** $\rightarrow$ **NP**
5. **owner_NP**$\rightarrow$ **NP** (e.g. "my dog")

An example of comparative NP's within a noun phrases are:

- "I eat more than Jim apples", i.e. "I eat more apples than Jim"
- "I eat more than limes green apples" or "I eat greener than limes apples", i.e. The apples I eat are greener than limes."

In the first sentence, the comparative NP pertains to the noun in the noun phrase. In the second sentence, the comparative NP pertains to one of the adjectives to the noun in the noun phrase, namely "green".

Specifying the proposition kernel:

- **PROP_KERNEL** $\rightarrow$ {aux-modal}**?** {Aktionsarten}**? VERB-KERNEL FREQ-OF-OCCUR? DURATION? ADV-UNIT? use_NP?**

1. **ADV-UNIT** $\rightarrow$ **(**{intensifier} **| COMPARATIVE)? MANNER-ADV**
   i.e. in what manner does the event take place, e.g. "very fast" and "faster than Jim"
2. **FREQ_OF_OCCUR** $\rightarrow$ **(**{quantity} "times" **(**"per" {time-unit}**)?) |** {adv-of-freq}
   i.e. how often does the event take place, e.g. "3 times a day" and "sometimes"
3. **DURATION** $\rightarrow$ **(**{quantity} {time-unit}**) |** "too long" **|** "too short"
   i.e. for how long does the event take place, e.g. "for 3 hours" and "for a long period of time". Duration is also presented by a slider with at the two extremes "a short period of time" and "a long period of time".
4. **MANNER-ADV** $\rightarrow$ **(**{adj/stativeV}"ly"**) |** {adv-of-manner} **| phrase**
   the first one is de-adjectival, e.g. "smoothly"
5. **use_NP** $\rightarrow$ **NP** , specifies the instrument case.
   i.e. what is used in the activity, e.g. "with a hammer"

The verb-kernel:

- **VERB-KERNEL** $\rightarrow$ **STATE_V_K | ACTIVITY_V_K**

This reflects our hierarchy of the verb system. As we stated, the top level division we made is a distinction between verbs that focus on the result of an action (the state and state-change verbs), and verbs that focus on the activity itself (activity verbs). For each class of verb we try to use its case predictive features to indicate what to potential arguments/cases are that the verb takes.

Verb kernel for verbs where the focus is on the activity itself

- **ACTIVITY_V_K** $\rightarrow$ **agent_NP theme/obj_NP?** {verb}

1. **agent_NP** $\rightarrow$ **NP**
2. **theme/obj_NP** $\rightarrow$ **NP**

Examples are "I ate an apple" and "I touch your arm", where "I" is the agent and "apple" and "arm" are the objects. The next figure illustrates the cases in the sentence "I ate an apple":

Verb kernel for verbs where the focus is on the result of the activity:
Verbs with focus on the result of the activity are divided into five top-level categories (see section 3.2.2.5). For each class of verb we try to use its case predictive features to indicate what to potential arguments/cases are that the verb takes. The five top-level categories are:

1. Verbs involving a physical transfer & location verbs (**PHYS_V_K**)
2. Alienable possession verbs (**POSS _V_K**)
3. Mental verbs (**MIND _V_K**)
4. Existence verbs (**EXIST _V_K**)
5. Identificational verbs (**IDENT _V_K**)

- **STATE_V_K** $\to$ **PHYS_V_K | POSS_V_K | MIND_V_K | EXIST_V_K | IDENT_V_K**

Physical transfer and location verbs:
- **PHYS_V_K** $\to$ **theme/obj_NP?** {phys_verb} **((source_NP? goal_NP?) | at_NP?) agent_NP?**

1. **goal_NP** $\to$ **at_NP**
2. **source_NP** $\to$ **at_NP**
3. **at_NP** $\to$ **NP**

Examples are:
  "I went from school to my house" (agent: I; source: school; goal: house)
  "I throw you the ball" (agent: I; goal: you; theme/obj: ball)
  "I stayed home" (agent: I; at: home)

The next figure illustrates the cases in the sentence "I went from school to my house":
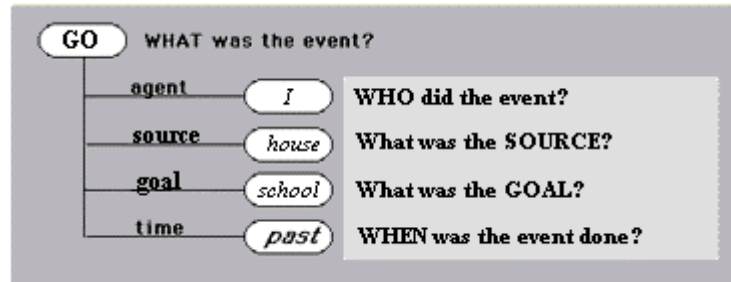
Alienable possession verbs:
- **POSS_V_K → theme/obj_NP ({possess-verb} at_NP) | ({atrans-verb} source/agent_NP? goal/agent_NP?)**

Examples are:

    "I give you an apple" (source/agent: I; goal: you; theme/obj: apple)
    "I took the apple from you" (source: you; goal/agent: I; theme/obj: apple)
    "I kept the letter" (at: I; theme/obj: letter)

In the first example the apple goes from me to you. I am not only the source but also the actor. In the second example the apple goes from you to me. I am not only the goal, but also the actor. In the last example, the letter stays with me, i.e. at my location.

The next figure illustrates the cases in the sentence "I give you an apple":
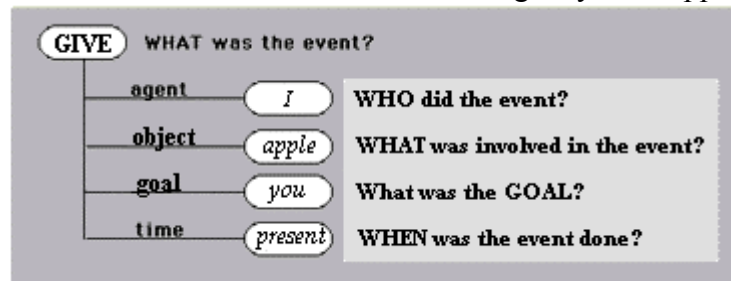
Mental verbs:
- **MIND_V_K → (theme/obj_NP | S) ({know-verb} at_NP) | ({mtrans-verb} source/agent_NP? goal/agent_NP?)**

Examples are:

    "I teach the children the material" (source/agent: I; goal: children; theme/obj: material)

"I know the material" (at: I; theme/obj: material)
"I ask you to eat your dinner" (agent: I; S: to eat your dinner)

In the first example I teach material from me to the children (a mental transfer). I am not only the source but also the actor. In the second example, the material stays with me, i.e. at my location.

Existence verbs:
- **EXIST_V_K** $\rightarrow$ **theme/obj_NP** {exist-verb} **agent_NP? source_NP? goal_NP?**

Examples are:
> "I sliced the bread" (agent: I; theme/obj: bread)
> "I made a statue from clay" (agent: I; source: clay; theme/obj: statue)
> "I boiled water in a pan" (agent: I; goal: pan; theme/obj: water)

In the last example the pan is the goal. For the goal, one can ask the question "to whom, to where". For the goal, one can ask the question "from what, from where".

The next figure illustrates the sentence "In the kitchen, I bake you a cake from apples 3 times a day". In this sentence, also the frequency of occurrence and place where cases are filled:
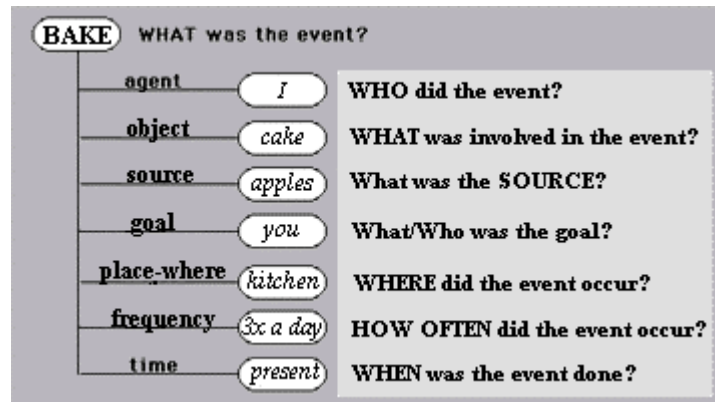
Identificational verbs:
- **IDENT_V_K** $\rightarrow$ **theme/obj_NP** {ident-verb} **(at_NP | {adj/stativeV})** **agent_NP?**

e.g. is, serves as, appoint/elect
Examples are:
> "They elected him president" (agent: They; at: president; theme/obj: him)
> "He is a mad" (agent: He; adj: mad)

In the second example "mad" is really an adjective or a noun modifying property that says something about the agent.

The terminals:

- {temporalizer} → *before | while | as soon as/until | after*
- {time_demonstrative} → *last | this | next*
- {time-unit} → *millenium | century | year | season | month | week | day | hour | minute | second | millisecond*
- {time-pronoun} → *past | future | when | long ago | just a moment ago | soon*
- {localizer} → *on | under | in(side) | outside | in front of | next to | in front of | on the left of | on the right of | north | south | east | west | northeast | northwest | southwest | southeast*
- {loc-pronoun} → *here | there | where*
- {nom-pronoun} → *I | you | he | she | it | you-all | they | who*
- {nom-demonstrative} → *this | that | which*
- {quantifier} → *all/every/whole | many | some | few | how_many/much*
- {number} → *0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 100 | 1000*
- {intensifier} → *very | most | least*
- {aux-modal} → *should | must | can*
- {Aktionsarten} → *start | stop | continue*
- {adj/stativeV}
- {noun}
- {verb}
- {adv-of-freq} → *never | sometimes | often | always*
- {subordinator} → *{temporalizer} | if (NEG) | only if (NEG) | because | in order (NEG) to | although/even if*
- {coordinator} → *and | or*
- declarative → *"."*
- command → *"!"*
- question → *"?"*

The terminals for noun, verb, and adjective are empty since they are organized hierarchically as trees.

4. Visual Representation of the Language

In this chapter the focus is on the visual representation of the system. Section 4.1 will research icon design and guidelines. Section 4.2 is about designing cross-cultural icons, and gives some dos and don'ts. In Section 4.3, we describe the icon language for VIL. First we determine what needs to be represented, then in section 4.3.2 we give principles for designing icons for visual elements and semantic roles, and in section 4.3.3 we give some rules for designing icons for the grammatical categories of verbs, nouns, and adjectives. Then, in section 4.4, we discuss the domain of cooking, eating and drinking, which is the domain that has been implemented for the first release of VIL. We also show the icons that have been designed for this domain.

4.1 Icon Design

Good graphic design is much more than just pretty colors and pictures. It can significantly improve the communicative value of the interface, which leads to increased usability (ease of learning, efficiency of use, memorability, reduced number of errors, and subjective satisfaction). Representation provides the basis for all communication. We can convey ideas about things that are not materially in our presence only by calling forth an appropriate mental representation. Icons are of central importance to the successful utilization and acceptance of a graphical user interface and its related applications. Icons can lead to faster decision making[127] and improve long-term memorability[128]. Furthermore, good icon design becomes more and more important nowadays, because icons are now commonly used within GUI's (Gittins, 1986; Horton 1994; Rogers, 1989; Sassoon and Gaur, 1997) and on the World Wide Web (Honeywill, 1999). Cross-cultural elements must be taken into account, though. For example, a system may represent the concept of a "queue" by an icon of a billiard ball (i.e. cue ball). Such visual representations might be fun to put in a design, but they are often harmful to the novice user trying to make sense of a new visual environment. Also, the product would be dead on arrival if it were ever exported to a non-English speaking country. Understanding what a thing *represents* (a queue) - as opposed to merely what it *depicts* (a billiard ball)- is a prerequisite for

---

[127] **Spence, R & Parr, M.** *"Cognitive assessment of alternatives,"* Interacting with Computers 3, No. 3, pp. 270-282, 1991.

[128] **Rogers, Y.** *"Icons at the interface: Their usefulness,"* Interacting with Computers 1, No. 1, pp. 105-117, 1989.

using the icon correctly. Users must recognize the icon on the display screen as a sign for the thing (e.g. an icon for a calculator program), rather than the thing itself.[129]

The next figure shows a simple model of icon processing [Barker and van Schaik, 2000], in which an icon can be processed in two ways. The simplest way is when a user directly associates the icon with the function it represents (least cognitive processing). Extra cognitive processing is involved when the user recognizes one or more objects that are embedded within the icon and he/she is able to deduce the function that the icon represents.[130]
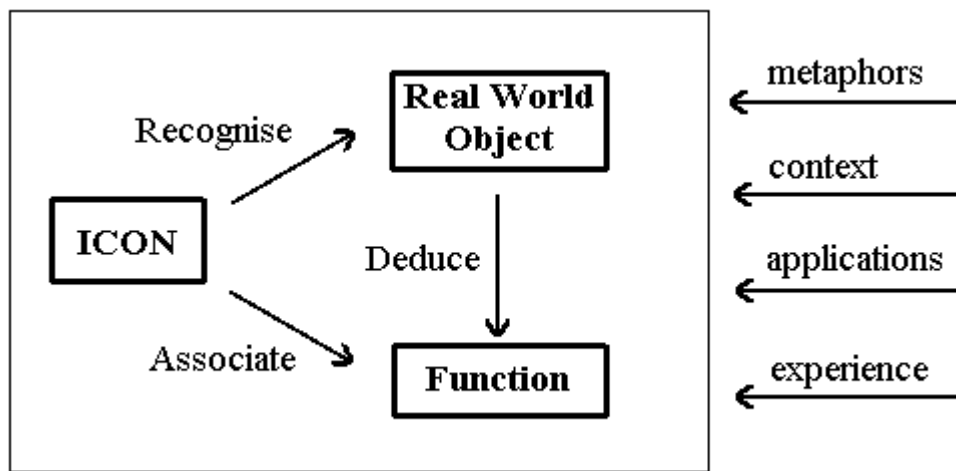
Barker and van Schaik state that icons never appear in isolation and that additional factors, which are often present, can influence the cognitive processing of an icon (e.g. metaphors, the context in which an icon appears, the application within which the icon is embedded, and experience).

Interpretation is the process of reconstructing the meaning of a sign by identifying the sign object and grasping the significance of the connection between the sign object and representamen. As shown in figure 22, the interpretation becomes easier up to a point as the representation becomes more schematic. As the level of abstraction increases, the sign becomes progressively more generic and less complex.

---

[129] **Nadin, M.** *"Interface Design and Evaluation-Semiotic Implications,"* in Hartson, R. and Hix, D. (eds.), Advances in Human-Computer Interaction, Volume 2, 1989, pp. 45-100.

[130] **Barker, P.** and **Schaik, P. van** *"Designing and Evaluating Icons,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.
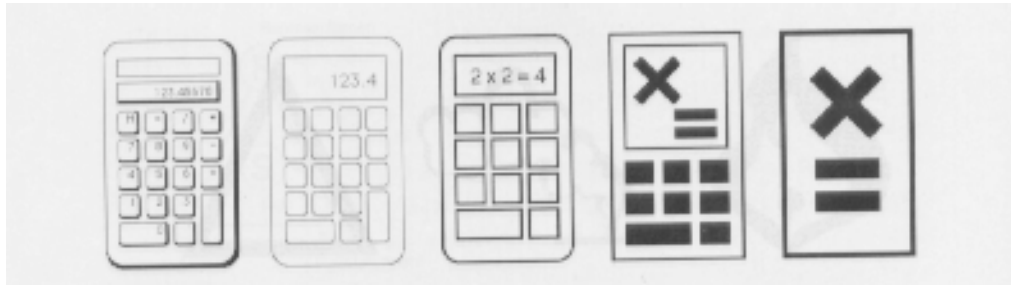
Figure 22: The ease of interpretation varies with the level of abstraction of the sign represented. As highly concrete, realistic representations are simplified, they become easier to interpret - up to a point beyond which further abstraction begins to obscure the sign's semantics (Nadin, 1989).

A given sign representation can be characterized by its degree of *abstraction* - the extent to which the *essential* qualities upon which the representation is based are isolated from the literal perceptual characteristics of the sign object. A photograph or realistic illustration provides a high degree of fidelity to a particular sign object, and is usually easy to recognize as a result. More schematic representations, because they permit the selective omission of detail, are better able to represent a broader class of objects (as opposed to one of its instances) or to focus on some characteristic aspect of the object (rather than on the perceptual reality of a specific individual).

No matter how important icons are to the usability of a system, icons are also one of the biggest problems in interface design. Imagery that distracts, confuses, or simply bewilders the uninitiated user is all too common in the current generation of graphical applications.

Visual metaphors can help users understand the working of their environment, but only to the extent that the mapping between image states and system states corresponds to the user's understanding of the real-world analogy.

4.1.1 Improving Design, Aesthetics, and Context Functionality

This section will give some guidelines that help in the design of better and more consistent icons. Shneiderman gives icon-specific guidelines for graphical user interfaces (GUI's):[131]

1. Represent the object or action in a familiar and recognizable manner.
2. Make the icon stand out from its background.

---

[131] **Shneiderman, B.** *"Designing the User Interface: Strategies for Effective Human-Computer Interaction,"* Addison-Wesley, 1992.

3. Consider three-dimensional icons; they are eye-catching, but also can be distracting.
4. Ensure that a single selected icon is clearly visible when surrounded by unselected icons. In VIL this is done by highlighting the icon when moving the mouse over an icon to indicate that the icon is selectable.
5. Make each icon distinctive from every other icon.
6. Ensure the harmoniousness of each icon as a member of a family of icons.
7. Design the movement animation: when dragging an icon, the user might move the whole icon, just a frame, or a black box.
8. Add detailed information, such as shading, thickness, color, or animation.
9. Explore the use of combinations of icons to create new objects.

Marcus applies semiotics as a guide to four levels of icon design:[132]

1. Lexical qualities: machine-generated marks - pixel shape, color, brightness, blinking.
2. Syntax: appearance and movement - lines, patterns, modular parts, size, shape.
3. Semantics: objects represented - concrete, abstract, part-whole.
4. Pragmatics: overall - legible, utility, identifiable, memorable, pleasing.

He recommends starting by creating quick sketches, pushing for consistent style, designing a layout grid, simplifying appearance, and evaluating the designs by testing with users.

"An icon can be seen first by its perceivable form (syntax), second by the relation between its form and what is signified (semantics), and third by its use (pragmatics)."[133]

For example, for an icon representing a man, the meaning may not simply be its denotation but rather its pragmatic effect. The symbol is likely to be found on a bathroom door and has the effect of "here is the men's lavatory".

Shneiderman considers a fifth level of icon design: [134]

5. Dynamics: receptivity to clicks - highlighting, dragging, combining, sound, animation.

[132] **Marcus, A**. *"Graphic Design for Electronic Documents and User Interfaces,"* ACM Press, New York, 1992.

[133] **Dormann, C.** *"Self-Explaining Icons,"* University of Brighton Grand Parade, http://www.intellectbooks.com/iconic/self/self.htm.

[134] **Shneiderman, B.** *"Designing the User Interface: Strategies for Effective Human-Computer Interaction,"* Addison-Wesley, 1992.

The American Institute of Graphic Arts (AIGA) uses three of Marcus' dimensions in evaluating over 300 pedestrian oriented symbols from all over the world. They evaluated the strengths and weaknesses of every symbol in relation to these dimensions:[135]

1. The *syntactic* dimension refers to the relation of one visual image to another: How well does this symbol relate to the other symbols? How well do the parts of this symbol relate to each other? Is the construction of this symbol consistent in its use of figure/ground, solid/outline, overlapping, transparency, orientation, format, scale, color, and texture? Are the most important elements recognized first?
2. The *semantic* dimension refers to the relationship of a visual image to meaning: How well does this symbol represent the message? Do people fail to understand the message the symbol denotes? Do people from various cultures misunderstand this symbol? Do people of various ages fail to understand this symbol? Is it difficult to learn this symbol? Does this symbol contain elements that are unrelated to the message?
3. The *pragmatic* dimension refers to the relation of a visual image to the user: Can a person see the sign? Is this symbol seriously affected by "visual noise" (poor lighting, oblique viewing angles, etc.)? Can this symbol be enlarged and reduced successfully?

4.1.2 Bad Icons

 The next few sections (adapted from Mullet, K. and Sano, D. *"Designing Visual Interfaces,"* SunSoft Press, 1995) show potential errors and bad icons that can occur at these three dimensions:

- *syntax*, which governs relationships between elements within the sign,
- *semantics*, or the meaning of the sign elements, and
- *pragmatics*, or the suitability of the image for a particular physical display and set of interpreters.

4.1.2.1 Misleading Syntax

The elements of any visual representation must combine properly to form a coherent sign. Every apparent logical relationship between elements in the image should reflect an analogous relationship between the corresponding sign objects. The images in figure 23 from an air terminal signage program are extremely confusing because the

---

[135] **AIGA,** *"Symbol Signs,"* The American Institute of Graphic Arts, 1993.

logical relation of the aircraft to the arrow is unclear. Because the viewer assumes that all elements of the image exist within the same coordinate system, the arrows appear to be "pushing" the tail of the plane to the left in each image. The intended meanings (Take Offs and Landings respectively) can be understood only when the viewer realizes that while the aircraft is being viewed from above, the arrow is being viewed from the side. While the selected viewpoint makes the aircraft easily recognizable, it is a highly unfamiliar vantage point for take-offs and landings. A more natural perspective presenting the plane in a side view (with the nose pointed up or down for take-off and landing respectively) would communicate more effectively. This representation would place the aircraft in the "appropriate attitude" and bring the elements of the sign into a more compatible spatial relationship. [136]
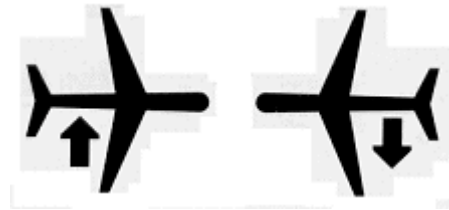
Figure 23 (Mullet and Sano, 1995).

4.1.2.2 Dominant Secondary Elements

Creating composite imagery presents syntactic problems. When secondary elements are balanced correctly, they can provide valuable context without interfering with the image. When secondary elements are too strong, they can so dominate perception that the rest of the information in the image can be extracted only with difficulty.

4.1.2.3 Images based on obscure references

When depicting abstract concepts, the designer often has no choice but to substitute a concrete object, which serves to bring to mind the meaning of the abstract concept. The semantic relation, however, is rarely a strong one, even with a well chosen metaphorical reference. When the reference is strained or obscure, the meaning is never apparent. The icons in figure 24 rely heavily on literal, concrete metaphors to suggest the corresponding functions. The magnet tool lets you move an entry from

---

[136] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

one page to another in an appointment book, and the footprints represent backtracking.[137]

Figure 24 (Mullet and Sano, 1995).

4.1.2.4 Cultural or language dependencies

The pragmatic aspect of an image describes assumptions about the viewer and the viewing environment that are implicit in any visual representation. An image may work for some viewers, but not for others. If the image relies on inside jokes, figures of speech, slang, or other terminology that is well-known only within a particular subculture, than it will be intelligible only to members of that group. In figure 25, the icon for screen dump uses visual and verbal puns that are only recognizable to those for whom the concepts are already familiar. Similarly, the origins of the term "debugger" are familiar to most developers, but not to many end-users.[138]
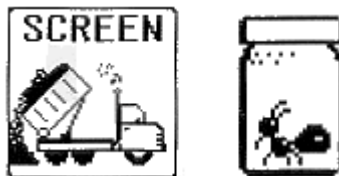
Figure 25 (Mullet and Sano, 1995).

4.1.2.5 Offensive or suggestive imagery

Imagery that some users may find personally or culturally offensive should always be avoided in commercial products. While examples as extreme as the *Kill* icon from NeXTStep in figure 26 are rare, it is not hard to find images that would be considered at least mildly upsetting in some cultures or to some users. Gestures too have different meanings in different cultures. The "thumbs up" sign does not mean "OK" in the

---

[137] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

[138] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

Middle East, for example. Images related to death, injury, or violence, in particular, are almost never appropriate in an office environment.[139]

Figure 26 (Mullet and Sano, 1995).

4.1.3 Icon Design Guidelines

Effective imagery must possess a perceptual immediacy that allows it to be recognized at a glance. For most images, this involves a process of careful abstraction in which all but the elements that most characterize the sign object are removed. When developing multiple images, care must be taken to maintain cohesion within the image set and to consider the physical, conceptual, and cultural context in which the images will ultimately be displayed.

Following are some guidelines for icon design (most of which are adapted from Mullet and Sano, 1995):

4.1.3.1 Simplicity

The importance of simplicity can hardly be overstated. Simple icons can be rapidly apprehended and understood well enough to support immediate use of a system or invite further exploration (this is what is called approachability). Simple icons can be recognized more easily than more complex ones. Since they represent less visual information to the viewer, they are more easily assimilated, understood, and remembered (recognizability). Simple icons have a greater impact than complex icons, because they can be immediately recognized and understood with a minimum of conscious effort (immediacy). Consequently, simple icons improve the usability of any visual system (usability). The quality of perceptual immediacy makes images ideally suited to public safety applications such as the labeling of hazardous material or the signage needed for traffic control.

---

[139] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

"Simplicity is a property pertaining to easiness of perception and identification."[140]

4.1.3.2 Reduction

Reduction through successive refinement is the path to simplicity. Anything that is not essential to the communication task must be removed from the icon. This way, significant elements will be noticed. The signs in figure 27 are typical of an entire class of objects rather than on the details of any one instance. All visual details except those needed to identify the object's category are removed. This *reduction of iconicity* makes the images more portable across cultural and linguistic boundaries.

Figure 27: Reduction plays the critical role of emphasizing canonical features in these public information icons developed for the U.S. Department of Transportation (DOT) by the American Institute of Graphic Arts (AIGA).

Mullet and Sano describe reduction as a three-step process:[141]

1. Determine the essential qualities (typically a short list of adjectives) that should be conveyed by the design (of the icon), along with any fixed formal elements, such as a name or label, a color, texture, or pattern.
2. Critically examine each element in the design and ask yourself why it is needed, how it relates to the design (identified above), and how the design would suffer without it. If you can not answer any of these questions, remove the element.
3. Try to remove the element from the design anyway. What happens? If the design collapses, either functionally or aesthetically, the element must be replaced. Otherwise, consider it omitting it from the final solution.

---

[140] **Meunier, J**. *"The Categorical Structure of Iconic Languages,"* Université du Quebec a Montré,
http://pluton.lanci.uqam.ca/membres/meunier/iconlang/iconlag.htm.
[141] **Mullet, K.** and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

4.1.3.3 Generality

Imagery is often used to represent a class of artifacts, rather than any particular instance of that class. Generality describes the ability of the individual sign to represent these higher-level groupings. Images with a high degree of abstraction automatically provide a general representation of a class of similar objects by removing visual details specific to any one instance. The generality of these images, moreover, allows them to be interpreted as representative of a broader class than might be suggested by a photograph or detailed drawing of the same object.

Generalization results in simpler forms that, because they contain less visual information, are easier to process, recognize, and react to. Generalization also allows elements of the original form to be selectively emphasized or de-emphasized to facilitate particular communication objectives as detail is removed from the image (Figure 28).
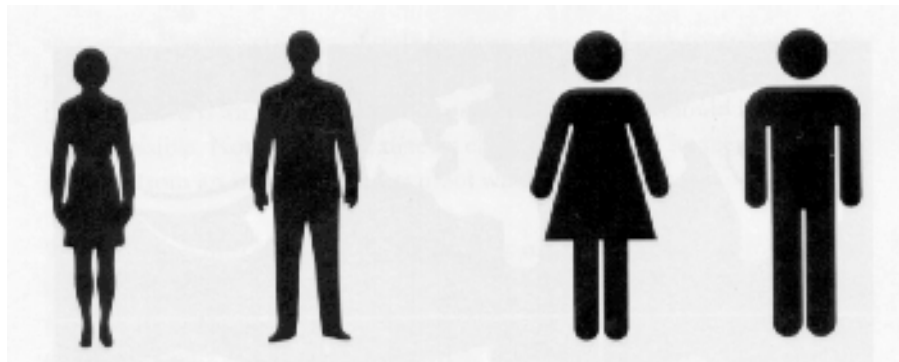
Figure 28: Two very different levels of abstraction are apparent in these icons. The realistic images on the left provide a representation that is more direct, but less general. Their accuracy strongly suggests that the icons in (a) represent typical individuals, rather than the more general populations seen in (b).

4.1.3.4 Cohesiveness

Images rarely appear in isolation. Each image generally forms part of a larger system in which many individual visual elements must work together effectively as a group. A cohesive system arises when shared formal qualities of the images themselves can be recognized effortlessly in early visual perception. Repetition of common forms throughout an image set helps users learn to "read" the resulting visual language and further enhance their processing of the information being presented. The waves in some of the Olympic pictograms, for example, make it easy to identify those icons

representing water sports. The icons in figure 29 represent wrestling, swimming, weightlifting, archery, rowing, and volleyball respectively.
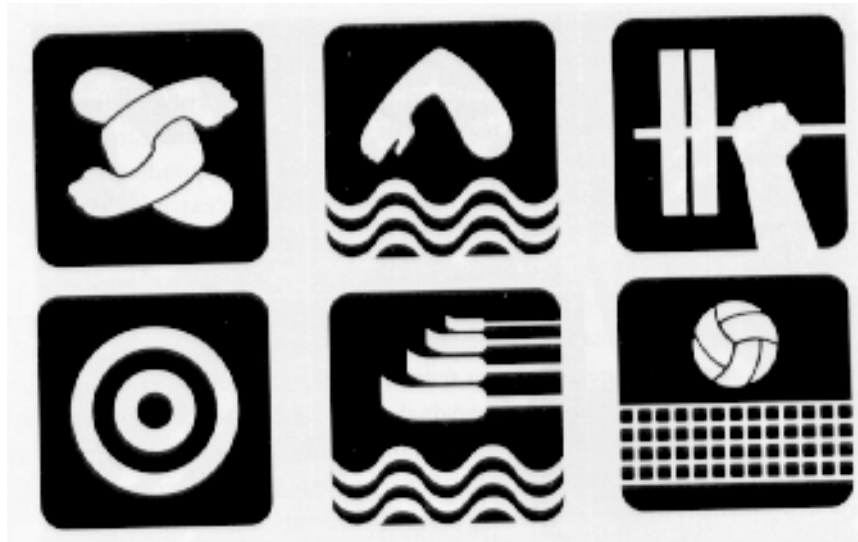

Figure 29: Cohesiveness based on a set of visual primitives can be seen in these icons for the 1968 Mexico City Olympics. Bold forms, close cropping, and light/dark balance are used consistently throughout the set (Mullet and Sano, 1995).

4.1.3.5 Regularity

Regularity reduces information by repeating elements according to some rule or principle. Human perception and memory operate more efficiently on regularized stimuli. The predictability of a regular pattern allows the viewer to "scan ahead" more easily when looking for a specific icon. Elements in different categories may all share some feature, for example, so the viewer knows that icons with this specific feature belong to that category. Features could be size, shape, color, texture, etc. Critical elements intended to stand out in the icon should not be regularized. Any irregularity will be interpreted as significant by the viewer, who will cheerfully ascribe a meaning to it even when none was intended. By regularizing non-critical elements throughout the icons, one will be able to attract the user's attention reliably to irregularities where one does want to make a distinction.

4.1.3.6 Viewpoints

Choosing the right point of view can be as important as determining which qualities to emphasize, since, from some vantage point, the parts that most effectively characterize the subject may be visible poorly or not at all. It would rarely be useful,

142

for example, to depict a chair from above or a radio from behind. The icons developed for the US Department of Transportation (see figure 30) also show how choosing the right viewpoint can help distinguishing objects.

Figure 30: Effective characterization depends on choosing the right point of view, eliminating non-characteristic details, and exaggeration of defining features, as in these icons from the U.S. DOT's Symbol Signs program (AIGA).

The viewpoints chosen for the DOT icons, for example, are those from which these vehicles are seen most frequently in the viewer's everyday experience. This strategy allows the image set to exploit the viewer's familiarity with a particular characteristic viewpoint.

4.1.3.7 Communicability

The communicability of any representation depends on a shared context between sender and receiver that allows signs to be interpreted within a pragmatics comparable to the one under which they were encoded. The experience of an image is affected not only by other images in the ensemble, but also by the physical, cultural, and conceptual environment in which it appears. The effect of context on our understanding of the world goes largely unnoticed. What is obviously a mailbox in one culture, for example, would be surprisingly unfamiliar to people from other cultures.

Figure 31: Cultural dependency is apparent in these icons representing mailboxes in the U.S. (a), Denmark (b), France (c), and Italy (d). Each could be interpreted as a control box or trash can by someone unfamiliar with the local convention (Mullet and Sano, 1995).

Saussure concluded that meaning is not inherent in things but is constructed through a social contract. He describes the linguistic sign as a kind of binary pair: a mental concept or idea, and the code element that represents it. Each helps to define and delineate the other from the formless chaos of infinite possibility, and the link between the two, and their differentiation from what surrounds them, is maintained by a kind of social contract or agreement.[142] This agreement can, however, differ from culture to culture.

Roscoe stated that a purely iconic language could not constitute a vehicle of communication, since we need marks that are purely arbitrary conventional signs, e.g. for past-tense. "If an iconic system is to be exploited, then the vocabulary of an iconic system will have to be supplemented with marks that do not have natural meanings."[143]

Iconic communication systems apparently cross language boundaries because they are based on a cultural consensus that is broader than that on which verbal languages are constructed.[144]

Effective visual representations for international audiences should be based on aspects of the sign object that are truly universal within the target population.

4.1.3.8 Consistency

The consistent appearance, placement, and meaning of important visual cues make it easier for viewers to interpret and respond to new situations as they arise.

4.1.3.8.1 Scale, Contrast, and Proportion

The effectiveness of a clear composition always depends at least as much on the relationships among the parts as it does on the parts themselves. When a single element is too large or small, too light or dark, too prominent or indistinct, the whole design suffers.

Scale describes the relative size or magnitude of a given design element in relation to the other design elements and the composition as a whole.

---

[142] **Saussure, F. de** *"Course in General Linguistics,"* trans. Wade Baskin, London, 1960.

[143] **Roscoe, J.** *"The limits Of Iconic Communication,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

[144] **King, A. J.** *"On the possibility and impossibility of a universal Iconic Communication System,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

Contrast results from noticeable differences along a common visual dimension, and provides the basis for visual distinctions. The dimensions include size, shape, color, texture, position, orientation, and movement, described by Bertin as *retinal variables*.[145] Contrasts can often be used to underscore essential qualities in an image.

Proportion deals in ratios rather than fixed sizes. Proportion is the metric that guides the choice of scales in a contrast relationship.

Benefits of effective use of scale, contrast, and proportion are the following:
- Differentiation - contrast is essential for differentiating elements from one another.
- Emphasis - scale and contrast can be used to emphasize important elements or areas in an icon.
- Activity - scale and contrast move the viewer's eye through the composition. Elements in contrast exert an influence on each other that exaggerates their contrasting qualities.

Figure 32 shows some adjectives from VIL in which the concepts are made clear by showing contrast and scale differences between two extremes.
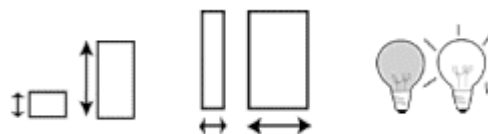
Figure 32: Icons representing the concepts of length (short-tall), thickness (thin-thick), and darkness (dark-light).

Icons can be categorized in the manner of McLuhan in terms of the temperature of information, which they can convey, according to their position on an axis stretching from pictures to symbols, and on their level of animation, see Figure 33.[146]

---

[145] **Bertin, J.** *"Semiology of Graphics,"* translated by P. Berg, Madison: University of Wisconsin Press, 1983.

[146] **Mealing, S.** *"Do you see what I'm Saying,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.
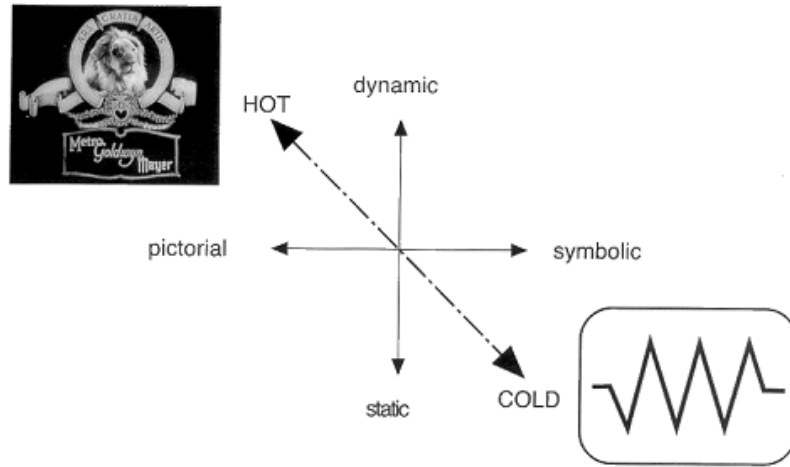
Figure 33: Temperature of Icons.

The 'hottest' presentation is not necessarily the best presentation, because it may become too specific to illustrate a general case. Too specific icons could be interpreted in multiple ways. Also, the differentiation from other icons becomes harder, i.e. the recognition is slower.

Figure 34 shows five different ways of referring to men, of which the second one is very symbolic and the last one too specific.[147]

Figure 34: Five ways of referring to man.

4.1.3.9 Subjectivity to Fashion

The style of icons is subject to fashion as any other design product. For example, in most applications of today, the icon with a diskette on it represents the save file option. With the way technology is going, diskettes may become obsolete. If diskettes cease to be used in the future, subsequent generations will require either a new icon or

---

[147] **Beardon, C.** *"IconText: An exploration of the Limitations of Iconic Languages,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

a short history lesson. It is however possible to design icons that have a long life span and that need no future stylistic updating without compromising continuity of recognition.

4.2 Designing Cross-Cultural Icons

To create icons that people around the world will recognize and understand, we must try to understand these people as well as we can. What symbols and objects will they recognize? Are they familiar with the same symbols as we are? Can we assume they have traveled extensively? Can we assume they watch movies and television?

From the examples in section 4.1.1, we have seen that if an icon relies on inside jokes, figures of speech, slang, or other terminology that is well-known only within a particular subculture, then it will be intelligible only to members of that group, and that icons that might be culturally offensive should always be avoided (e.g. the "thumbs up" for "OK").

"If (like most Western codes, such as tourist amenity symbols) the meaning of an icon depends on knowledge of experience accessible to relatively few, then it will be seen by the rest of the world as a secret, exclusive code."[148]

Horton describes some general principles for creating international icons[149]:
- Remove or translate text
- Avoid culture specific symbols like
    - puns and verbal analogies
    - gestures and body parts
    - mythological and religious symbols
    - animals
    - colors
    - provincialism
- Consider reading and scanning direction
- Show best-known version
- Use abstract symbols

---

[148] **King, A. J.** *"On the possibility and impossibility of a universal Iconic Communication System,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

[149] **Horton, W.** *"The Icon Book: Visual Symbols for Computer Systems and Documentation,"* John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

4.2.1 Remove or Translate Text

Generally, icons that rely on text for their meaning must be redesigned or translated for users who speak another language. Since VIL is a pure iconic language, we will not have any text in the program. The only place we use text in icons is for proper names, for example "John" or "Paris", and labels of measurement, for example in weight we use "g" for gram. Since we are using the SI-system (see chapter 3), these labels are considered universal.

4.2.2 Avoid Culture-Specific Symbols

The meaning of an icon depends on the association it triggers in the mind of the user. The same image can have vastly different associations in different cultures. We must choose symbols that have the same meaning for all users.

4.2.2.1 Puns and Verbal Analogies

Visual puns show objects whose name sounds like that of the concept you want to represent. Visual puns should be avoided, because they require subtle knowledge of language, and are impossible to translate. They are not universal. For example, let us assume that we represent the "mouse" that controls the pointer on the screen by an image of the animal mouse. The problem with this is that in some languages the name for the box that controls the screen is not the same as that of a small rodent.

4.2.2.2 Gestures and Body Parts

We must be very careful when using hands in icons. There is almost no arrangement of the human hand that is not an obscene, rude, or impolite gesture somewhere in this world. Consider the various meanings of a few gestures common in the USA (taken from William Horton, 1994):

Icon	Intended Meaning	Problems
	Yes, OK	In Sicily, this is an invitation to insert the thumb into a private part of the anatomy.
	Precisely, yes	In France, this means zero or worthless. In Japan it is a reference to money. In South America it means that the viewer is a part of the anatomy of this shape.

If one must show hands, show them holding, pressing, or moving something as in the following icons from VIL representing the categories: verbs representing an exchange of possession, put verbs, hold/keep verbs, and verbs of exerting a force:

4.2.2.3 Mythological and Religious Symbols

We do not all share the same religious and mythological symbols, and consequently should avoid such symbols.

4.2.2.4 Animals

Tribal societies, sports teams, political parties, and nations commonly adopt animals as emblems. Such totems imply the group possesses the desirable characteristics of the animal: the courage of a tiger, the ferocity of a lion, the swiftness of a hawk, etc. We should not likewise use images of animals to represent characteristics traditionally associated with these animals, because many such associations are local to a single culture:[150]

---

[150] **Horton, W.** *"The Icon Book: Visual Symbols for Computer Systems and Documentation,"* John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

Icon	Intended Meaning	Problems
	Loyalty Search, retrieve	To Americans, dogs are pets or hunting companions; to many Asians, they are food.
	Savings account	To Muslims and many Jews, the pig is considered unclean and unholy.
	Deception	In the orient, the snake is a symbol of life and rebirth.

Also, many animals have religious significance, e.g. cows to Hindus, pigs to Muslims and Jews, lamb and fish to Christians. We will only use images of animals when we want to represent the animals themselves, i.e. an icon of a cow is only used when we are talking about cows.

4.2.2.5 Colors

Conventional meanings of colors often vary. In the UK, the first place is often awarded a red ribbon, in the US first place earns a blue ribbon. The use of green to symbolize money works in America only, because American paper money is colored green. The meanings for color come from a culture's religion, literature, and fine arts. Here are some common associations for colors in different cultures:[151]

Color	Western European	Japanese	Chinese	Arabic
Black	Death, evil			
White	purity, virtue	death, mourning	death, mourning	
Red	danger	anger, danger	joy, festive occasions	
Yellow	caution, cowardice	grace, nobility, childish, gaiety	honor, royalty	happiness, prosperity
Green	safe, sour	future, youth, energy		fertility, strength
Blue	masculinity, sweet, calm, authority	villainy		virtue, faith, truth

---

[151] **Horton, W.** *"The Icon Book: Visual Symbols for Computer Systems and Documentation,"* John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

The question is: do we need color, other than to represent the colors themselves (adjectives in 3.2.2.4)? Can we represent all the icons in black and white? Color certainly makes icons and therefore the whole system livelier. Colors could be very helpful to show the distinction between a tomato and an orange. However, black and white icons are more neutral and more generic, since one does not presuppose a color. Also, specific objects have different colors in different cultures. Consider for example something as simple as an apple. It can be yellow, green, or red. Giving it a color could make it less recognizable to some cultures than others. Furthermore, it is also more consistent to use black and white images throughout the whole system.

4.2.2.6 Provincialism

Universal concepts do not always lead to universal symbols. The same general idea may appear quite different in different countries. An activity may be performed in a different manner and the appearance of particular objects can vary too. Consider the universal concept of reminder represented by the next provincial symbol: [152]
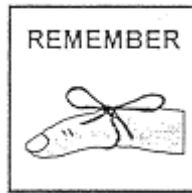
4.2.3 Consider Reading and Scanning Direction

The accustomed reading direction of the user influences the interpretation of the sequence of events in a graphic and the relative importance and virtue of objects in a graphic. We are not talking about the display of the icons on the screen, but about the elements inside a single icon. Just as we propose having no order in the interpretation of messages (the icons on the screen that represent a message can be placed in any order), we should strive for order independence inside of icons as well. We cannot assume left-to-right, right-to-left, or top-to-bottom order. Let us assume we want to make an icon for the abstract concept represented by the verb *to fill* by having an empty glass on the left of the icon and a full glass on the right side. In Arabic countries, where people read right to left, this icon would be confusing, unless we add an arrow pointing away from the empty glass and towards the full glass.

---

[152] **Barker, P.** *"Human Communication Processes,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

Figure 35: Icons representing the abstract concept "to fill". The left icon is ambiguous, since it does not consider reading direction. It could easily be seen by an Arabic person as "to empty". The right icon has an arrow indicating reading direction.

In VIL, we have consistently adopted to indicate direction or scanning sequence by an arrow, as shown in the previous figure.
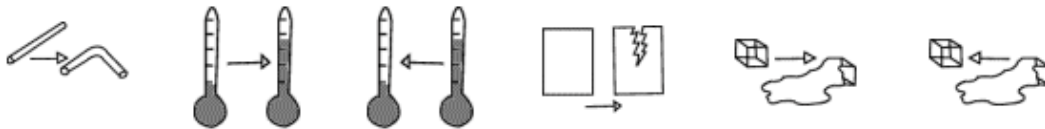

Figure 36: Icons representing the abstract concepts to bent, to heat, to cool, to tear, to melt, and to freeze.

If the icon does not represent a state change, one could indicate scanning direction in two other ways:

- make the icon read from top to bottom (all languages read from top to bottom of a page)
- draw images symmetrically or in other ways that do not depend on a left-to-right sequence, for example:

4.2.4 Show the Best-Known Version

Many objects come in several versions. When using such an object in an icon, we must pick the version most widely recognized. If there is no international version, then pick the most common version familiar to the largest number of users. Also consider the oldest of several contending versions. Older versions are more likely to have appeared in movies seen abroad. Older versions may be more common in some

less affluent countries. Here are some old, but recognizable objects (taken from William Horton, 1994):

4.2.5 Use Abstract Symbols

Most simple geometric shapes are free of cultural connotations. Icons using abstract shapes and lines often work well, but are conventional:

verbs "to unformat/scramble" and "to rotate" (taken from William Horton, 1994) .

In this section we have tried to give some general guidelines in icon design and have also showed examples of bad icons and things to avoid. The next section describes in detail the icon language of VIL.

4.3 VIL's Icon Language

In chapter 3 we discussed the grammar and vocabulary of the communication language VIL. In this section, we discuss a different sort of grammar and vocabulary, namely that of the icons themselves. First, in section 4.3.1, we determine what needs to be represented. In section 4.3.2 we give principles for designing icons for visual elements and semantic roles. We will discuss the process of creating a vocabulary of visual elements, which are elements of an icon that we can combine according to some rules to produce the icons and grammatical entities that we need. In section 4.3.3 we give some rules for designing icons for the grammatical categories of verbs, nouns, and adjectives (both category and terminal elements).

Just like VIL, the icons consist of two parts: a vocabulary (i.e. visual elements) and a grammar. The vocabulary is a collection of visual elements that are combined according to the rules of the icon grammar to form VIL icons. For VIL itself, the grammar records how to put icons together to form a message. For icons, the grammar records how to put visual elements together to form a comprehensible icon. This icon then represents an entity in VIL's grammar, e.g. it might represent the agent noun.

Besides iconic representations, VIL will also have entities that are not represented iconically, but graphically. For example, if the location is a country, one can select the country from a map of the world. This map is not an icon but an image, and we call this a visual or graphical representation.

Principle 4.1: *Iconic vs. Visual Representation in VIL*

> In VIL, in addition to icons, visual representations are used in elements the user can interact with. These are selecting a location from a map, selecting a time on a clock, and selecting a date from a calendar.

In order to design the icons, we must design the visual elements and set rules for combining them. The next few sections describe the steps involved in designing the visual elements and setting up the rules.

4.3.1 Determine what needs to be represented

The first step in the design of the icons is to determine all the things for which we need icons and visual representations. In order to determine these, let us first give the grammatical entities again:

> TIME-WHEN,
> PLACE-WHERE,
> VERB,
> AUX-MODAL,
> AKTIONSARTEN,
> ADVERB,
> INTENSIFIER,
> NOUN,
> NOM-PRONOUN,
> NOM_DEMONSTRATIVE,
> QUANTIFIER,
> NUMBER,
> ADJECTIVE,

SUBORDINATOR,
COORDINATOR.

Most of these grammatical entities are represented iconically. The time-when and the place-where cases can, besides being represented iconically, also be represented graphically by a clock, calendar, and place on a world map respectively.

We need to represent the following:
- **Visual elements**
- **Grammatical cases**
- **Entries in the Iconicon:**
  - Category and terminal icons
  - Concrete and abstract meaning
- **Program icons**

The visual elements are the building blocks for the icons. Icons for the grammatical cases are needed, since they are not only used as visual element, but also as an icon when the case is not yet instantiated. Visual elements and grammatical case icons are discussed in the next section. The majority of icons will be part of the browsable part of the iconicon, which is the vocabulary of the language. These are the verb, noun, and adjective hierarchies. Besides icons for abstract meanings of terminal nodes in the hierarchies for the substantive categories (e.g. *"memory"* or *"to throw"*), we also need abstract icons for non-terminals of the hierarchy (category names), e.g. an icon for *"verbs of motion"* (where the generalized form is the verb *"to move"*) or *"living things"*. Last, but not least, we need some program icons that are used to guide the user to the completion of a message, e.g. new message, save message, send message, exit, etc.

4.3.2 Design icons for visual elements and grammatical entities

In commercial products, visual elements are often combined with icons to associate them with a particular manufacturer or product line of which they are a part. For example, assume a company is designing icons for a program used to create multimedia presentations. The visual element:

which could be some company trademark, may be incorporated into the icons for the various media handled by the product (taken from William Horton, 1994)[153]:

In VIL, visual elements can be put to use in several ways:

Semantic Roles of Nominal Constituents
Visual elements are used to distinguish between the grammatical cases in a message. For example, without visual combining features, it could be difficult to see the difference between the icons for the subject, direct object, and indirect object of a message. Consider the message "The boy gave the book to the teacher". The subject of the sentence is *the boy*, and he is also the source. The indirect object is *the teacher*, and he/she is also the goal. Visual elements are used to represent the cases.

When a case is uninstantiated, these visual elements are icons representing the grammatical case. Once the case is instantiated, the visual element becomes part of the icon, although at a much smaller size.
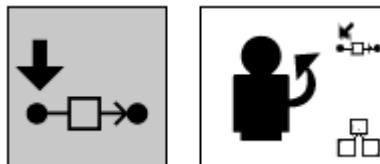
Figure 38: An uninstantiated icon representing the source NP on the left and its instantiated version on the right, where the pronoun "I" was chosen.

Principle 4.2: *Visual elements as case distinguishers*

> Icons representing semantic roles are used as visual elements in instantiated icons to clearly identify the case of entity that the icon represents.

---

[153] Horton, W. "The Icon Book: Visual Symbols for Computer Systems and Documentation," John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

Principle 4.3: ***Visual elements to distinguish instantiated icons from uninstantiated icons***

> Background color is used to distinguish instantiated from uninstantiated icons. An instantiated icon is an icon for which the slot has been filled. All uninstantiated icons in VIL have a gray background, all instantiated icons have a white background.

Derivational Morphology: Grammatical Categories
Visual elements are used to distinguish between instances of grammatical categories in a message, like verbs, nouns, adjectives, etc. For example, consider the noun *key* and the verb *to open*. They can both be represented by an image of a key, if we had some visual element to distinguish between verbs and nouns.

Principle 4.4: ***Visual elements as category distinguishers***

> A border is used to distinguish the main grammatical entities verb, noun, and adjective.

The previous principle makes the next principle possible.

Principle 4.5: ***Visual elements to represent derivational morphology***

> Borders are also used to represent derivational morphology. When a grammatical entity is derived from another grammatical entity (e.g. a de-nominal verb), then the outer border denotes what entity the icon represents (e.g. a verb), and the inner border denotes what entity it is derived from (e.g. a noun).

Derivational morphology has not been implemented in the first release of VIL. The application, however, has been designed with derivational in mind, and it will be included in the second release of VIL. Consequently, borders around adjectives are not implemented in the first release (adjectives are always shown in isolation in the noun properties dialog).

Other functions for visual elements

Besides visual elements for semantic roles and derivational morphology, there are some other functions that visual elements are used for.

Principle 4.6: *Visual elements to indicate degree of gradable concepts*

Gradable concepts are represented by a value along a slider with at the ends the two extremes/antonyms. Among these are most adjectives and de-adjectival adverbs. The grammatical case of duration can also be indicated by a slider (for a short time - for a long time).

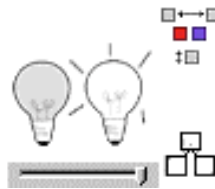The adjective "light" is then represented as follows:

Figure 39: VIL icon representing "light". Notice the value of the slider being all the way to the right. The visual elements on the right side of the icon are the grammatical entity (adjective) and the category visual element. Note that the same icon with the slider on the left would mean "dark".

Principle 4.7: *Visual elements for negation*

Visual elements are used to indicate negation, for example a diagonal bar through an icon.

Principle 4.8: *Visual elements for direction, state and choice*

In VIL, arrows are used as visual elements with the following three functions, represented as sub-principles:

Principle 4.8.1: *Visual elements for direction/motion*

Direction or motion in VIL is indicated by arrows. These arrows show the direction an object is moving or being moved.

Principle 4.8.2: *Visual elements for change of state*

> Change of state in VIL is indicated by arrows. These arrows point from the before state to the after state.

Principle 4.8.3: *Visual elements for choice*

> Arrows are also used for choice. This occurs, for example, with contradictory adjectives in which the arrow points to the antonym of choice. Another way to indicate choice is by using arrows as "part-of" indicators. These arrows point to the part of the icon that is meant.

Principle 4.9: *Visual elements to indicate a selected icon*

> A selected icon is clearly made visible by beveling it, i.e. raising it from the screen on a mouse over.
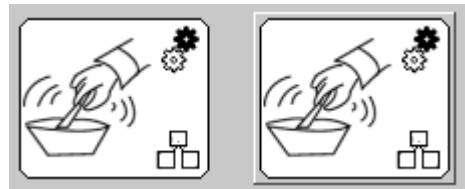
Figure 39: An unselected and a selected icon representing the verb "to stir".

In VIL, not every visual element is used to add meaning to an icon. Some visual elements are meant to help explain the user something. Consider the next icon of a source NP "the big red apple", where the head noun is "apple":
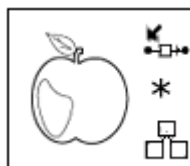
There are four visual elements present in this icon. The border indicates that we are dealing with a noun. Then there are three visual elements on the right. The visual element at the top indicates that the grammatical case is a source. The other two visual elements are not adding meaning to the icon. Instead they are there to help the user.

Principle 4.10: ***Visual elements to indicate modification***

> An asterisk is used to notify the user that a head noun has some noun modifying properties set. In the above example, the NP is not just "apple", but "the big red apple". The user is encouraged to click the icon to find out about its (case) modifying properties. This can also indicate that the case is a conjunction of two or more nouns.

Principle 4.11: ***Visual elements to allow the user to find where an item occurs in the iconicon***

> The visual element at the bottom of the icon represents a tree structure. When the user click on this visual element, a popup window appears showing the user where in the hierarchy the icon belongs, i.e. the path from the root of the noun classification to the terminal apple is shown. To even further help clarify the icon, a set of icons that are siblings of apple is shown, i.e. they belong to the same category as apple, namely that of fruits.
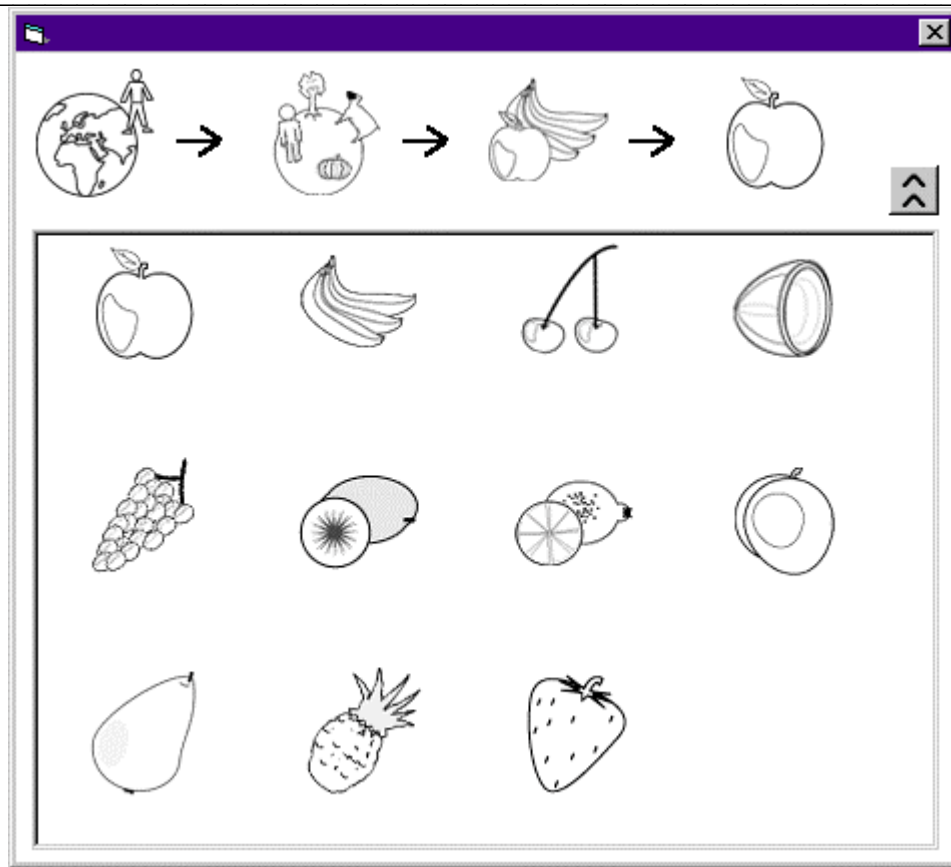
4.3.2.1 Rules for representing visual elements

The following table shows which rule, i.e. what convention, is applied to the previously mentioned principles:

Nr	Principle	Rule
4.2	*Visual elements as case distinguishers*	Each noun has a case distinguisher in upper-right corner. Verbs and adjectives have verb and adjective marker elements in the upper-right corner.
4.4	*Visual elements as category distinguishers*	Each verb, noun, and adjective has a category distinguisher element (as a border around the icon)
4.5	*Visual elements to represent derivational morphology*	Border around border
4.6	*Visual elements to indicate degree of gradable concepts*	Slider below icon

4.7	*Visual elements for negation*	Each case can be negated by a negation symbol on top of icon, e.g. "I did *not eat*…", "*not I* ate…" (i.e. it was not I, who ate…), or "*not a 3 o'clock*"
4.10	*Visual elements to indicate modification*	Nouns may have a modification element (in middle right)
4.11	*Visual elements to allow the user to find where an item occurs in the iconicon*	Each noun, verb, and adjective has a hierarchy review element (in the lower right)

Figure 40 shows a VIL icon for the source noun phrase "apple" and explains its various visual elements:
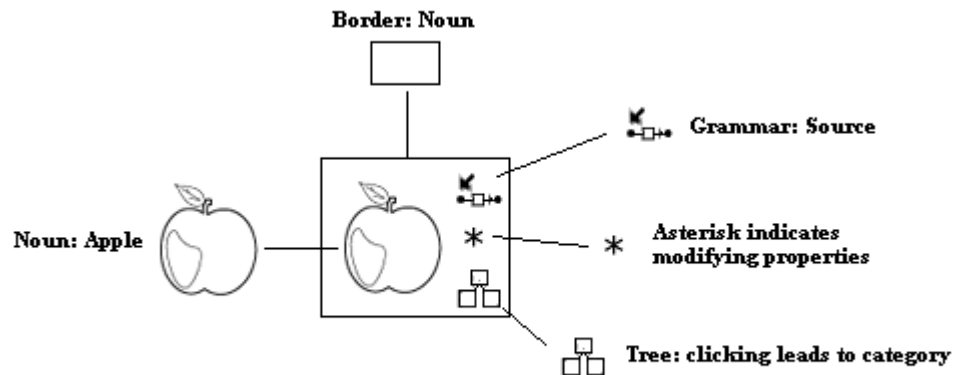

Figure 40: Example of various visual elements. The icon for the VIL source NP combines various visual elements

4.3.2.2 Relationships among Objects

Often the meaning of an icon is not in the objects or actions shown, but in the relationship among the objects in the icon. Some common relationships in VIL are:

- *Relationship: Ownership*

One object may possess or control another (Figure 41).

Figure 41: Icons representing hold/keep verbs, and alienable possession verbs

- *Relationship: Inclusion*

Some objects include other objects. To put it differently, some objects are part of other objects (Figure 42):

Figure 42: Icons representing geography, volume, and a kg

- *Relationship: Part of/Association with*

    Some objects are a part of an object or generally go together with another object. Usually this is indicated by pointing to the item or making it more emphatic (Figure 43):

Figure 43: Icons representing steam and water

- *Relationship: Position*

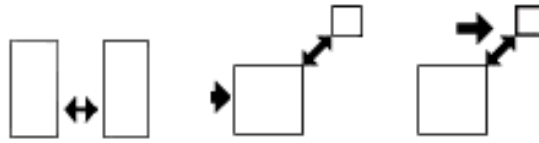Icons often show physical or spatial relationships among objects (Figure 44):

*Figure 44: Icons representing the abstract concept distance and
the nominal demonstratives here and there*

- ***Relationship: States***

  Objects can exist in various states or conditions. Often this is reflected by a choice
  between two antonyms (e.g. *on* vs. *off* for a light switch) or, for gradable concepts,
  by a value among a scale with at the ends two antonyms (e.g. the scale for the
  gradable concepts of speed would consist of a slider with at the two ends images
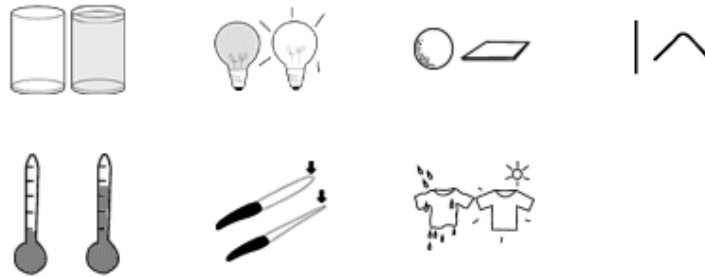 for "slow" and "fast"), see Figure 45.

*Figure 45: Icons representing gradable concepts: empty-full, dark-light, round-flat,
straight-bent, cold-hot, blunt-sharp, and wet-dry.*

One thing to notice with icons for antonyms is that they make sense when seen
together, but are not always clear when seen in isolation. For example, the icons for
empty and *full* only make sense when seen together. Try only looking at the icon for
full. A problem would arise if the users pick the adjective *empty*. Some other user
who has to interpret the message, and who does not see the two antonyms together,
won't understand that this is not just *a glass* but stands for the adjective *empty*.

4.3.3 Icon Representation Principles for Verbs

Below are some principles that apply to various types of verbs.

Principle 4.12: ***Showing the tool or activity to represent verbs of creation and destruction***

Verbs of creation and destruction are represented by showing the tool used in the creation/destruction process or the activity being performed (Figures 46 and 47):

Figure 46: verbs of creation: to present (pointing device and chart), to reveal (napkin), to create (brush and palette), and to record (camera).
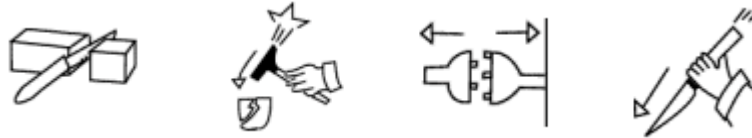

Figure 47: verbs of destruction: cut verbs (knife), destroy verbs (hammer), disconnect verbs (plug), and kill verbs (knife).

Principle 4.13: ***Showing before and after states to represent verbs of change of state***

Many actions transform something from one condition into another. What these icons usually have in common is that they show the before and after states connected by an arrow (Figure 48). If the before state is well known, the icon may show only the destination state (e.g. the verb to boil):

Figure 48: transformation verbs: to bend, to heat, to melt, and to tear.

Principle 4.14: ***Showing concrete action to represent manipulation verbs***

Actions that modify, delete, or rearrange objects, typically show a concrete action upon an image of the data plus the tool that is used in the activity (Figure 49):

165

Figure 49: manipulation verbs: to chop, to cut, to slice, to peal, to pour, and to mix.

Principle 4.15: ***Showing arrows to represent verbs of motion or conceptual movement***

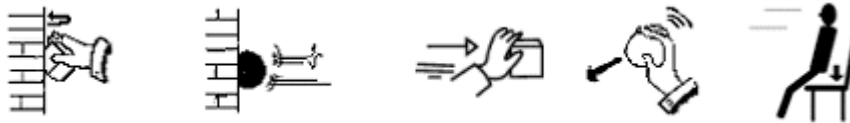Changes in physical space are typically shown by arrows (Figure 50):

Figure 50: motion verbs: conceal verbs, verbs of contact by impact, verbs of exerting a force, propel verbs, and verbs of moving without changing location.

Principle 4.16: ***Showing the activity itself to represent verbs of activity***

An activity is shown by an image of the activity being performed (Figure 51):
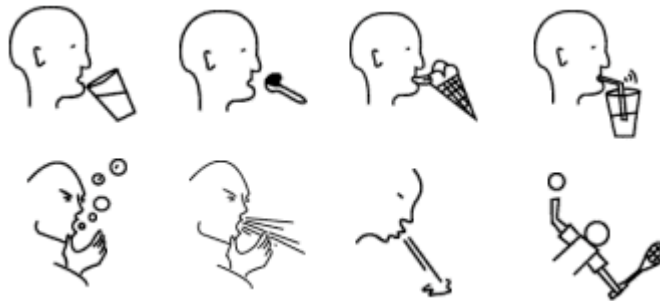
Figure 51: activity verbs: to drink, to eat, to lick, to suck, to burp, to cough, to spit, and to play tennis.

4.3.4 Icons Representation principles for Nouns

Nouns can be concrete objects or abstract concepts. Below are some principles that apply to various types of nouns:

Principle 4.17: ***Showing the object itself to represent concrete nouns***

Concrete objects are always shown by a picture of the object itself (Figure 52):

Figure 52: concrete objects: hamburger, fork, pan, coffee maker, rooster, eggplant, and apple.

Principle 4.18: ***Showing an associated concrete object to represent abstract concepts***

Abstract concepts are often shown by an analogous or associated object (Figure 53):

Figure 53:abstract concepts by showing associated object: measurement, temperature, volume, and weight.

Principle 4.19: ***Using symbols to represent abstract concepts***

Sometimes abstract concepts are shown using abstract symbols (Figure 54):
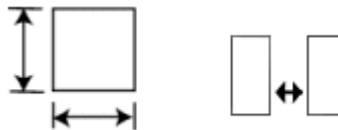

Figure 54:abstract concepts by using symbols: size/dimension and distance.

Principle 4.20: ***Showing several children of a category to represent noun categories***

Categories are generally shown containing several elements from its category (Figures 55 and 56):

Figure 55: categories in the kitchen category: electric kitchen machines, cooking utensils, pots & pans, and silverware.

Figure 56: categories in the food category: dairy products, grain products, cooking liquids, drinking liquids, dishes, type of meals (e.g. breakfast), and spices.

4.3.5 Designing Icons Representing Noun Modifiers or Adjectives

Adjectives sharpen the reference to an object. They modify attributes of a nominal concept from their default or stereotypical value. Below are some rules that apply to adjectives:

<u>Principle 4.21:</u> ***Showing a slider in combination with two antonyms to represent gradable adjectives***

> Gradable concepts are represented by a value along a slider with at the ends the two extremes/antonyms. Since these antonyms often do not make sense when presented alone, the icon for the adjective shows both extremes. However, if one extreme can be inferred, then only the most important extreme is shown (Figure 57):

Figure 57: Gradable adjectives: straight-bend, round-flat, blunt-sharp, wet-dry, and sticky.

Principle 4.22: ***Showing one of our senses to represent adjective categories**

> Most adjectives generally have to do with one of our senses or our ability to evaluate or compare thing. For example, we can see something is dark or red, we can feel is something is sticky, we can taste if is something is sweet, or we can evaluate is something is right or if someone is touchy. These categories can be shown by showing the body limb used to perform these actions (Figure 58):

Figure 58: categories representing: properties one can taste, properties one can see, properties one can smell, properties one can hear, properties one can evaluate, and properties one can feel.

4.4 The Domain of Cooking, Eating and Drinking

In order to show the viability of a computer-based iconic communication language like VIL is, we have decided to implement one particular domain. Our initial goal is not to implement a whole system, but just to show that it is feasible to communicate with icons. Another reason for a limited domain was to allow testing and evaluation. The domain we have chosen to implement is that of eating and cooking, since this is one of the domains that people talk about a lot in simple day-to-day conversation. This section describes the full vocabulary used in the domain along with examples of icons from the iconicon.

The domain implements part of the hierarchies for verbs, nouns, and adjectives, as described in chapter 3. However, the program preserves the category structure for nouns, verbs, and adjectives as we have described them in the previous chapter. This means that in the tree view, all sibling categories are shown, but only the ones for which we have implemented sub trees can be selected. The icons that can be selected, i.e. of which sub trees are implemented, are enabled and the sibling categories that cannot be selected are disabled, but visible. Disabled icons are shown in gray instead of black and white.

Figure 59: The root of the verb tree with categories: verbs involving an exchange of exchange of possession (not implemented), existence verbs, identity verbs, verbs involving an involuntary activity (not implemented), verbs involving a mental activity (not implemented), physical transfer and location verbs, and verbs involving a voluntary activity.

In the Figure 59 only the categories for existence, physical transfer and location, and voluntary activity are enabled.

This section only describes icons from the domain of cooking, eating and drinking.

4.4.1 Nouns

Almost all of the nouns for our domain fall under the top-level (root) categories **physical world** and **science & technology** (third and fourth icon):

Figure 60: The root of the noun tree with categories: arts & architecture (not implemented), communication (not implemented), physical world, science & technology, beliefs & society (not implemented), sports (not implemented), and transportation (not implemented)

Within *physical world* the nouns fall under *living world* (third icon):

Figure 61: The categories in physical world: geography (not implemented), geology (not implemented), living world, universe & space (not implemented)

Within *science & technology* the nouns fall under *household* (third icon):

Figure 62: The categories in science & technology: computer (not implemented), electronic communication (not implemented), household, machines (not implemented), time & measurement (not implemented), physics & chemistry (not implemented), and transportation (not implemented)

4.4.1.1 Physical World: Living World Categories

In *living world* we find, among other categories,
- Animals
- Fruits
- Humans
- Vegetables

Figure 63: The categories in living world: animals, evolution (not implemented), fruit, humans, micro organisms (not implemented), plants (not implemented), and vegetables

The sub-categories for the *animals* category are, among others:
- Birds
- Fish
- Insects & Crustaceans
- Mammals
- Mollusks
- Reptiles

Figure 64: The categories in animals: birds, fish, insect & crustaceans, mammals, mollusks, pets (not implemented), and reptiles & amphibians

The implemented sub-categories for *humans* are:
- Food

- Household

Figure 65: The categories in humans: body (not implemented), food, household, people (not implemented), and psychology (not implemented)

The *household* category is cross-classified and added for convenience. It really belongs with s*cience & technology*.

The *food* category contains the following sub-categories:
- Dairy products
- Grain products
- Cooking liquids
- Drinking liquids
- Dishes
- Meals
- Spices

The *spices* category is cross-classified and really belongs to *plants*. The *dishes* category contains the following sub-categories:
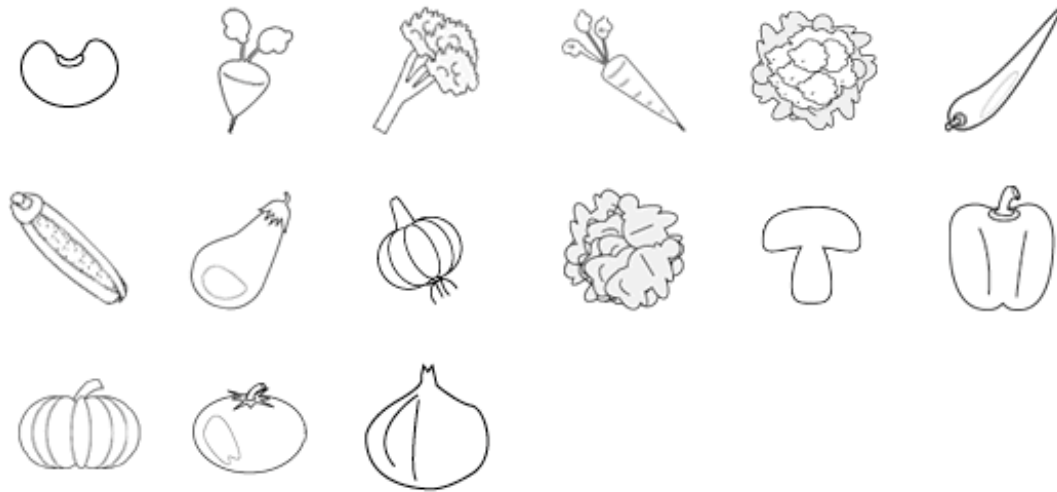- Cake, candy, & deserts
- Entrees
- Fast & fried food
- Fish dish
- Meats
- Sandwiches

4.4.1.2 Physical World: Living World Terminals

The above mentioned categories contain the following terminals:

Vegetables: been, beet, broccoli, carrot, cauliflower, chili pepper, corn, eggplant, garlic, lettuce, mushroom, pepper (paprika), pumpkin, and tomato.

Fruits: apple, banana, cherry, coconut, grape, kiwi, lemon, peach, pear, pineapple, and strawberry.
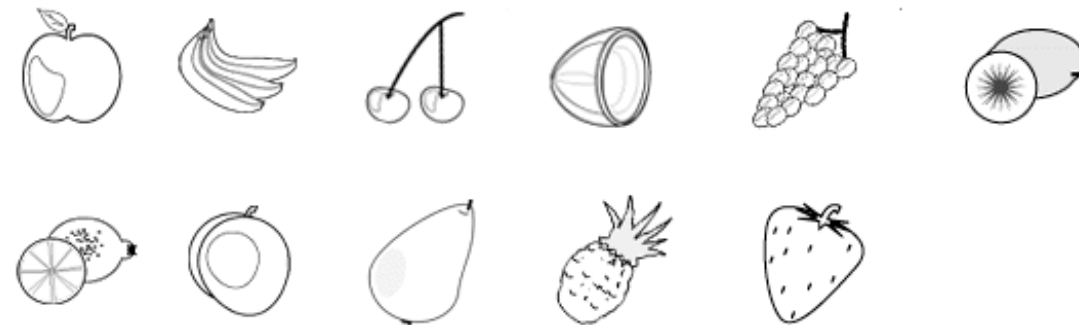
Birds: chicken, duck, ostrich, penguin, rooster, turkey

Fish: flat fish , shark, swordfish, tuna, whale

Insect & Crustaceans: crab, fly, grasshopper, shrimp, lobster, spider

Mammals: cow, elephant, horse, pig, rabbit, sheep

Mollusks: shell, snail, squid

Reptiles: alligator, frog, lizard, snake, turtle

Dairy products: butter, cheese, egg, honey, milk

Grain products: bread, bun, croissant, bagel/donut, French bread, fusile ("spiral macaroni"), muffin, tortellini

Spices: garlic, ketchup, mustard, pepper, salt, tomato paste

Cooking liquids: butter, sunflower oil, olive oil, tomato paste

Drinking liquids: beer, carbonated beverage, champagne, coffee, tea, water, wine

Meals: breakfast, dinner, lunch

Sandwiches: sandwich, hamburger, hot dog

Entrees: salad, soup

Meats: kebob, pork chop, ribs, rolled meat, sausage

Fish dish: sushi

Fast & fried food: hamburger, fried egg, fondue, French fries, hot dog, pizza

Cake, candy, and deserts: chocolate, cookie, ice cream, lollipop, cake, ice popsicle (water ice on a stick)

4.4.1.3 Pronouns

Users can also select a pronoun from the pronouns I, you, he/she/it, we, you all, they:

___

4.4.1.4 Science: Household Categories

Under *household* we find, among other categories,
- Kitchen

Figure 66: The categories in household: cleaning (not implemented), furniture (not implemented), kitchen, and personal articles (not implemented)

The *kitchen* category has sub-categories:
- Electric machines
- Cooking utensils
- Pots & Pans
- Silverware & eating utensils

4.4.1.5 Science: Household Terminals

The above mentioned "kitchen" categories contain the following terminals (if there are too many items in a category, then only 6 items are shown):

Electric machines: beater, blender, clock, coffeemaker, rice cooker, toaster.

Cooking utensils: bottle opener, bowl, cake server, can opener, colander, corkscrew, scale, spatula, jar, kettle, lemon squeezer, measuring cup, peeler, pizza slicer, and whisk.

Pots & Pans: pot, grill, bunt pan, loaf pan, pie pan, frying pan, Chinese pan ("wok").

Silverware & eating utensils: bottle, bread knife, Chinese kitchen knife/cleaver, chopsticks, cup, fork, wine glass, drinking glass, knife, plate, rice scoop, soup cup, spoon.
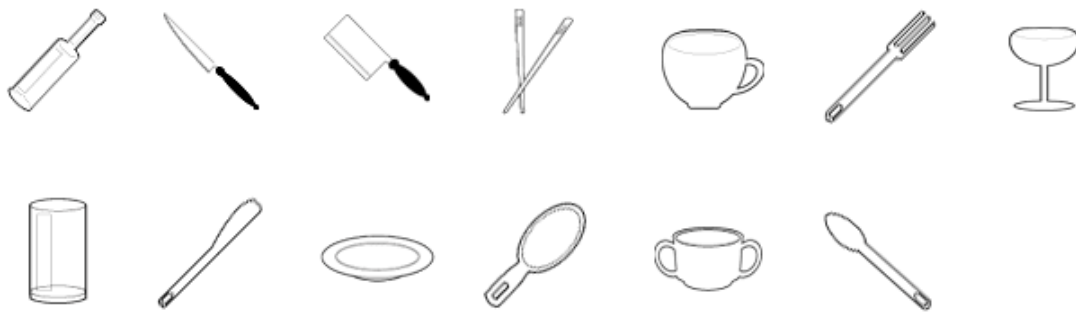
4.4.2 Verbs

The top-level (root) categories that are implemented for the domain are **voluntary activity** (eventually leading to verbs like eat, drink, etc.), **physical transfer & location** (eventually leading to verbs like put, pour, etc.), and **existence** (eventually leading to verbs like cut, chop, mix, stir, cook, bake, etc.).

Figure 67: The root of the verb tree with categories: verbs involving an exchange of exchange of possession (not implemented), existence verbs, identity verbs, verbs involving an involuntary activity (not implemented), verbs involving a mental activity (not implemented), physical transfer and location verbs, and verbs involving a voluntary activity.

4.4.2.1 Voluntary Activity Categories

In *voluntary activity* we find, among other categories,

- Expel
- Ingest

Figure 68: The voluntary activities categories: verbs of exerting a force (not implemented), expel verbs, hold/keep verbs (not implemented), ingest verbs, verbs of motion without changing location (not implemented), sports verbs (not implemented), and touch and poke verbs(not implemented).

4.4.2.2 Voluntary Activity Terminals

The above mentioned categories contain the following terminals.

Ingest: breathe, chew, drink, eat, lick, suck, swallow.

Expel: breathe, burp, cough, spit, vomit.

4.4.2.3 Existence Categories

Under *existence* we find (along with the exist category) the categories of *create* and *destroy* verbs:

- Create
- Destroy

The category of *create* verbs has (among sub-categories like calve, present, record, and reveal) the following sub-categories:

- Connect
- Transform

Figure 69: The create categories: calve verbs (not implemented), connect verbs, verbs of creation (not implemented), present/show verbs (not implemented), record verbs (not implemented), reveal verbs (not implemented), and transform verbs.

The category of *destroy* verbs has (among sub-categories like destruction, disappear, disconnect, and kill) the following sub-category:

- Cut

Figure 70: The destroy categories: cut verbs, destruction verbs (not implemented), perish verbs (not implemented), disconnect verbs (not implemented), and kill verbs (not implemented).
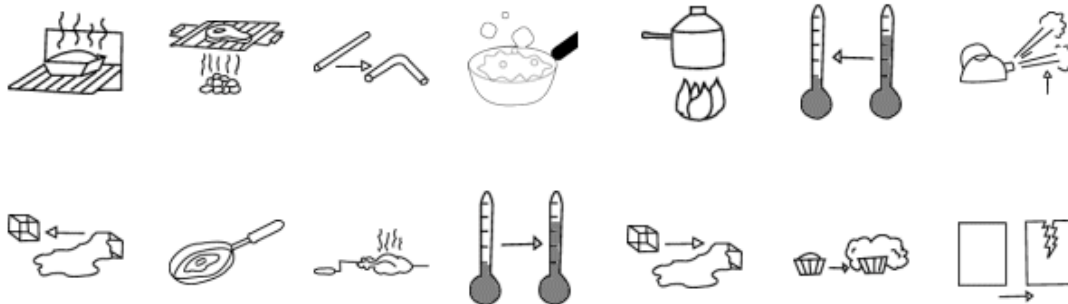
4.4.2.4 Existence Terminals

The above mentioned "create" and "destroy" categories contain the following terminals.

Connect: marinate, mix, puree, stir, whip.

Some of the verbs mentioned are de-nominal, e.g. to whip can also be the verb version of a beater.

Transform: bake, barbeque, bend, boil, cook, cool, steam, freeze, fry, grill, heat, melt, rise, and tear.

Cut: chop, cut, dice, grind, peel, slice.

4.4.2.5 Physical Transfer & Location Categories

Under **Ptrans & Location** we find the category of **put** verbs.
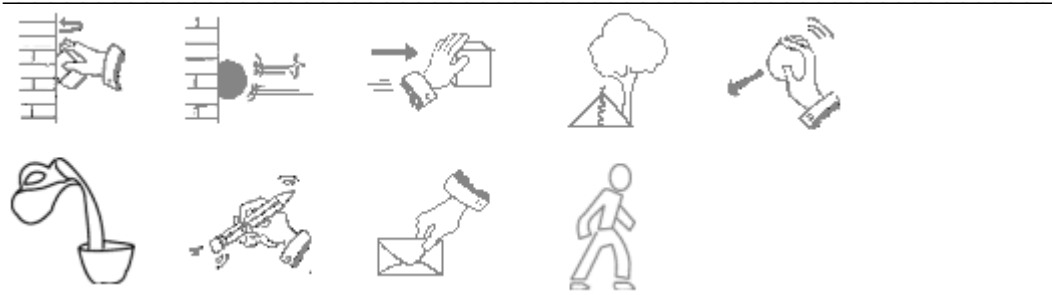
- Put

Figure 71: The physical transfer and location categories: conceal verbs (not implemented), verbs of contact by impact (not implemented), verbs of exerting a force (not implemented), lodge verbs (not implemented), propel verbs (not implemented), verbs of putting and pouring, erase and remove verbs (not implemented), verbs of sending and carrying (not implemented), and ptrans verbs (not implemented).

4.4.2.6 Physical Transfer & Location Terminals

The above mentioned physical transfer and location categories contain the following terminals.

Put: empty, fill, pour, put, spray.

4.4.3 Adjectives

All nouns have the possibility to modify their properties by using adjectives. Adjectives that can be added will fall into the category of perceptive adjectives. The perceptive categories (root) interesting to our domain are:

- Visual
- Color
- Tactile
- Olphactory
- Taste
- Measurement

Figure 72: The root of the adjective tree with categories: aural (not implemented), taste, visual, mental (not implemented), smell, tactile, measurement, and color.

The above mentioned categories contain the following terminals.

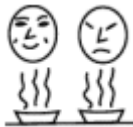**Visual**: empty-full, dark-light, round-flat, straight-bent

Color: black, blue, brown, gray, green, orange, pink, purple, red, yellow, white

Tactile: cold-warm, light-heavy, round-flat, blunt-sharp, sticky, straight-bent, wet-dry

Olphactory: fragrant -smelly

Taste: bitter, sour, hot, sweet

4.4.3.1 Measurement Categories

Under *measurement* we find categories:
- Size/Dimension
- Distance
- Temperature
- Volume
- Weight.

4.4.3.2 Measurement Terminals

The above mentioned categories contain the following terminals.

Size/Dimension: big-small, short-tall, minimum-maximum, thick-thin

Temperature: Celsius, Fahrenheit

Volume: liter, deciliter, centiliter, pint, quart, fluid ounce, gallon, cup, table spoon, tea spoon

Weight: kilogram, gram, pound, ounce, handful

In the next version of VIL, we will build in some conversion capability, maybe some sort of calculator, to convert one measuring unit to another.

4.4.4 Program Icons

Besides visual elements, icons for the grammatical entities, and icons for the vocabulary, we also need icons that are used in the computer program, which implements VIL. These are general toolbar icons like in any other program and some more specific icons.

4.4.4.1 Toolbar icons

There are three toolbars used in the VIL program:

- **Main screen toolbar**: the toolbar icons in the main screen perform the following functions:
    - New message
    - Open message
    - Save message
    - Print message
    - Update message
    - Send message
    - Help
    - Exit program

The toolbar icons light up (one at a time) when the user moves over them with the mouse:

- **Iconicon browser screen toolbar**: the toolbar icons in the browser screen for the iconicon perform the following functions:

    - Go up one level
    - Show history (actually last 10 picked icons)

185

---

o Print current icon page
o Close window

The toolbar has one additional icon when selecting nouns:

o Show pronouns

- **Noun properties screen toolbar**: the toolbar icons in the noun properties screen perform the following functions:

  o Add adjective
  o Delete adjective
  o Add noun
  o Delete noun

4.4.4.2 Other program icons

Other program icons are:
- Browse button, which allows the user to select a different language for tooltip help:

Having the user select a language of his/her choice does not necessarily make the program less universal. The composed messages are still icons only. It does, however, greatly enhance learnability, encodability, and even decodability. See chapter 5 for a more detailed discussion on tooltip help.

- Flags representing the language for the tooltip help:

- Confirmation and Cancel buttons for closing screens:
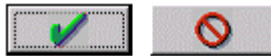
The next chapter discusses the program in detail. It also explains how to compose a message.

5. VIL: The program

This chapter describes the computer program accompanying the language in detail. In section 5.2, a sample message is composed and in setion 5.3, the architecture is discussed.

5.1 Introduction

The program exactly follows VIL's grammar. So, it is verb centered and the verb determines what its potential arguments are. Composing a message in VIL is not just picking some icons from the iconicon and putting them in some order. The grammatical elements in VIL are order independent, i.e. there is no rule stating that the subject must precede the verb in placement on the screen. Unlike spoken/written languages, VIL is non-linear. For consistency and speed of encoding and decoding, the icons will be placed in a predefined (default) location. If wished, they may be moved around, since clear visual elements will specify which icon represents which grammatical entity. The only order that VIL knows is that the verb has to be picked before other grammatical cases, like source or goal. This is because VIL is verb centered and the verb determines which potential arguments will be shown on the screen for the user to instantiate. So, composing a message consists of a sequence of steps. This also makes VIL less prone to syntactical errors than normal languages, i.e. it is for instance not possible for one message to have two verbs as in the syntactical incorrect "I give swim a book".

5.2 VIL Message

In this section, we describe how the system works by composing the following message:

"I put two porkchops in a frying pan."

When composing a message in VIL, several steps need to be taken:
- Select tooltip language
- Select a message type
- Select a verb
- Select NP's made available by selected verb
- Select a time when the event takes place
- Select a location where the event takes place
- Fill out remaining parts of message

- Finish the message

The steps after the selection of the verb do not need to be taken in the order specified. One can fill in the location before the NP's, if one wishes to do so.

5.2.1 Select tooltip language

The login screen is added to allow the user select the language (default English) he/she wants the tooltips and hypertext help in.

Having the user select a language of his/her choice does not necessarily make the program less universal. The composed messages are still icons only. It does, however, greatly enhance learnability, encodability, and even decodability:

- Learnability: the user learns the system and his/her way around the application much faster
- Encodability: the user composes messages faster
- Decodability: the user decomposes the messages faster

In the words of Stuart Mealing, we tend to look at tooltips as VAT, meaning Value Added Text.[154] Furthermore, Yazdani argues to consider the combination of text and icon. He states "Is it not true that we see in our everyday life people combining spoken language with gestures, hand and eye movements, intonational variations?"[155]

The advantages of tooltips are most clearly available in a learning phase. After the user gains more experience, the tooltips are less likely to be used. By clicking the "browse the world" button, the user is presented with a screen identifying the (current) languages available for tooltips and help:

---

[154] **Mealing, S.** *"Do you see what I'm Saying,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

[155] **Yazdani, M.** *"Communication through icons,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

For now, only English, Dutch and Spanish are supported. After the user has selected the language of their choice, VIL's main screen comes up.

5.2.2 VIL's Main Screen

VIL's main screen, the message composition screen, contains a toolbar, two containers, and a status bar. In version one, only one container will be used, namely the one for composing messages. The second container can hold previously composed messages, and is useful for reference when replying to a message or for copying items to the message being composed. It is read-only.

When the screen appears, it is initially empty. Let us start our message off, by selecting "New message" from the toolbar:

The toolbar is a *Coolbar* (TM Microsoft) type of toolbar (as in Internet Explorer). The images in it look disabled, but become colorized when you move over them with the mouse. VIL actually has two toolbars. The top one lets the user do things like:
- Compose a new message (**"New"**)
- Open an existing message ("**Open**")
- Save a composed message to disk ("**Save**")
- Printing a message ("**Print**")
- Updating a message ("**Update**")
- E-mailing/sending a message ("**Send**")
- Asking for help ("**Help**")
- Quitting VIL ("**Exit**")

The second toolbar holds the names of the five last composed messages, and lets the user open one by simply clicking on it:

By selecting "New" from the toolbar, a message type dialog comes up. This dialog asks the user what kind of message he/she will be composing:

- Declarative: a simple assertive sentence
- Interrogative: a question
- Imperative message: a command

In this case, we select the icon for declarative sentence. Once we have selected what type of message to compose, the screen gets filled with 4 icons. The top right corner contains a small icon denoting what the message type is for this message (in our case declarative). The icon of the clock is a placeholder for the time when the message/event takes place, and the icon of the globe is a placeholder for the location where it takes place. The fourth icon is the icon for verb or actually for rest of the sentence, i.e. kernel. This icon sets the wheel in motion. By instantiating this icon (i.e. selecting a verb), placeholders for the rest of the message are put on the screen. The verb determines what arguments are shown.
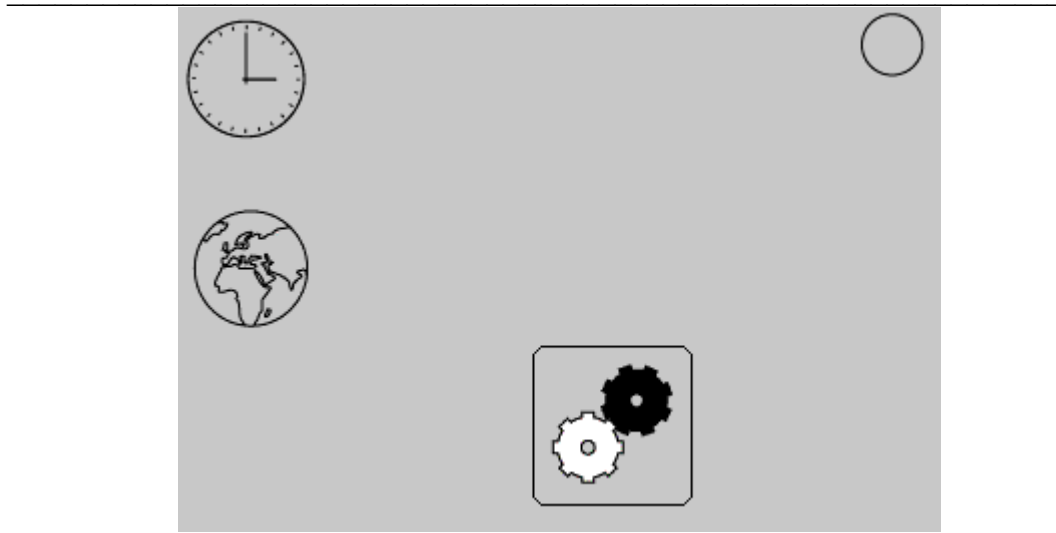
5.2.2.1 VIL Icons

In the previous chapter it has been explained in detail how a VIL icon looks and what its visual elements are. A VIL Icon that is not instantiated is square and has a size of 80x80 with a gray background. The following image shows uninstantiated VIL icons for verb and source:
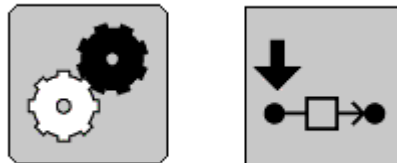
An instantiated VIL image has dimensions 95x80, where the main icon area is 64x64, and the rest is for visual elements. The following VIL icon shows the verb "to put". For an explanation of its visual elements, see chapter 4. The white background in the main icon lets us know that the verb has been instantiated:
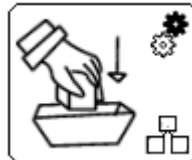
The large uninstantiated icon representing the grammatical case, in the above case "verb", becomes a much smaller visual element in the instantiated version.

5.2.3 Select a verb

When you double click on the verb icon in the main screen, a verb chooser comes up, showing the root or top-level categories of the verb hierarchy. This screen contains an icon view and a toolbar:

The leftmost icon on the toolbars shows you the immediate parent of the icons shown in the icon view. In this case we are showing the top-level, so the parent has been taken to be the icon for "verb" itself. The other buttons on the toolbar let you go up one level in the tree, show the ten last picked verbs (history), print the currently visible icons, and close the dialog.

By selecting "history" from the toolbar, the icon view shows the ten last selected verbs:
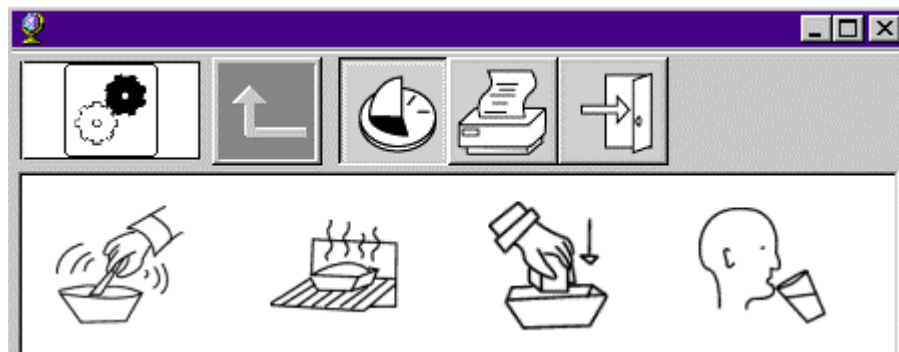
Here the last selected verbs were "to stir", "to bake", "to put", and "to drink".

Let us pick our verb "to put". It falls into the category of "verbs of physical transfer and location". Then we choose "verbs of putting & pouring". From the terminals, we pick the verb "to put" by double clicking on it, and the dialog closes and we return to the main screen. So, now that we have picked our verb "to give", let us see what the main screen looks like now.

5.2.3.1 VIL's Kernel

After we have picked the verb "to put" the main screen will load placeholders for the rest of the sentence. The arguments that it will load depend on the type of verb chosen. In this case the verb is of type "physical transfer" and takes as arguments an agent, a source, a theme/object, and a goal. The other grammatical entities that are put on the screen are instrument, aktionsarten, frequency of occurrence, and duration.
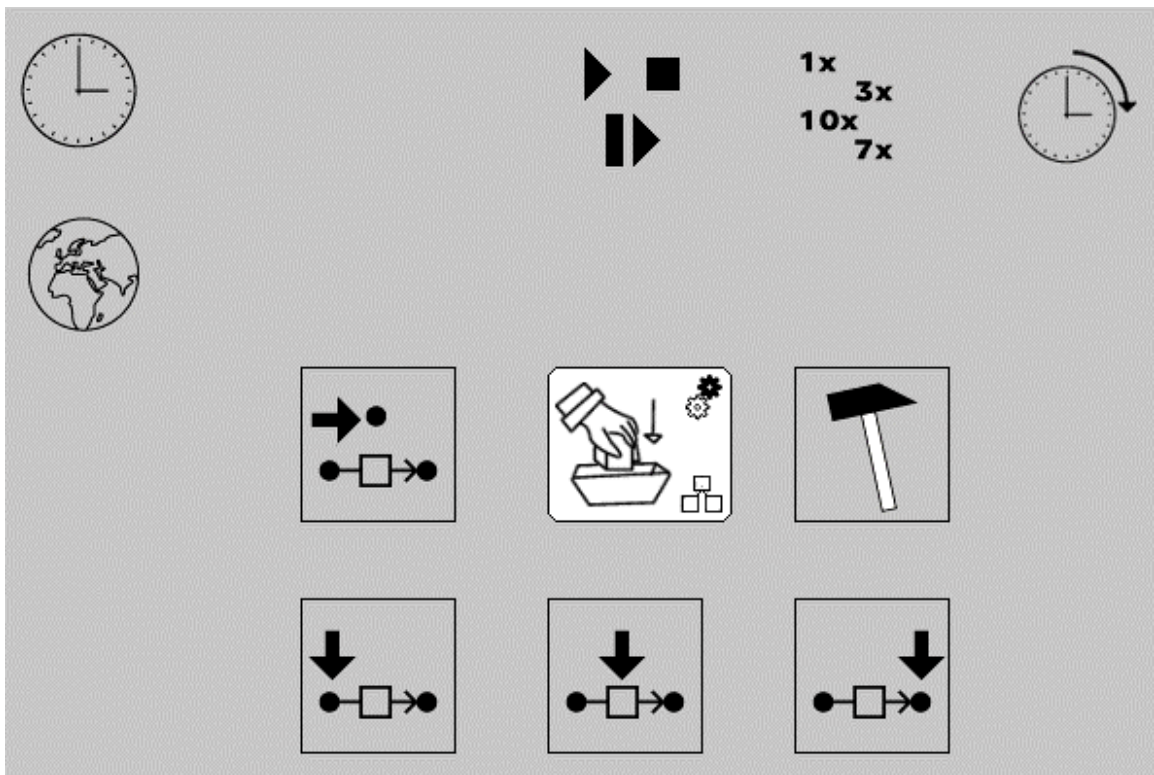
The "verb icon" gets instantiated: it contains the icon for "to put", and visual elements for grammatical entity (i.e. verb) and for category help. The tooltips of the icon also help in decoding the icon. Since it is instantiated, the background color of the icon changes from gray to white.
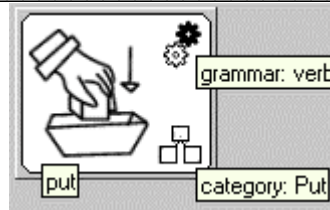
Let us instantiate the noun phrases for our sample sentence "I put two porkchops in a frying pan".

5.2.4 Select Noun Phrases (NP's) Made Available by the Selected Verb

As stated before, the verb *to put* takes as arguments:

- An agent
- A source
- A theme/object
- A goal/destination

These are shown as:
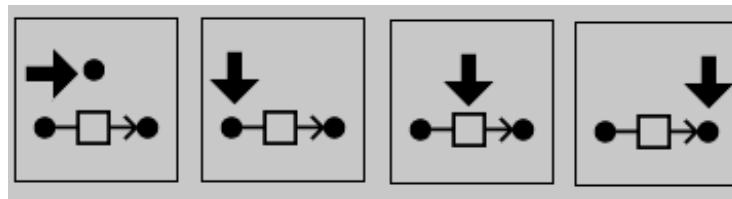
By double clicking on any of these icons, the user is taken to the noun chooser dialog. This dialog is the same as the verb chooser, but for nouns. Also, the toolbar has one extra button: show pronoun.
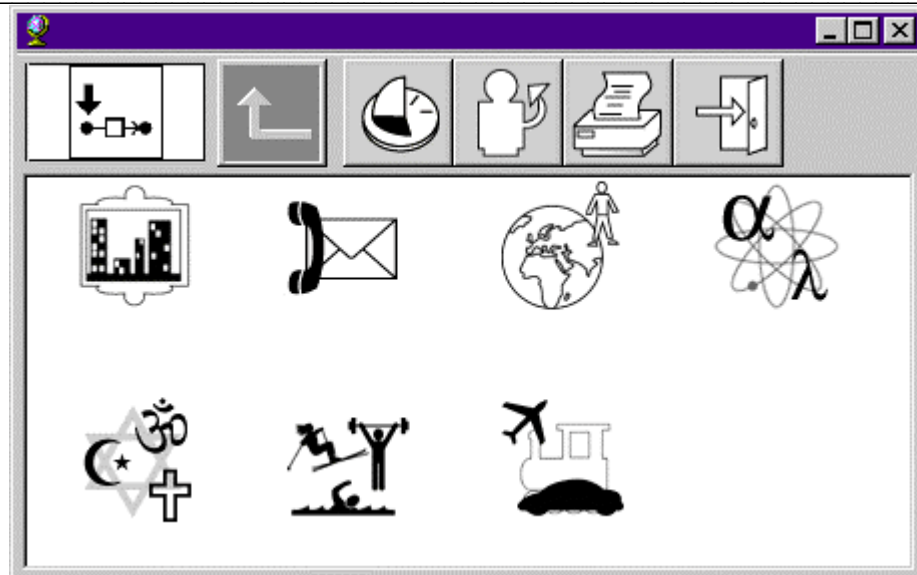
Let us instantiate our agent noun phrase: I. We click on the show pronoun button in the toolbar, and the icon view shows all pronouns: I, You, He, We, You All, and They:
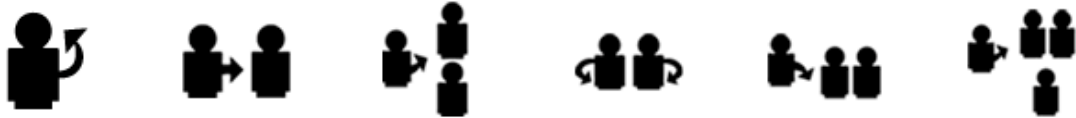
We choose the icon for "I" by double clicking it. Next we instantiate the goal noun phrase "a frying pan". We double-click the goal NP in the main screen to bring up the noun chooser dialog again. We follow the following path:

Science & Technology -> Household -> Kitchen -> Pot & Pans

We finally select "frying pan" from the pots & pans category. To select the theme/object "pork chop", we follow the following path, where we select "pork chop" from the category of meats:

Physical World -> Living World -> Humans -> Food -> Dishes -> Meats

Now that we have chosen icons (i.e. nouns) for our three arguments (agent, goal, and theme), let us see what the main screen (partly) looks like:
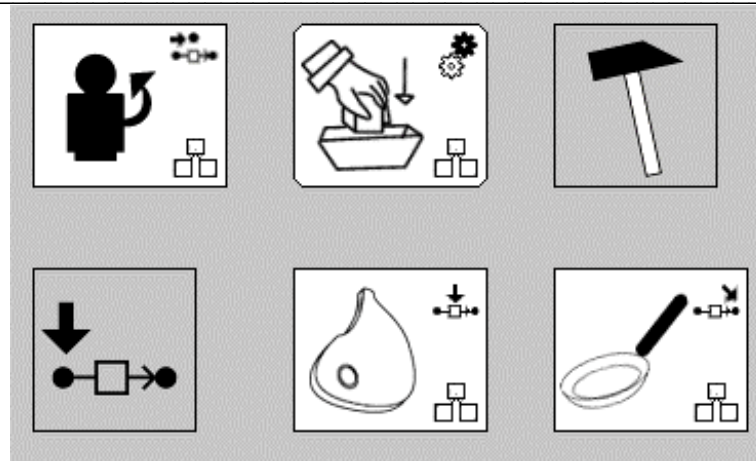
The border distinguishes the nouns from the verbs, and the grammar visual element distinguishes the cases from each other.

Now we can start completing the noun phrases by filling in the various properties for the nouns.

5.2.4.1 Select noun modifying properties

Now that we have chosen the nouns "I", "pork chop", and "frying pan", we can fill in the properties for them. For generality "I" is called a noun here even though it is a pronoun.

When a noun has been instantiated, its double-click behavior changes. Where double-clicking an uninstantiated NP brought up the noun chooser dialog, double-clicking an instantiated NP will bring up the noun properties dialog. In this dialog the user can modify the following properties about the noun:

- Nominal demonstrative
- Quantity
- Owner noun phrase
- Comparative noun phrase
- One or more adjectives.
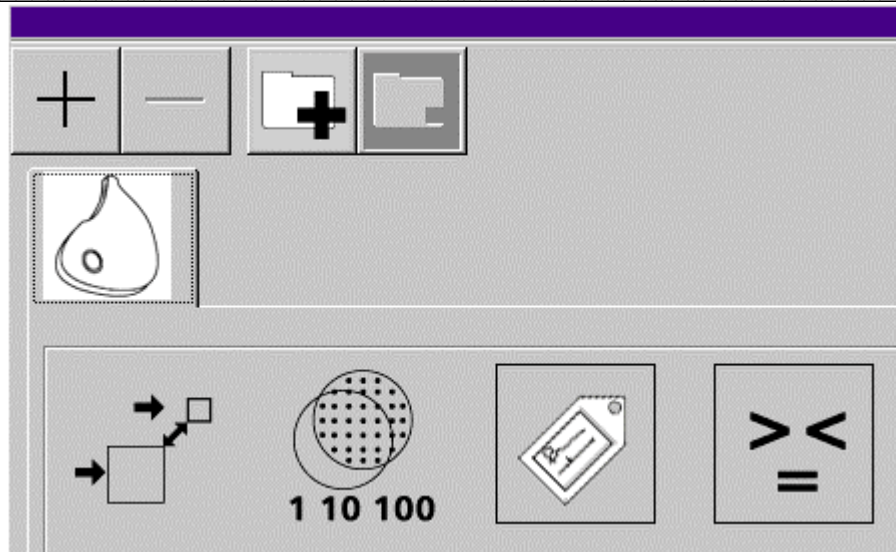
The toolbar contains buttons for adding an adjective, deleting an adjective, adding a noun, and deleting a noun. The dialog contains a tab strip, which contains a tab for each noun in the noun phrase if the noun phrase is a noun conjunction. If the noun phrase in our message was: "porkchop and sausage", then we had to add the noun "sausage" to the theme noun phrase by selecting "add noun" from the toolbar. An extra tab would have been added for the new noun for which we would also have been able to change properties separately (by bringing its tab to the front):
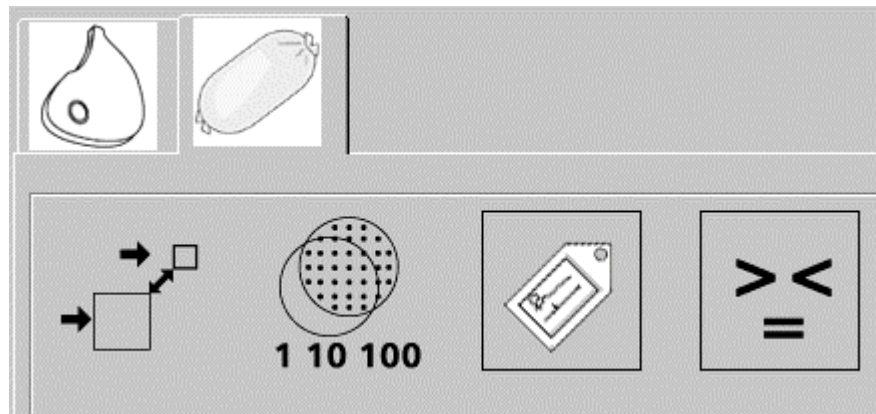
Now, let us fill in the properties for our three arguments. The agent "I", and the goal "frying pan" do not have modifying properties. So, let us concentrate on the theme NP "two porkchops".

The first icon placeholder in the tab is *nominal demonstrative*. When we double-click the icon, the following dialog comes up:
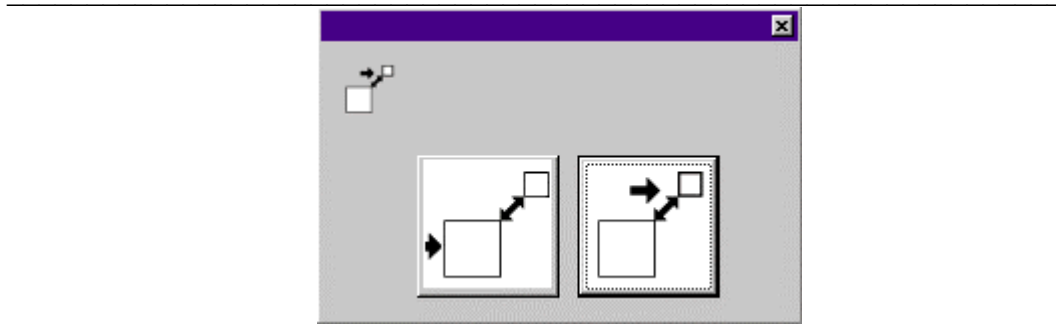
Here we can choose between "this" and "that". We do not need a *nominal demonstrative*, so we move on to *quantity*, and we double-click the "quantity icon". The following dialog comes up:
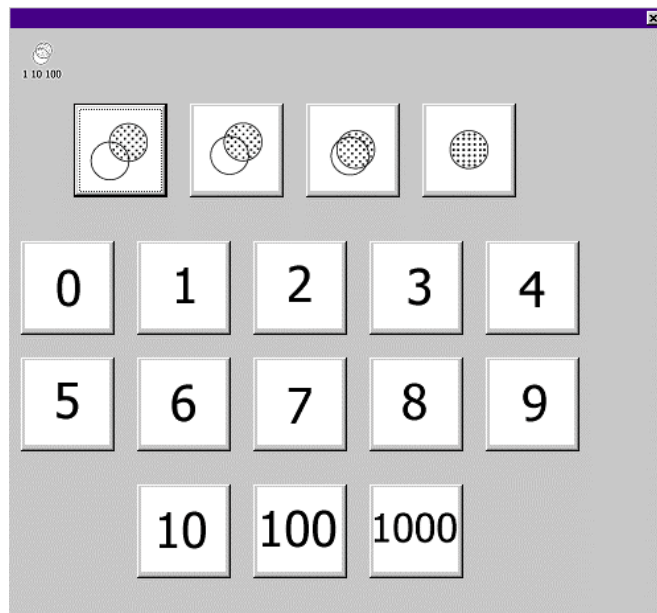
In this dialog we can choose between a quantifier like "some", "half", "most", and "all", or a number like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 1000. So, we choose "2". Now we have instantiated "two porkchops" and the noun properties dialog looks like:
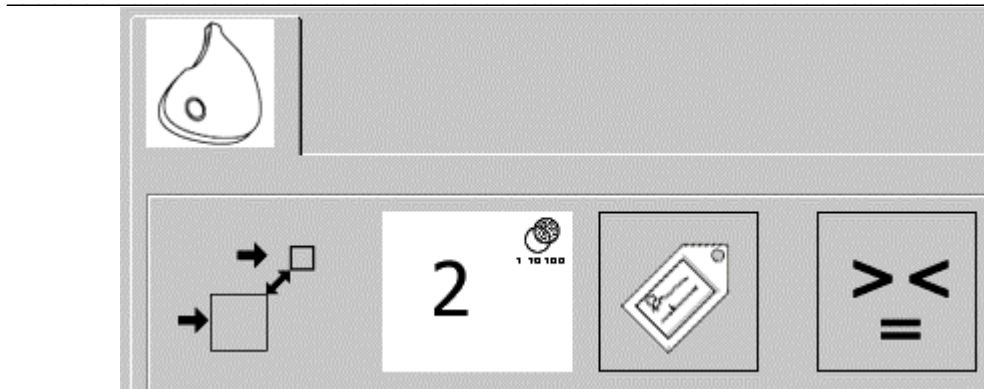
Next, we close the dialog and return to VIL's main screen. The theme NP icon gets and extra visual element (an asterisk) to let the user, but moreover the interpreter (i.e. decoder) of the message, know that the noun "porkchop" has modifying properties. The interpreter of the message can then double-click the icon to find out about the properties and find out that the theme NP is not just "porkchop", but "two porkchops":
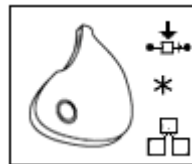
Should we wanted to have instantiated the property *comparative NP*, then double-clicking the "comparative NP icon" would have brought up the following dialog:
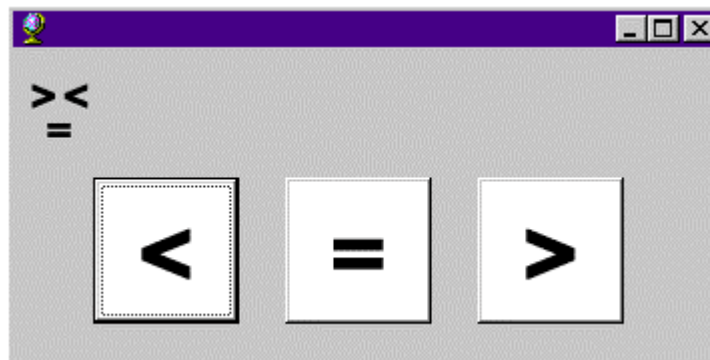
Consider the sentence "I put more porkchops than ribs in a frying pan ". In this case we would have had to make the theme NP "more porkchops than ribs ", in other words "more than ribs porkchops" (as in I put more than ribs porkchops in a frying pan). Then we would have chosen to instantiate the "greater than" icon of the

previous screen. The other choices in the comparative NP dialog are "less" and "as much/equal".

Now that the arguments for the verb have been filled in, we can complete our message by instantiating the rest of the sentence.

5.2.5 Fill out remaining parts of the message

Now that we have instantiated the verb and its arguments, let us complete the message by instantiating the rest of the message. Let us see what grammatical entities can still be filled in (remember that uninstantiated grammatical entities can be recognized by a gray background):
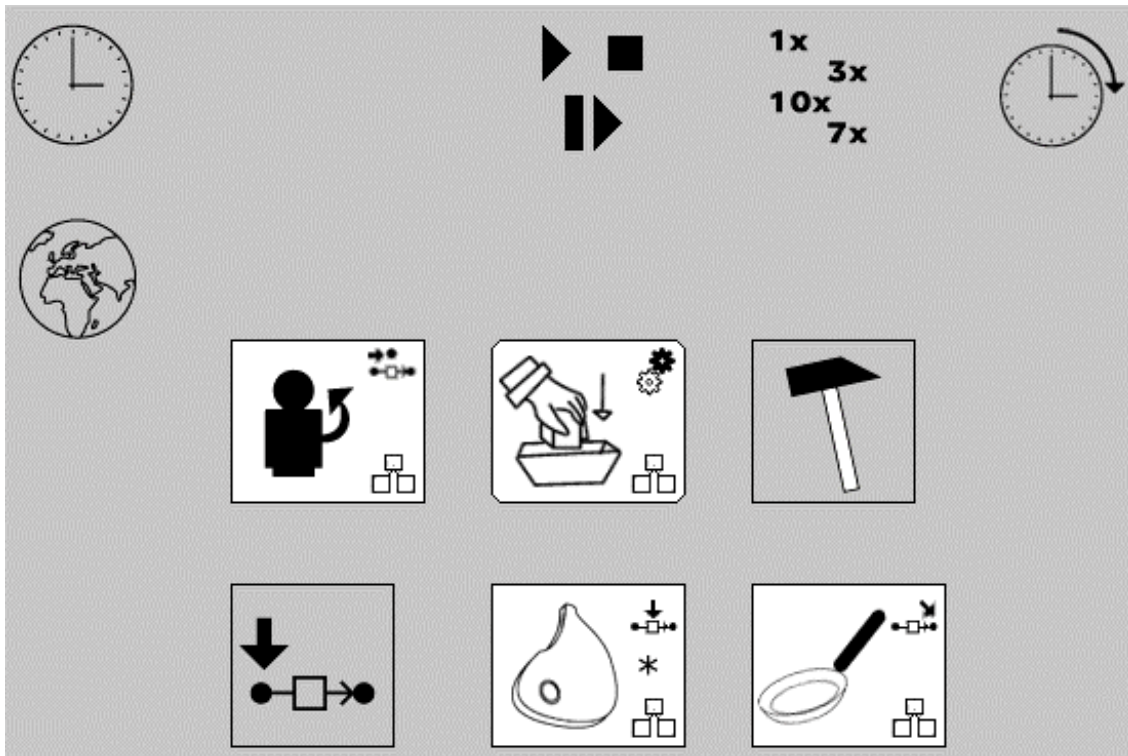
One of the remaining entities is *aktionsarten*, i.e. what is the status/progress of the event. The aktionsarten available are "start", "stop", "continue":
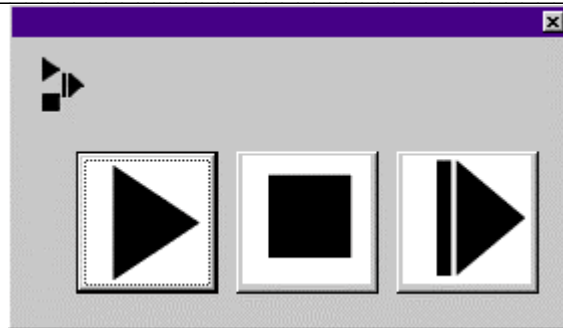
Next, there is *duration*, i.e. how long does the event take place:
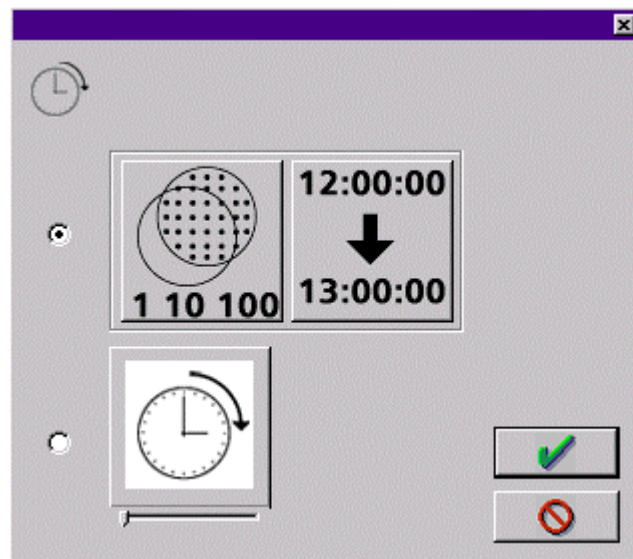
Duration is chosen by either selecting a quantity and a time unit (e.g. for 3 hours), or by a setting the "duration slider" to a specific value. Setting the slider in the leftmost position means "for a very short period of time", the slider in the middle means "for a normal/regular amount of time", and the slider in the rightmost position means "for a very long period of time".

Then there is *frequency of occurrence*, i.e. how many times does the event take place:
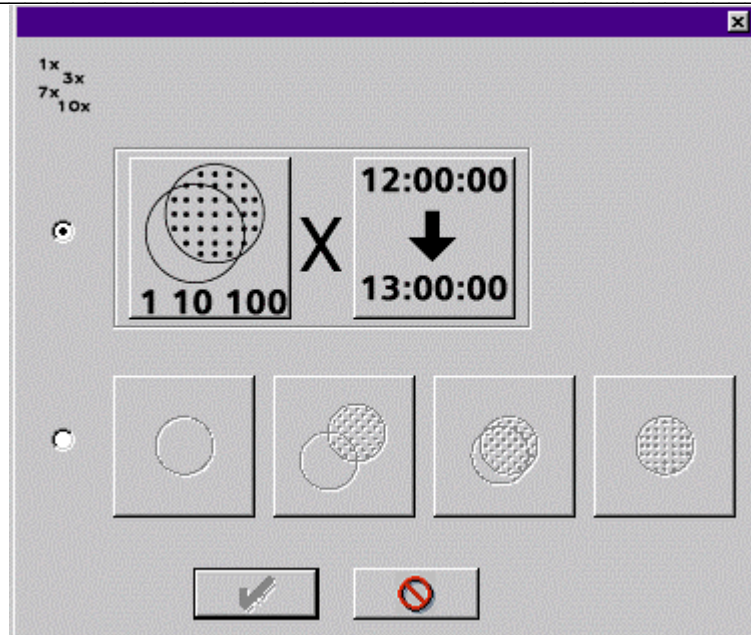
Frequency of occurrence is chosen by either selecting a quantity and a time unit with an X in between, as in "3 *times per* day", or selecting an adverb of frequency. Available adverbs of frequency are never, sometimes, often, and always.

Adverb/Manner has not been implemented yet. Then there is *instrument NP*, which describes what instrument was used in the message, as in "I put two porkchops in a frying pan *with a spatula*". Aktionsarten, duration, frequency of occurrence, adverb/manner, instrument, and are not needed to complete our message.

The next two sections describe the *time when* and *place where* cases.

5.2.6 Select a time when the event takes place

The "time when icon" is a placeholder for the time when the event being described takes place. Double-clicking the (uninstantiated) "time when icon" brings up the "time when dialog". This dialog consists of a *temporalizer* and/or one of the following: a *time pronoun*, a *time demonstrative* plus a *time unit*, a *time when sentence*, or a *time expression*:
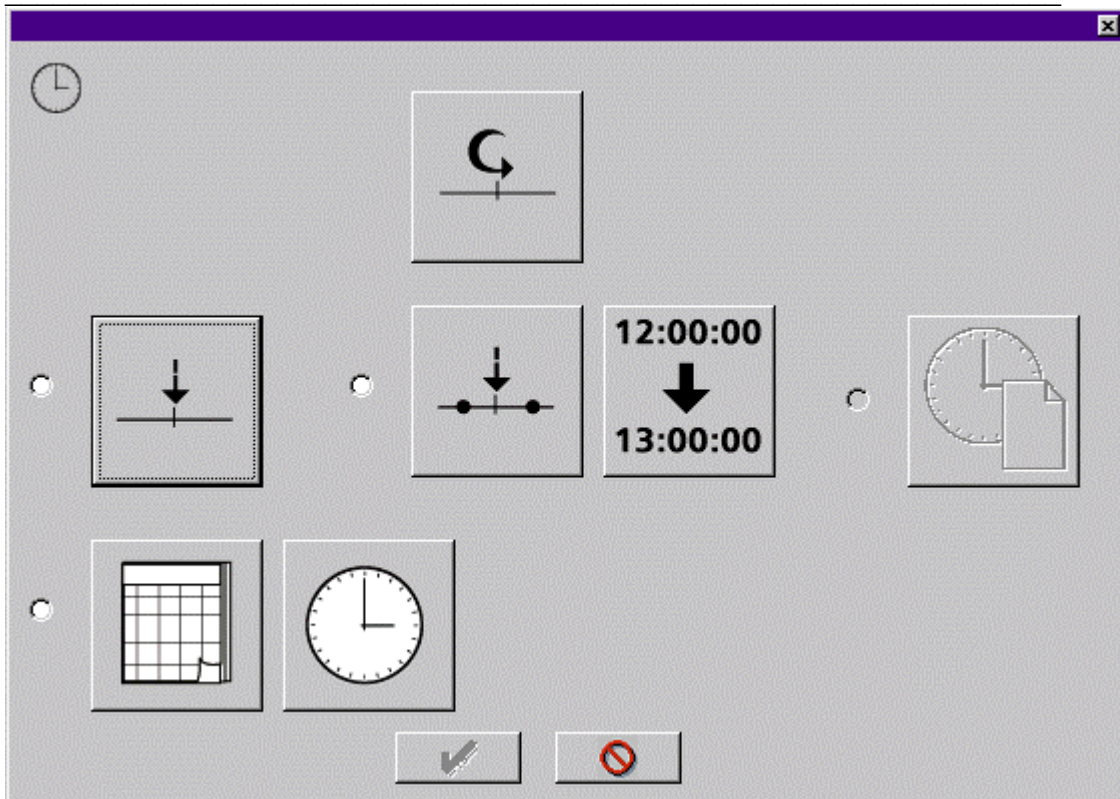
When one radio button is selected, the other options become disabled (in our sample message the user chose to instantiate the time expression). The *temporalizer* has as possible values: "before", "at/on/in", "after", and "while":
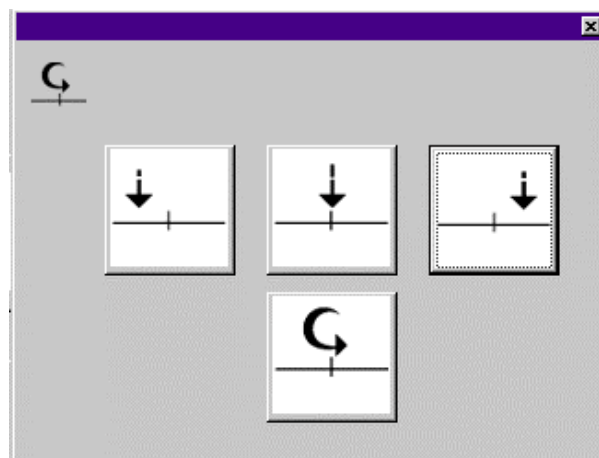
Probably the most used entity in the "time when dialog" is the *time pronoun*. If the user wants to make sentences like our example "I put (past tense) two porkchops in a frying pan ", then he/she would choose to instantiate the *time pronoun*. The *time*

pronoun allows for values like "long ago", "just a moment ago", "past tense", "present tense", "soon", and "future tense":
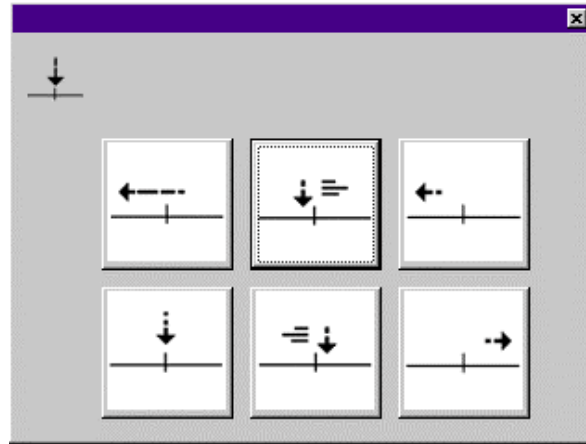
If the user chooses to instantiate the second option, then he/she has to instantiate a *time demonstrative* and a *time unit*. The *time demonstrative* consists of the values "last", "this", and "next":
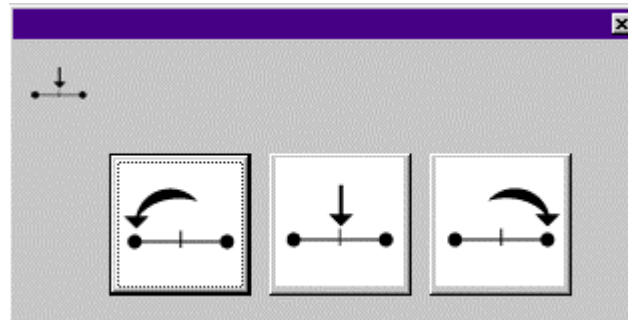
Then the user has to instantiate the *time unit*. The *time unit* is either a season, i.e. "spring", "summer", "autumn", or "winter", or one of the following: a "second", a "minute", an "hour", a "day", a "year", a "decade", a "century", or a "millenium". So in combination with the *time demonstrative* this could yield something like "this year" or "next winter":
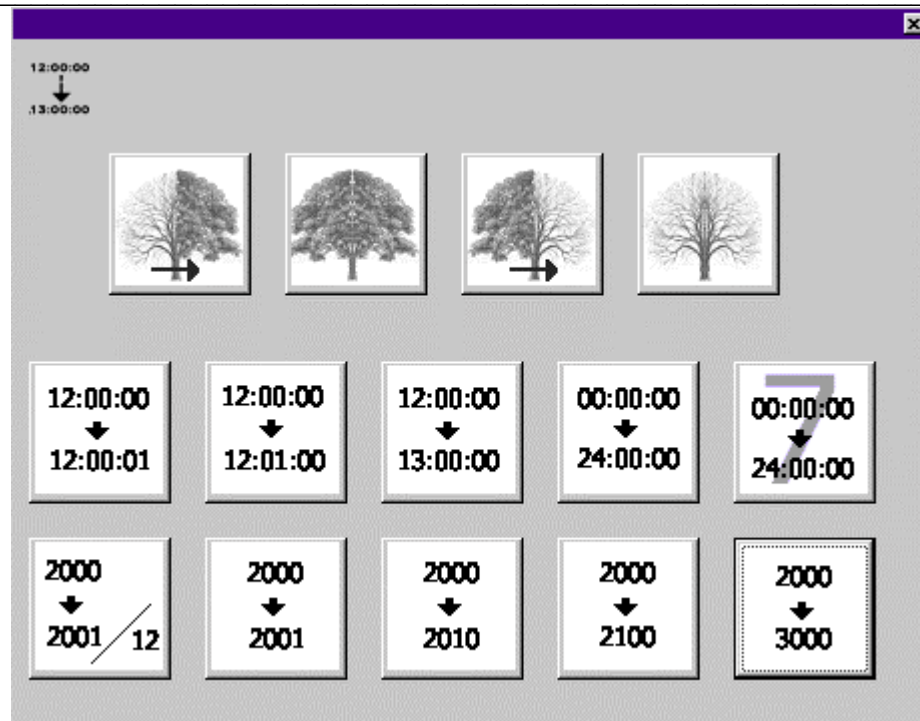
Double-clicking the "time when sentence icon" lets the user compose a new sentence, which together with the *temporalizer* could yield something like "before I went to the dentist" or "while I was eating". The last option in the "time when dialog" is a *time expression*. Here it is possible to, for example, choose an actual date or time. As opposed to the "time unit dialog" where only the word "hour" could be chosen, now "three o'clock" can be chosen. The date and time options are not iconic but graphical. The date is chosen from a calendar. It uses the system calendar, and even though it looks culture specific, it adapts to the operating system language. So for a Dutch person, the date originally selected by an American, is displayed in a Dutch way when the Dutch user runs a Dutch version of Windows:

The clock is also dynamic. It allows the hands of the clock to be dragged to the appropriate location to indicate the time wished (e.g. 7:25):
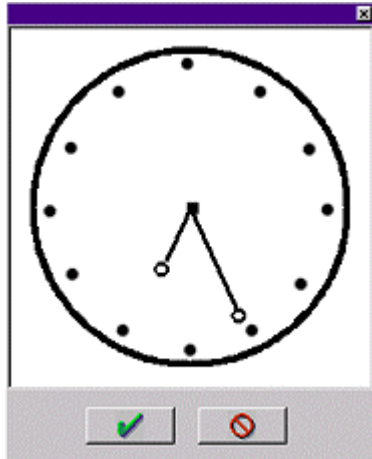
Let us instantiate the *time pronoun* for our sample sentence by picking the icon for "past tense". This completes our sentence. For completeness the next section discusses the *place where* case.

5.2.7 Select a location where the event takes place

The "place where icon" is a placeholder for the location where the event being described takes place. Double-clicking the (uninstantiated) "place where icon" brings up the "place where dialog". This dialog consists of a *localizer* and/or one of the following: a *place where NP*, a *locative pronoun*, a *location from a MAP*, or a *place where sentence*:
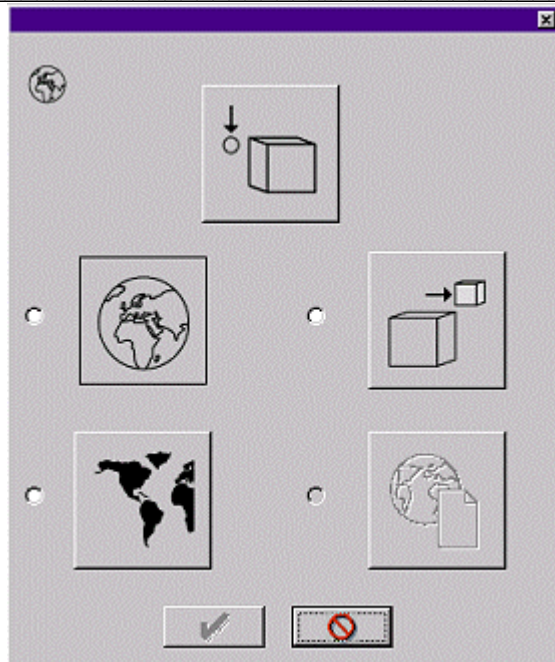
The *localizer* consists of one of the following values: "behind", "between", "in front of", "before", "at", "after", "to the left of", "to the right of", "above/over", "below/under", "inside", and "in the corner":
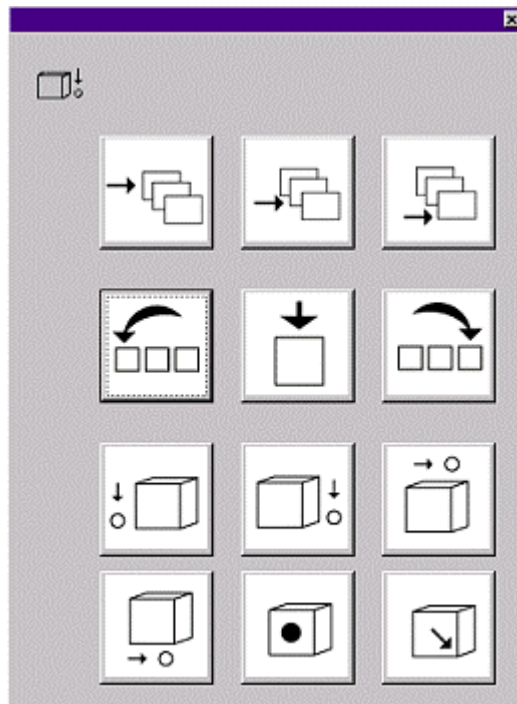
As with the *time when* case, selecting one radio button will disable all others. So, clicking on the place where sentence disables the MAP, locative pronoun, and "place where NP buttons". Clicking the "place where NP" allows the user to select a noun from the noun chooser dialog. This allows the user to make a location like "in the garden". Clicking the "place where sentence" will (in a future release) allow the user to compose a new sentence as *place where*. Clicking the *locative pronoun* button allows the user to select a *location pronoun* like "here" or "there":
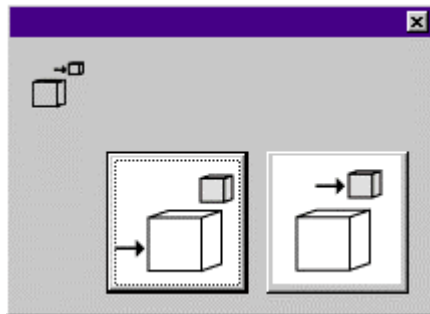
Lastly, it is also possible to select a location from a map. Clicking the MAP icon will bring up a picture of a map of the world. It is the possible for the user to click on any location on the map:

5.2.8 Finish the message

By now, we have instantiated everything necessary to compose our message "I put two porkchops in a frying pan". The VIL icons for the uninstantiated grammatical entities are still on the screen. This is unnecessary and even makes decoding more complicated. When we click on the "Update Message" icon in the toolbar,

all uninstantiated icons will be hidden. This way we are only left with the entities we want, i.e. the ones that we need to communicate our message. The following screen shows the completed and updated message "I put two porkchops in a frying pan". The icons for "Aktionsarten", "Frequency of occurrence", "Duration", "Manner/Adverb", "Instrument", and "Place Where" have been made invisible:
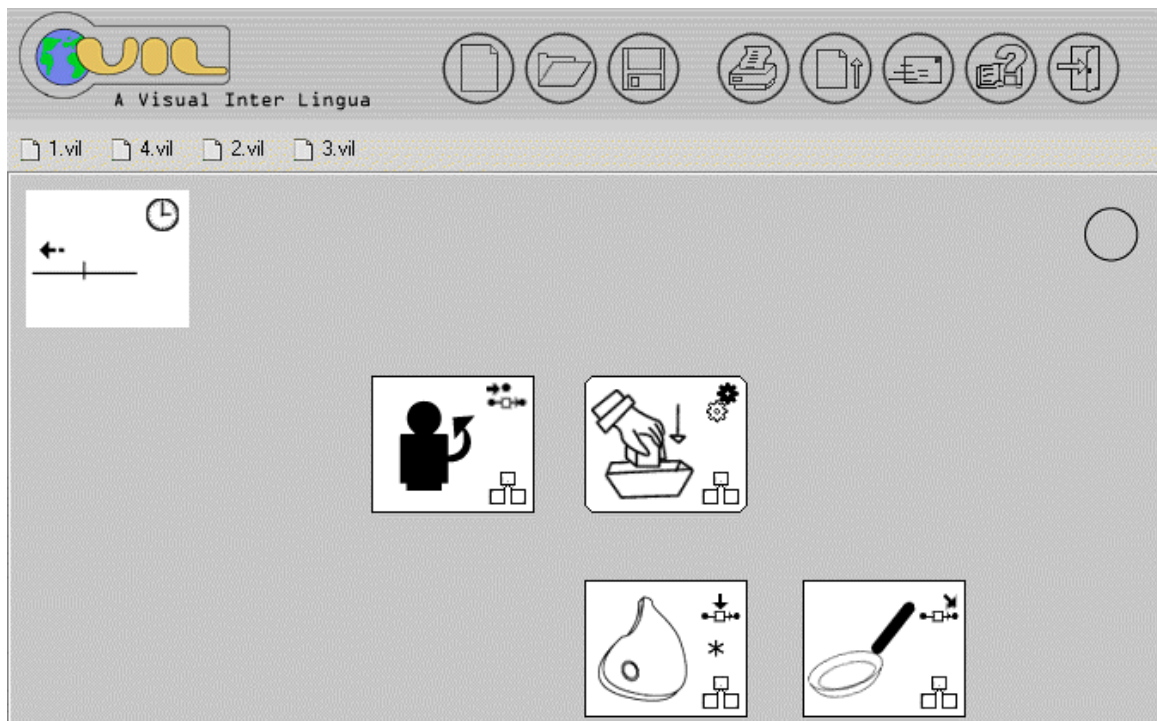
The message can be saved to disk, but can also be sent to someone else anywhere in the world by e-mail. The person receiving the file, just double-clicks it and, since the computer recognizes the ".vil" extension, it will startup VIL and load the message for him.

5.2.9 Decoding the message

The recipient would decode the message something like this: In the top-right corner he/she sees the sentence is declarative. Then he/she looks at the verb, recognizes "to put". When the meaning of a verb icon is not immediately clear, the user can right

click the grammar visual element. This leads to an animation being shown of the represented verb.

Next, the user inspects the slots for the agent, goal, and theme NPs. So, far the message "I put (present tense) a porkchop in a frying pan" is decoded. Then he/she looks at tense, and sees that a "past tense" is indicated. Now the sentence becomes "I put (past tense) a porkchop in a frying pan". Then the recipient looks at each NP in turn to see if any of them has an asterisk (visual element), indicating that the noun has modified properties with it. He/she sees that only the porkchop icon has an asterisk, double-clicks the icon, and finds that the NP is in fact not "a porkchop", but "two porkchops". Now the recipient is happy to have decoded the message "I put two porkchops in a frying pan".

5.3 Help

Throughout the chapter we have seen various methods of providing help that augment learnability, encodability, and decodability. These help features are in addition to VIL's specific features that help in encoding and decoding, such as borders, colors for instantiation, visual elements indicating case, and asterisks for inviting to further investigation.

They are summarized below:
- Animated help for verbs showing an animation of the action being performed

"Animation makes the meaning of the icon more self-evident. We can say that the icon automatically emphasizes its grammar and semantics by showing clearly that it signifies an action and by displaying the meaningful features of the action."[156]

- Language specific tooltips & help file
- Tooltips for each visual element
- The history button on the iconicon browser, showing the ten last picked icons
- Category help for verb, noun, and adjective showing the path to the selected icon and also siblings from its immediate parent category

The last form of help has not been shown yet and is addressed now.

---

[156] **Dormann, C.** *"Self-Explaining Icons,"* University of Brighton Grand Parade, http://www.intellectbooks.com/iconic/self/self.htm.

5.3.1 Verb, Noun, and Adjective Category help

Verbs, nouns, and adjectives are selected through an iconicon browser. When a verb, noun, or adjective is instantiated, a category visual element (representing a tree structure) is displayed in the bottom right corner of the icon. Right clicking this visual element results in a popup window showing the user where in the hierarchy the icon belongs, i.e. the path from the root of the classification in question to the terminal represented by the icon. To even further help clarify the icon, a set of icons that are siblings of icon is shown, i.e. these are icons that belong to the same category as the icon in question.

Next figure shows the popup category help for the verb "to put":
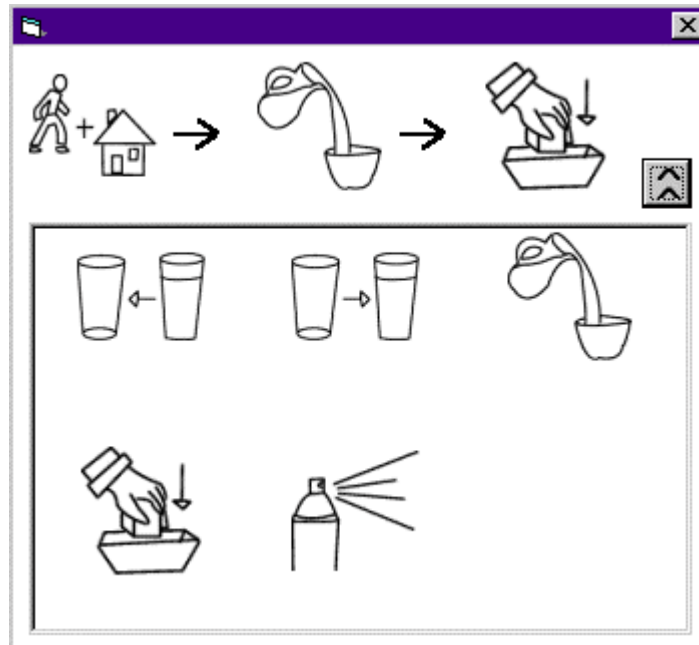
The upper part shows the path to the icon, i.e. physical transfer & location -> put & pour verbs -> to put. The lower part shows siblings from the put & pour category: *to empty, to fill, to pour, to put, and to spray*.

5.4 VIL Technical Architecture

The application is developed following the 3-tier model as much as possible. A 3-tier model is a model, in which the goal is to separate presentation, application logic, and data. The presentation layer knows a lot about how the user works and how to deal

with application results, but knows nothing about the application logic of the application. The application logic layer is responsible for transforming business and data access rules in services that can be used by the presentation layer to find the information it needs. The application logic layer has no idea how the user interacts with the interface or how the results are displayed to the user. This lack of user knowledge helps to keep the functionality of the application logic components general, which increases their potential reusability in other applications. The application logic components also use class modules to structure the logic of the application logic components and the data access rules in a way that helps in developing, debugging, readability, maintainability, and source code reusability. Figure 77 shows the architecture of the VIL application.
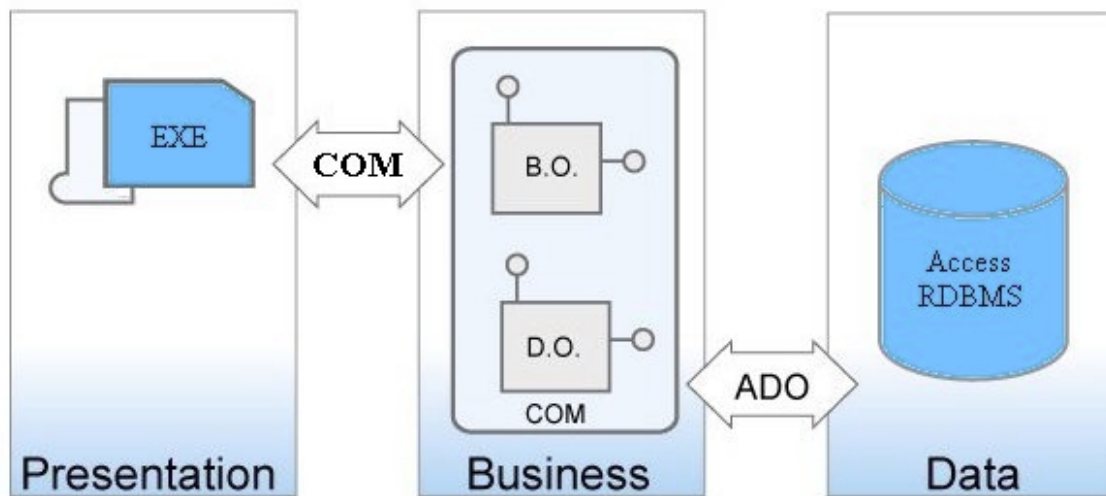
Figure 77. The architecture of the VIL application

The presentation layer consists of an executable. This executable communicates through COM (Component Object Model) with the application logic components. These application logic components are built as Visual Basic 6 ActiveX DLLs. The goal is to make the presentation layer as light as possible and designate all calculations and traffic to and from the database to the application logic layer. Of course, a web-or-page-based application contains a much thinner presentation layer than an executable (Figure 78). However, the complexity of the graphical presentation and user interface prevents a page-based application. It does not prevent an Internet version, since we can use Microsoft techniques like ActiveX controls and documents to create a version of VIL that runs within Internet Explorer. If we do this, then it is unfortunately not possible for Netscape users to use VIL. However, an Internet version could still add value, since more people living in more countries can make use of the language.
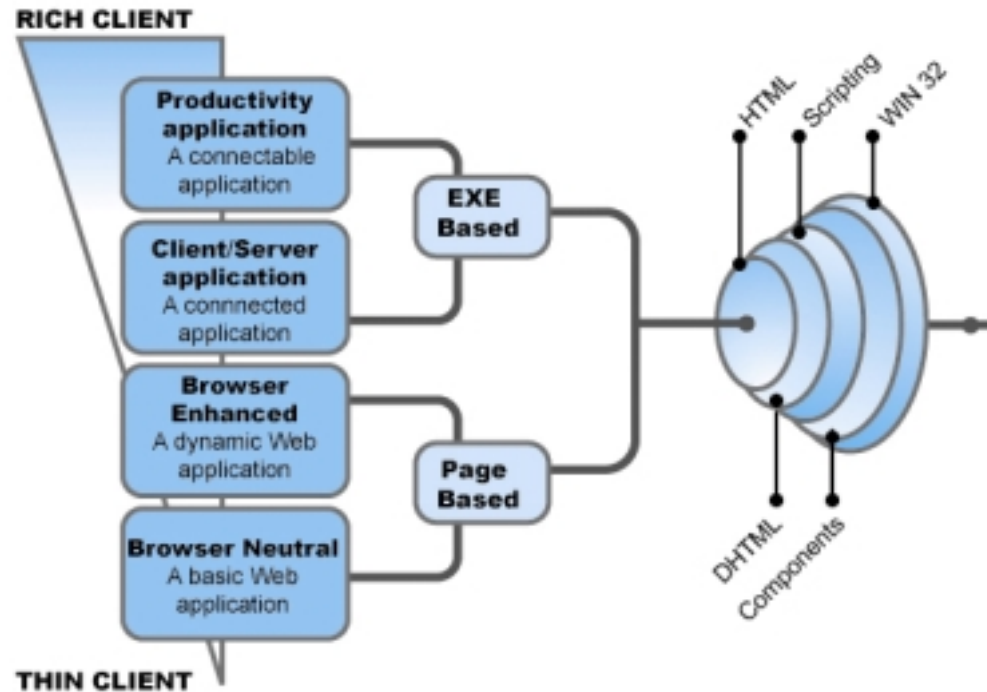
Figure 78. Thin and rich clients in the presentation layer

The advantage of splitting up presentation application logic, and data is that components can be developed separately and that the application logic components are reusable. The application logic components communicate through ADO (ActiveX Data Object) with the database. It does not matter which RDBMS (Access, SQL Server, or Oracle) is being used, ADO ensures a consistent interface to the database.

Advantages of component-oriented development are:
- Very scalable: new functionality can be added by just adding a new component.
- Reusable components: components can easily be put to work in other applications.
- Simplifies adaption to new technology: new components can be written that contain new developments in technology (e.g. new multimedia effects).
- Can be developed parallel: more people can work simultaneously on the application, each with their own expertise.
- Easy in maintenance: an error in a component means that only that specific component can be adjusted and that the rest of the application can remain intact (no recompilation necessary).

- Easy to move to internet: component development is the standard in Internet technology, which makes the step from an executable-based local application to an Internet application less complex.

Figure 79 shows two applications that share components on all three levels. One of these applications could be the VIL executable-based application, and the other could be the Internet based application.
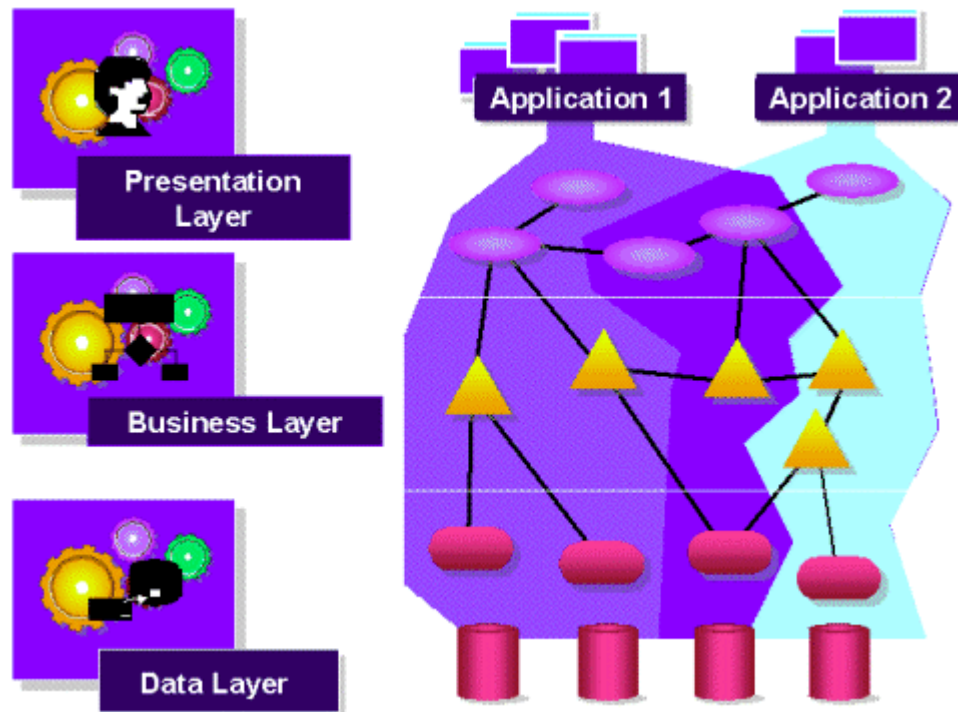
Figure 79. Two applications showing reuse of components in all three layers (courtesy of Microsoft Corporation).

5.4.1 Hardware and Software

VIL is written in Visual Basic 6.0 and runs on a Windows 95/98/NT 4.0 platform. It makes use of ActiveX component technology. The VIL grammar component is an ActiveX DLL, and the VIL icons and VIL toolbar icons are ActiveX controls (OCX). All the information concerning the grammar of the language VIL as well as the iconicon, including all images, is stored in a Microsoft Access 2000 database. VIL uses ADO with ODBC to communicate with the database.

6. Evaluating and Testing VIL

In this chapter we evaluate VIL. We reiterate the expected contributions from section 1.3 and evaluate them. In addition to arguing for principles such as learnability, encodability, decodability, and extensibility, we will also try to show that VIL satisfies Richards' criteria for a universal language to be accepted: the political, the psychological, and the linguistic aspects (as mentioned in section 1.4).

The expected contributions from section 1.3 were:
1. Creating a set of principles.
2. Designing the language and icons.
3. Implementing the language.
4. Demonstrating the viability of a visual inter language.
5. Identifying areas, which need additional work.

The set of principles is addressed in section 6.1. The design of the language and the icons, and the implementation of the system are evaluated by the two tests described in section 6.2 and 6.3. The viability of VIL and its future plans are described in the conclusion in the next chapter.

6.1 The Principles

As stated in section 1.3, most of the principles of the language are hard to test and have to be argued for. This was done in the previous chapters. In this section we argue that VIL meets the four primary capability principles learnability, encodability, decodability, and extensibility, and also Richards' three criteria, the political, the psychological, and the linguistic aspect.

Let us look at Richards' criteria first:

6.1.1 The Political Aspect

One of the major reasons that a natural language cannot serve as a universal language is the political problem. It would never be accepted by other countries. Nationalism and racial pride are insuperable obstacles. With an iconic language such as VIL, there is no political problem, since visual representations are universal. The actions in pictures take the place of references to any mother tongue. The following quote from Richards refers to using pictures to illustrate Basic English:

"By and large pictures are universal. People in the remotest corners of the earth and living in cultures least akin to ours learn our pictorial dialects with surprising ease and become thereby accessible to what we find to say with our pictures."[157]

6.1.2 The Psychological Aspect

The psychological aspect entails that if people have to learn a new language, then it should be easy to learn. The root criticism of any revived or artificial written language is that it is too hard to learn. VIL uses a restricted grammar and vocabulary to make the system not only universal but also easy to acquire. Also, the arguments given below for learnability, encodability, and decodability suggest that the system is easy to learn and use, so the user should have no fear to learn the system.

6.1.3 The Linguistic Aspect

Ogden argued that "languages are shaped by use rather than by design.[158] Long-established languages have been hammered and wrought, broken and remade, in countless ways which only the biggest and best of dictionaries can show in detail."

He stated that "no artificial language can acquire a tithe of such richness of interdependence, without centuries of wide and varied use. With a simplified form of a living language these discrepancies are reduced. Its meanings are held in place by the extent of the common use its words have been put to."[159]

It is true that only a non-artificial natural language can accomplish this. However, as stated, political aspects keep a natural language from becoming universally accepted. For an artificial language, the principle of extensibility could help extending and reshaping the language. Furthermore, we have to keep in mind the objectives of VIL. It is not intended as a full language that will replace natural language. It is not intended as a language in which to write poetry or scientific papers. It has been developed to allow people to communicate about a limited range of topics. In this respect, the linguistic aspect is not as relevant as Ogden suggests.

Colin Beardon stated it well when he said, "We prefer to see iconic systems as potentially enabling some level of communication between people with different viewpoints and have learned to be wary of any attempt to claim universality or objectivity for what we are creating".

---

[157] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[158] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

[159] **Richards, I. A.** *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

Let us now argue for how VIL meets the four primary capability principles:

- Learnability
- Encodability
- Decodability
- Extensibility

6.1.4 Learnability

Learnability means that the system should be easy to learn. We argue that with a visual language the learning process will be faster than with a written language. We believe it is easier for people to recognize visual representations than to come up with words for them. Richards and Ogden supported this view. Richards argued that the proportion of the meanings of a language that can be visually presented is an enormously important factor in determining the ease with which it can be learned and retained. Furthermore, in VIL, learnability is further supported by the following:

- VIL uses a drastically reduced grammar and vocabulary.
- Icons in VIL are easy to recognize (as test results later in this chapter will show).
- VIL uses an easy to use interface.
- VIL has numerous help functions that help explaining the meaning of an icon, like animation, category help, and tooltips.

As we shall see, the tooltips increase the learning curve a lot. They are used in the program and give tooltip help in the user's preferred language about not only the icons, but also all the visual elements present in an icon. We recognize that this is not in full agreement with our vision of a language free iconic communication system. However, the icons and messages produced by the system are text free. The tooltips are only used to make the system much easier to acquire (psychological condition). We expect and have observed during our tests that, once users gain some experience working with the system, they tend to use the tooltips less and less.

6.1.5 Encodability

Encodability means that it should be fairly easy to generate messages with the language and its delivery system. For a visual or iconic language messages are generated faster and with greater ease if the language is computer-based, since the user is not required to draw icons by hand. The icons can simply be chosen from the screen. The design of the grammar (we take the greatest common denominator of the things to be expressed, i.e. no inflection and no gender, etc.), the design of the hierarchies for our grammatical categories, and the design of the program itself are all designed to facilitate encoding.

Also, the "history" option in the program lets the user quickly choose one of his/her last used icons. Furthermore, the intelligent capabilities talked about in section 6.5 on possible enhancements could further speed up message generation.

6.1.6 Decodability

This means that the messages that are composed with the system should be easy to comprehend. Again, the simple grammar also helps in messages being easier to comprehend. VIL's self-explanatory icons and consistent visual combining features also greatly facilitate decoding of messages. Furthermore, the program has some help features, which make individual icons easy to decode:

- Animated explanation of verbs.
- Category help: showing where an icon stands in the hierarchy, i.e. the path from the root of the hierarchy to the icon in question is shown. Furthermore, siblings from the category the icon in question belongs to are shown. Showing these siblings greatly help recognizing an icon.
- Explanation of the different visual combining features of an icon by tooltips.

6.1.7 Extensibility

Extensibility means that the system should have the possibility to be extended by users. The design of the hierarchies makes it possible for the system administrator to easily add more icons, without it having to undergo a drastic redesign. Once these icons are added to the database, they can automatically be seen and used in the program for VIL.

We could let the user create new icons by combining icons. One way could be using icon compounding. Icons for adjective and noun could be compounded to form a new word that has a transferred or figurative meaning (e.g. *large* and *mouth* for *conceited*). Icons for verb and adverb could be compounded to form a new verb (e.g. *walk* and *fast* for *run*). These features, however, are not implemented in the current version of the system.

6.2 Testing the Icons

The design of the language and the grammar can best be evaluated when people are actually learning and using the system. This will be addressed in the next section. But before we can test the system, we must ensure that the icons are developed well and are understandable to people from different cultures. Did we follow the principles we stated for icon design in chapter 4? Are the icons truly universal? How many of the

icons are conventional (have to be learned), and are they consistent enough so one can predict the meaning of other conventional icons?

To answer all these questions, we needed to test the icons with people from various parts of the world. We need to set up a survey for testing the icons. This survey would serve the following purposes:

1. Determine how recognizable the icons are
2. Determine whether there are cross-cultural differences in recognition of icons

   Hypothesis: There is no difference in recognition of icons for people of western and non-western culture.

3. Find out which icons need to be redesigned

The creation of a survey consists of several steps.

Step 1: Determining the Target Audience

Ideally, the icons would be tested with speakers of all languages in the world. This is, of course, impossible to do. We need ways to test the icons people from as many cultures as we can. America is a multi-cultural country and even in our own neighborhood we find various clubs or organizations of people from different cultures (Italians, Greeks, Koreans, etc.). There are also lots of ESL classes (English as a Second Language) where we find a diversity of people.[160] The problem with this, however, is that these people may already be biased since they live in the US and may have become familiar with our culture or at least our way of representing things.

Step 2: Determining the Distribution Medium

To be sure that the tests are not culturally biased towards the U.S., we need to test with people that not only speak different languages, but also live in other countries. The best way to reach a wide variety of people from different cultures is an Internet icon test. By having the survey posted on the Internet, we can potentially reach any person in any country that has access to the Internet. We have created an electronic survey and put it on the Web. We recognize that this approach is not ideal in that people that use the Web are technologically literate people, and presumably relatively aware of other cultures. Furthermore, these people, being familiar with computers,

---

[160] **Bohne, S., Bunar, M., Gardella, S. and Hynes, P.** *"Cross-Cultural Icon Test"* An Interactive Qualifying Project Report, WPI, May, 1996.

_____

have an unfair advantage of already being familiar and exposed to lots of icons and user interfaces.

Step 3: Designing Questions

There are several different kinds of questions one could ask when testing the icons:

1. *open ended* (e.g. what best describes the meaning of this icon?),
2. *multiple choice* (e.g. does this icon represent: a... b... etc.),
3. *image-for-meaning* (i.e. give some word and ask the subject to pick the appropriate icon),
4. *image-meaning* (i.e. let the subject match icons and meanings from two lists, one of icons and one of meanings/words),
5. *icon set-category* (i.e. given a specific set of icons, ask the subject what category do these icons belong to?).

In our survey, questions of type 1, 4, and 5 occur. One major disadvantage with our icon test is that the instructions as well as the questions assume that the test subjects are familiar with the English language. Ideally, it would have been better to translate the survey into different languages to ensure more people understand and participate in the survey.

When considering concrete icons, besides the types of questions mentioned above, there is another type of question (about prototypicality) we could have asked the subjects:

6. *prototypical image*: Give the subject several versions of icons that are generally recognized everywhere, like a *tree*, and ask him or her "which icon do you think best represents a tree?" Also, give the subject several versions of icons that are represented differently across cultures, like a *mailbox*, and ask the same question.

This last type of question could have been usefull for determining the best-known version of an icon. It is often the case that for a concrete object there is no universal representation. In such cases, it is useful to use the best known version, i.e. "which of the icons is recognized by most people?"

Step 4: Advertising the Survey

Once our web survey was set up and put on the Web, we needed to let as many people know about it as we could. We asked students at the university and friends to test the icons. But more importantly, we advertised the survey by posting to various

newsgroups that deal with various cultures. The following is a list of newsgroups, to which we posted our advertisement:

soc.culture.arabic	soc.culture.argentina	soc.culture.australian
soc.culture.belgium	soc.culture.british	soc.culture.china
soc.culture.cuba	soc.culture.egyptian	soc.culture.filipino
soc.culture.greek	soc.culture.indian	soc.culture.irish
soc.culture.israel	soc.culture.iranian	soc.culture.lebanon
soc.culture.malaysia	soc.culture.netherlands	soc.culture.polish
soc.culture.new-zealand	soc.culture.pakistan	soc.culture.puerto-rico
soc.culture.romanian	soc.culture.russian	soc.culture.singapore
soc.culture.south-africa	soc.culture.spain	soc.culture.turkish
soc.culture.thai	soc.culture.usa	soc.culture.vietnamese
soc.culture.zimbabwe	soc.culture.yugoslavia	soc.culture.venezuela
soc.culture.ukranian	soc.culture.swiss	soc.culture.native
soc.culture.kenya	soc.culture.algeria	soc.culture.costa-rica
soc.culture.cambodia	soc.culture.guinea-conakry	soc.culture.french
soc.net-people	misc.education	alt.cooking-chat
alt.cooking-chien	alt.mexico	

6.2.1 The VIL Website and the Web-based Icon Test

In order to be able to put the survey on the Internet, we created a website. This website was designed by Rishi Thakoerdat, a graphic design student, who also designed most of the icons. On this website one can:

- Learn and read about VIL.
- Download VIL papers, screensavers, and the application.
- Read about other work on iconic and visual communication.
- Read about artificial languages.
- Read about language families and natural languages.
- Read about general linguistics topics and dissertations in linguistics.
- Find multi-lingual and normal dictionaries.
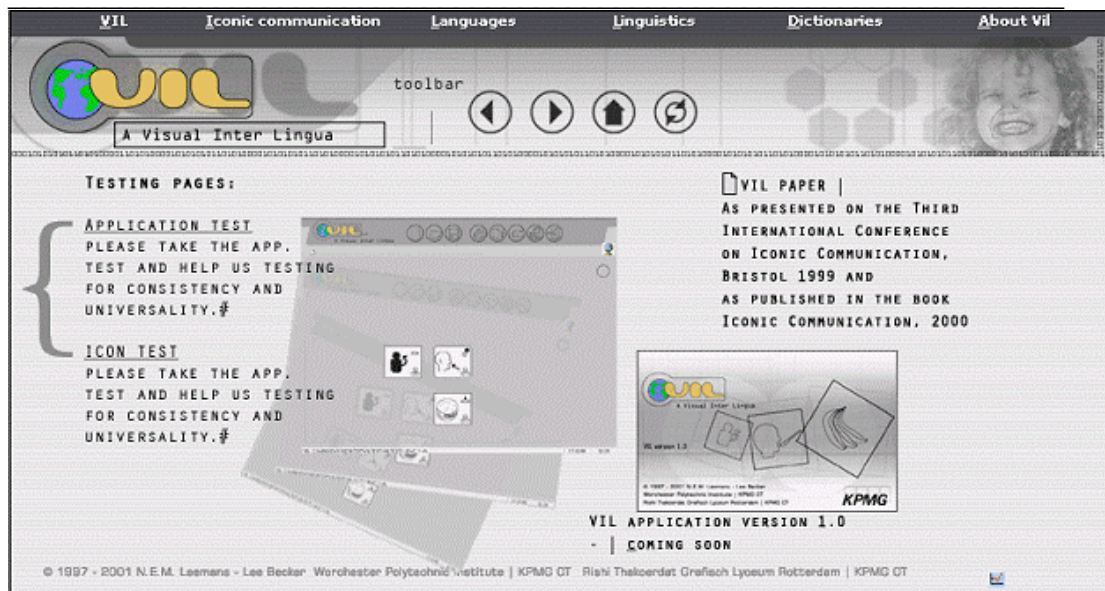
Figure 80. The main page of the VIL website.

The website, of course, also has links to the VIL icon test and the VIL language/application test, the latter of which is the discussion of section 1.3. The icon test is divided into three parts:

- Testing verbs
- Testing nouns
- Testing adjectives
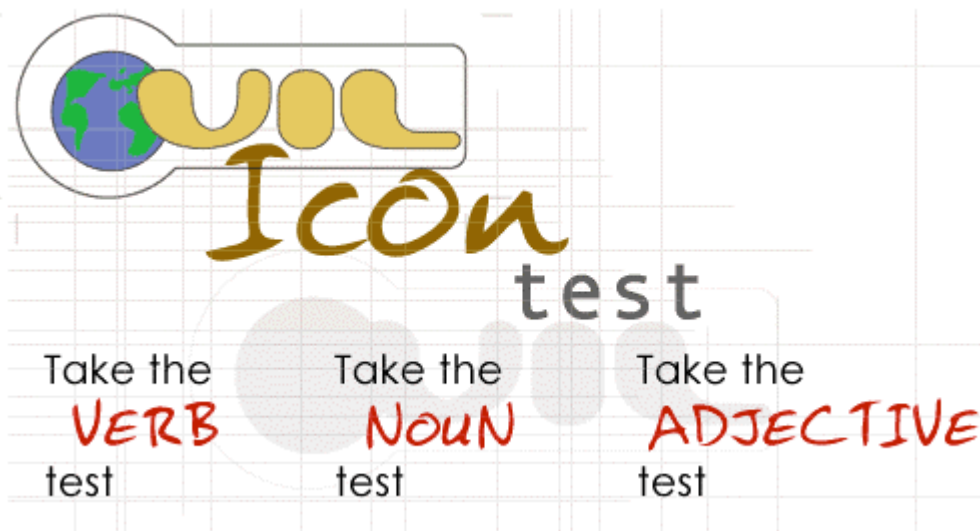
Each test is further subdivided into two parts. In part A the user is asked to identify icons that represent categories of verbs, e.g. "cooking verbs". In part B the user is asked to identify icons that represent single verbs, e.g. "to bake". Within each part these are image-meaning questions and open ended questions. For categories, the image-meaning questions ask the subject to match icons and meanings from two lists and the open ended questions ask the subject to identify the categories the icons represent. For terminals, i.e. single icons, the open-ended questions ask the subject to identify the single icons representing concrete or abstract meaning.

We designed 11 test pages for verbs, 22 test pages for nouns, and 8 test pages for adjectives. The following figure shows the verb test page:
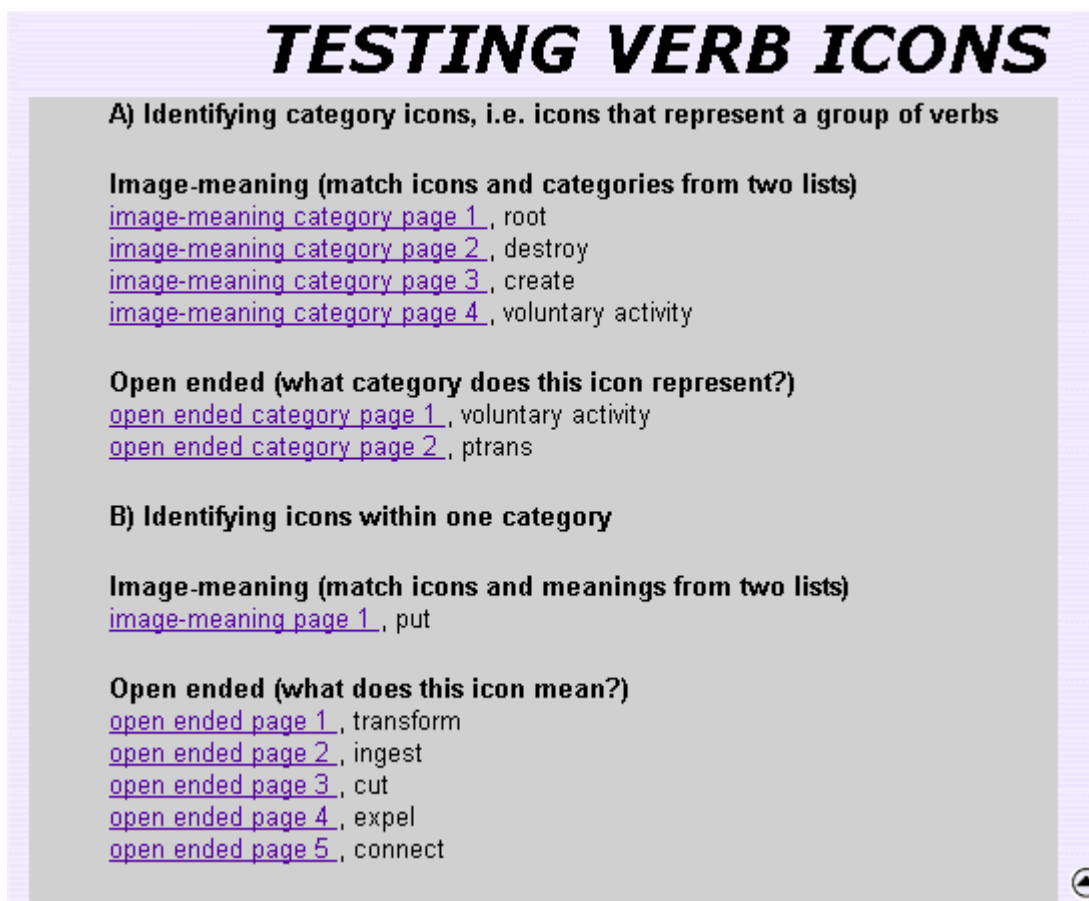
The following four images show screenshots of image-meaning category, open-ended category, image-meaning terminal, and open-ended terminal, question pages.

The following figure shows the image-meaning category questions page for the create verbs category:

IMAGE-MEANING CATEGORY QUESTIONS PAGE 3

For each of the verbs below, please fill in the number that you think that corresponds to the meaning in the list of meanings below. All verbs represented on this page are verbs that represent categories of verbs. Please also write down what you think these groups of verbs are themselves categories of, i.e. its parent category, for example the categories all represent "verbs of creation".

In order to get more accurate results,
it would be greatly appreciated if you'd fill in your country
of birth and country of residence, if different.

What is your country of birth?:

What is your country of residence (if diffferent?:

Match the meaning of the following icons?

Number	Meaning
1	Reveal verbs.
2	Calve verbs.
3	Connect verbs.
4	Create verbs.
5	Record verbs.
6	Transform verbs.
7	Present verbs.

These categories of verbs all belong to the (parent) category:

The next figure shows the open-ended category questions page for the voluntary activity verbs category:

OPEN ENDED CATEGORY QUESTIONS PAGE 1

For each of the verbs below, representing groups of verbs (categories), please write down what you think they mean. All verbs represented on this page are verbs that represent categories of verbs (so one must think very general), e.g. for the first icon, do not think "push verbs", but "verbs of exerting a force". The verbs here are verbs that represent, i.e. are children of the category, "verbs representing a voluntary activity".

In order to get more accurate results,
it would be greatly appreciated if you'd fill in your country
of birth and country of residence, if different.

What is your country of birth?:

What is your country of residence (if diffferent?:

What is the meaning of the following icons?

The following figure shows the image-meaning terminal questions page for the put/pour verbs category:

Match the meaning of the following icons?

Number	Meaning
1	Spray.
2	Fill.
3	Pour.
4	Put.
5	Empty.

These verbs all belong to the category:

The next figure shows the open-ended terminal questions page for the cut verbs category:

We expect that open-ended questions are more difficult, since the choices are not specified and one cannot use elimination. On every test page, the subject is asked to fill in his/her country of birth and country of residence. In case of incorrect answers, this information helps us identifying icons that are not specific to a culture. Furthermore, at the end of each test page, the subject is asked to identify the category that the icons on that page belong to.

6.2.1.1 What if the icons are not 100% universal?

For a complex system like VIL, designing icons that are totally language independent and culturally neutral may prove impossible.

Fortunately users do not require perfect icons, only recognizable ones. One can use images not common to every culture provided that users can still recognize them.

Most users will still recognize the following icons, though not common to every culture: [161]

An Arab will still recognize this icon as being an icon of a book, even though books in Arabic, which reads right-to-left, would have the spine on the right.

A Japanese will still recognize this icon as being a symbol of greeting among people, even though in the orient a bow is more common as a greeting than a handshake.

A Chinese user will still recognize this icon as an icon for silverware (concrete) or food or restaurant (abstract), even though chopsticks and a bowl may be more common in China.

If a large proportion of people recognize what a particular icon stands for then we can say that it is a successful icon.

6.2.1.2 What if there are many errors on particular icons?

Once we have determined, which icons were identified incorrectly, we need determine when to redesign an icon. Should we redesign every incorrect icon? Should we redesign icons that accounted for more than 10% of the errors? We need to determine some cutoff point beyond which icons have to be redesigned. To determine this cutoff point we need to compare the number of incorrect answers for an icon against the number of correct answers for that icon. For an individual icon it is not good enough to say that if the icon accounts for more than 6% of the errors then it needs to be redesigned. Let us assume this 6% amounts to 80 incorrect answers. What if it was identified correctly 7920 times? Then it was actually interpreted incorrectly 1%. Redesigning icons is discussed more in section 6.2.1.3.3.

6.2.1.3 The VIL Icon Test Results

The icon test was conducted for about two months. This yielded the following feedback:

- *Number of tests*: 687 pages were tested

---

[161] **Horton, W.** *"The Icon Book: Visual Symbols for Computer Systems and Documentation,"* John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

- *Number of answers*: this yielded 4960 answers for individual icons (image-meaning and open ended) and 319 answers for categories sets of icons belong to (icon set-category question)
- *Nationality*: people from at least 29 countries took the test (89 people did not fill in their nationality). The following table shows the nationalities of the subjects along with the number of icons tested by them:

Nationality	# icons tested
Australia	6
Azerbaijan	83
Belgium	434
Brazil	7
Canada	227
China	7
Choctaw Nation	19
Cyprus	18
Egypt	7
England	115
France	93
Greece	149
India	48
Iran	17
Ireland	140
Malaysia	7
Mexico	7
Netherlands	318
oz	34
Pakistan	11
Poland	46
Romania	4
Scotland	7
Singapore	4
Surinam	6
Switzerland	4
Turkey	353
United Kingdom	21
USA	2126
Yugoslavia	4

6.2.1.3.1 Interpreting the test results

In order to see whether there are any cultural differences in the answers, we divided the answers up into "Western" people and the rest of the world based on the nationalities of the subjects, which responded to the survey, we somewhat arbitrarily divided them into Western vs. non-Western. Western people are people from USA, Canada, Australia, and western Europe. From the 4960 received answers, 3644 answers were from Western people, and 1316 from the rest of the world.

For all 4960 questions, 3914 answers were correct and 1046 answers incorrect. This is around 79% correct answers for all questions.

The following table splits the results into answers to image-meaning and answers to open ended questions. Within these we make a distinction between western and non-western cultures, and also between questions in which the subject had to identify icons representing categories and questions in which the subject had to identify icons representing terminals.

Type of question	Western culture	Category vs. terminals	nr. of answers	nr. of correct answers	nr. of incorrect answers	Percentage correct
Image-meaning	Yes	Category	1384	1065	319	77%
Image-meaning	Yes	Terminal	225	217	8	96%
Image-meaning	No	Category	527	430	97	82%
Image-meaning	No	Terminal	75	65	10	87%
Open ended	Yes	Category	416	270	146	65%
Open ended	Yes	Terminal	1619	1324	295	82%
Open ended	No	Category	122	73	49	60%
Open ended	No	Terminal	592	470	122	80%

If we filter out the results for western versus non-western cultures from the above table, we see the following:

Type of question	Category vs. terminals	Percentage correct Western	Percentage correct Non-Western
Image-meaning	Category	77%	82%
Image-meaning	Terminal	96%	87%
Open ended	Category	65%	60%
Open ended	Terminal	82%	80%

From the above tables the following conclusions can be drawn:
- If we focus on western versus non-western culture, we see that the results are close together for all types of questions. This means that the icons for terminals

are generally recognized equally well across cultures (no statistically significant differences). These findings are confirmed by statistically results from t-tests with α = 0.05 (95%):

Type of question	Category vs. terminals	t-test result
Image-meaning	Category	t(16) = 0.789793
Image-meaning	Terminal	t(8) = 1.502955
Open ended	Category	t(8) = 0.863064
Open ended	Terminal	t(42) = -0,30565

- As expected, the open-ended questions are harder than the image-meaning questions. We accounted this to the fact that for open-ended questions one cannot apply the process of elimination.
- Category questions are more difficult than terminal questions. This is due to the fact that it is easier to recognize a single object than to recognize the icon for a group of objects. It is also safe to assume that categories higher up the tree are harder to recognize than categories that are lower in the tree. This is because they have to represent a broader range of objects. Consider having to recognize the icon for the category "fruit" versus the icon for the category "physical world", which is a top-level category.

If we split up the results into questions about verbs, nouns, and adjectives, we see the following:

Type of question	# answers	# correct answers	# incorrect answers	Percentage correct
Verb categories	1181	721	470	61%
Verb terminals	722	542	180	75%
Noun categories	1144	1035	109	90%
Noun terminals	1585	1368	217	86%
Adjective categories	114	82	32	72%
Adjective terminals	204	166	38	81%

If we split this up into western versus non-western we get the following results:

Western culture				
Type of question	# answers	correct answers	incorrect answers	Percentage correct
Verb categories	940	565	375	60%
Verb terminals	509	386	123	75%
Noun categories	772	703	69	91%
Noun terminals	1176	1027	149	87%
Adjective categories	88	67	21	76%

Adjective terminals	159	128	31	81%

Non-Western culture				
Type of question	**# answers**	**correct answers**	**incorrect answers**	**Percentage correct**
Verb categories	251	156	95	62%
Verb terminals	213	156	57	73%
Noun categories	372	332	40	89%
Noun terminals	409	341	68	83%
Adjective categories	26	15	11	58%
Adjective terminals	45	38	7	84%

If, from the above table, we compare the percentage correct of western versus non-western people, we see the following:

Type of question	**Category vs. terminals**	**Percentage correct Western**	**Percentage correct Non-Western**
Verb	Category	60%	62%
Verb	Terminal	75%	73%
Noun	Category	91%	89%
Noun	Terminal	87%	83%
Adjective	Category	76%	58%
Adjective	Terminal	81%	84%

We see the following:
- Verb and adjective terminal icons are recognized better than verb and adjective category icons. This is as expected since it is harder to recognize an image representing a category than an image representing a single verb or adjective.
- Noun category icons are recognized better than verb category icons. An explanation for this could be that for nouns the category icons consist of several images of children of the category and these are generally recognizable concrete objects, e.g. an apple. For verbs and adjectives the categories are single icons that represent the icons in its category. These are generally harder to recognize.
- Nouns are recognized much better than verbs and adjectives. This is because nouns contain the most concrete objects.
- Adjectives are recognized well because they generally show two antonyms in the icon, which allows them to be recognized better than each of the antonyms in isolation.
- There is very little difference between the test results for western and non-western people. The only noticeable difference is for adjective categories (18%). There is no apparent logical explanation for this and this may be due to the fact that there were not many adjective category questions (only two pages). Results from t-tests show that there are no statistically significant differences (with $\alpha = 0.05$):

Type of question	Category vs. terminals	t-test result
Verb	Category	t(8) = 0.315254
Verb	Terminal	t(10) = -0.10464
Noun	Category	t(12) = -0.610185
Noun	Terminal	t(28) = -1.504452
Adjective	Category	t(2) = 1.006349
Adjective	Terminal	t(10) = -0.60311

One thing that we could not measure is the time it took the subjects to fill in a form. From a few subjects that we were able to watch as they took the tests, we concluded the following:

- It took longer to recognize icons representing abstract concepts like categories than to recognize icons representing concrete objects, like an apple or a glass
- For abstract concepts, image-meaning tests were completed faster than open ended questions
- For concrete objects, the time to complete a test page depended on the number of icons in the page. If there were a lot of icons, then open-ended questions were completed faster, if there were a few questions, then image-meaning questions were completed faster. This is largely due to the time lost scrolling back to the top of the page to see the list of meanings to match.
- Concrete objects are recognized much faster than abstract objects
- Nouns are recognized faster than verbs and adjectives
- Adjectives are recognized faster than verbs
- Individual (terminal) icons are recognized faster than category icons

These results are in agreement with the test results from the above table.

6.2.1.3.2 Errors

In this section we look closer at the errors made by the test subjects. The following table shows what percentage of the icons accounted for what percentage of the errors. There did not appear to be any statistically significant differences between Western versus non-Western people.

Type of question	# icons incorrect answers	total # icons in test	# incorrect answers for these icons	total # incorrect answers	% icons	% errors
Verb categories	6	42	282	470	14%	60%
Verb terminals	6	41	65	180	15%	36%
Noun categories	5	45	71	109	11%	65%
Noun terminals	8	140	63	217	6%	30%
Adjective categories	3	13	15	32	23%	47%
Adjective terminals	3	35	13	38	9%	34%

When only a few of the answers account for most of the errors, the test is even more successful, because this indicates that most other icons are recognized pretty well. We only need to adjust the most erroneous icons, for example for verb categories, only 6 icons accounted for 282 errors, which is 60% percent of the errors.

In the rest of this section we identify the icons that were recognized incorrectly the most for both category and terminal icons split up into verbs, nouns and adjectives. The following table shows verb categories that were incorrectly identified 10 or more times:

Verb Category	Icon representing	Nr Incorrect Answers	Most common mistakes
Physical transfer	Put	14	Pour
Physical transfer	Lodge	12	Camp
Physical transfer	Conceal	12	?, knock
Physical transfer	Sending	12	Mail
Root	Voluntary activity	17	Physical transfer
Root	Existence	12	Voluntary activity, stand
Root	Physical transfer and location	10	Voluntary activity, go home
Voluntary activity	Move without changing location	68	Sit
Voluntary activity	Hold & keep	54	Carry
Voluntary activity	Expel	46	Speak
Voluntary activity	Exerting a force	45	Send, throw
Voluntary activity	Sports	36	Play tennis, serve
Voluntary activity	Touch	33	?, slap, spank

The following table shows verb terminals that were incorrectly identified 7 or more times:

Verb Category	Icon representing	Nr Incorrect Answers	Most common mistakes
Connect	Puree	12	Brush, sweep
Cut	Grind	12	Stir, mix, African food preparation
Cut	Chop	8	Cut
Expel	Breathe	8	Reject, vomit, spit out
Expel	Burp	7	Blow bubbles
Expel	Spit	7	Blow, sneeze
Transform	Barbeque	17	Grill, cook, roast
Transform	Cook	7	Heat, simmer
Transform	Rise	7	Overflow, overdone
Transform	Grill	7	Cook, roast, burn

The following table shows noun categories that were incorrectly identified 6 or more times:

Noun Category	Icon representing	Nr Incorrect Answers	Most common mistakes
Food	Meals	8	Dishes
Food	Dishes	8	Meals
Physical world	Living world	21	Environment, ecosystem, world
Physical world	Geology	15	Mining, crystals
Animals	Mollusks	11	Sea creatures
Animals	Mammals	16	Farm animals
Animals	Birds	8	Chicken, poultry
Animals	Reptiles	6	Wild animals, lizards

The following table shows noun terminals that were incorrectly identified 6 or more times:

Noun Category	Icon representin	Nr Incorrect Answers	Most common mistakes
Fish	Tuna	8	Fish
Birds	Ostrich	6	Flamingo
Cooking utensils	Plate	6	Dish, saucer
Cooking utensils	Peeler	6	Cup
Silverware	Soup cup	7	Cup, baby drinking cup
Silverware	Rice scoop	6	Measuring spoon
Vegetables	Beet	9	Radish
Fruit	Coconut	7	?
Fruit	Peach	7	Plum, avocado

Mixed noun page	Salad dish	7	Pizza, entrée
Pots & pans	Bump pan	9	Pot
Pots & pans	Pie pan	9	Bowl
Pots & pans	Grill	6	Waffle iron

The following table shows adjective categories that were incorrectly identified 4 or more times:

Adjective Category	Icon representing	Nr Incorrect Answers	Most common mistakes
Measurement	Distance	6	Comparison
Root	Aural	5	Ear, listen
Root	Taste	4	Tongue

The following table shows adjective terminals that were incorrectly identified 4 or more times:

Adjective Category	Icon representing	Nr Incorrect Answers	Most common mistakes
Size & dimension	Min-max	5	Length, larger
Taste	Bitter	5	Tangy, sour

6.2.1.3.3 Redesigning Icons

Now that we know the incorrectly identified icons, we need to ask ourselves, when should we redesign an icon? Should we redesign every incorrect icon? Should we redesign icons that accounted for more than 10% of the errors? To determine a cutoff point for when an icon needs to be redesigned, we actually need to compare the number of incorrect answers for an icon against the number of correct answers for that icon. For an individual icon it is not good enough to say that if the icon accounts for more than 6% of the errors then it needs to be redesigned. Let us assume this 6% amounts to 80 incorrect answers. What if it was identified correctly 7920 times? Then it was actually interpreted incorrectly 1% (80/(80+7920)). If we look at the verb terminals, for which there were incorrect answers, in this manner we see the following:

Icon	# Correct	# Incorrect	Total answers	Percentage incorrect
Barbeque	2	17	19	89%
Puree	4	12	16	75%
Grind	5	12	17	71%
Chop	9	8	17	47%
Breathe	19	16	35	46%
Rise	9	7	16	44%

Icon	# Correct	# Incorrect	Total answers	Percentage incorrect
Burp	11	7	18	39%
Cook	11	7	18	39%
Grill	11	7	18	39%
Spit	11	7	18	39%
Dice	10	6	16	38%
Boil	12	6	18	33%
Steam	13	6	19	32%
Mix	12	5	17	29%
Suck	12	5	17	29%
Slice	12	5	17	29%
Cut	12	5	17	29%
Eat	13	4	17	24%
Freeze	15	4	19	21%
Bake	15	4	19	21%
Melt	15	4	19	21%
Marinate	13	3	16	19%
Empty	14	3	17	18%
Fry	15	3	18	17%
Bend	16	3	19	16%
Fill	15	2	17	12%
Lick	15	2	17	12%
Whip	16	2	18	11%
Peel	15	1	16	6%
Put	16	1	17	6%

If we look at the noun terminals, for which there were incorrect answers, in this manner we see the following:

Icon	# Correct	# Incorrect	Total answers	Percentage incorrect
Beet	3	9	12	75%
Coconut	5	7	12	58%
Ostrich	5	6	11	55%
Peeler	6	6	12	50%
Flatfish	6	5	11	45%
Corn	8	4	12	33%
Broccoli	8	4	12	33%
Pear	9	4	13	31%
Plate	17	7	24	29%
Chocolate	8	3	11	27%
Chili	8	3	11	27%

Icon	# Correct	# Incorrect	Total answers	Percentage incorrect
Rooster	9	2	11	18%
Snake	9	2	11	18%
Swordfish	9	2	11	18%
Shell	9	2	11	18%
Grasshopper	9	2	11	18%

The tables only show icons for which there were more than 10 answers given. Although, to be able to make a fair judgment, we should consider icons for which there were more than 15 answers. We believe that we definitely need to redesign icons that were interpreted incorrect more often than correct. For icons that were interpreted correct more often than incorrect, we believe the cutoff percentage is hard to determine. Consider the verb "to bend":

Icon	# Correct	# Incorrect	Total answers	Percentage incorrect
Bend	16	3	19	16%

An incorrect percentage of 16% sounds high. However, 16 correct answers and only 3 incorrect sounds pretty good. Considering this information, we believe a cutoff percentage of 20% is reasonable.

However, some icons will remain conventional and have to be learned. However, if these icons are designed well, then they are probably remembered well. Even though the significant task of recognition is usually assumed to occur when an icon is first encountered, there is a strong case to be made that it is more important that the meaning of the icon is remembered. An initial struggle to determine the intended meaning of an icon (or some instruction or tutorial) may be a small price to pay if it is then remembered accurately.

6.2.2 What Was Learned from the Icon Test: Summary

Let us look at the purposes of the icon test again and for each summarize the conclusions.

6.2.2.1 Determine how recognizable the icons are

From the results of the test we have seen that the icons were recognized very well, around 79% correct answers for all questions.

Summarizing, the following observations were made:

- Category icons are generally harder to recognize than terminal icons.

The following summarizes the results for the image-meaning and open-ended questions:

Type of question	Category vs. terminals	Percentage correct
Image-meaning	Category	80%
Image-meaning	Terminal	92%
Open ended	Category	63%
Open ended	Terminal	81%

- Nouns are recognized much better than verbs and adjectives.
- Adjectives are recognized better than verbs.
- Noun category icons are recognized better than noun terminal icons.

The following summarizes the results for the questions about verbs, nouns, and adjectives:

Type of question	Percentage correct
Verb categories	61%
Verb terminals	75%
Noun categories	90%
Noun terminals	86%
Adjective categories	72%
Adjective terminals	81%

6.2.2.2 Determine cross-cultural differences in recognition of icons

The following hypotheses was made:

- **Hypothesis**: There is no difference in recognition of icons for people of western and non-western culture.

The hypothesis was confirmed. We saw that the results were close together for all types of questions. This meant that the icons for terminals are generally recognized equally well across cultures (no statistically significant differences). These findings were confirmed by statistically results from t-tests with $\alpha = 0.05$ (95%):

Type of question	Category vs. terminals	t-test result
Image-meaning	Category	t(16) = 0.789793
Image-meaning	Terminal	t(8) = 1.502955
Open ended	Category	t(8) = 0.863064
Open ended	Terminal	t(42) = -0,30565

Type of question	Category vs. terminals	t-test result
Verb	Category	t(8) = 0.315254
Verb	Terminal	t(10) = -0.10464
Noun	Category	t(12) = -0.610185
Noun	Terminal	t(28) = -1.504452
Adjective	Category	t(2) = 1.006349
Adjective	Terminal	t(10) = -0.60311

6.2.2.3 Find out which icons need to be redesigned

We determined a cutoff point for when an icon needs to be redesigned. We actually compared the number of incorrect answers for an icon against the number of correct answers for that icon. By carefully studying all the errors we believed that a cutoff percentage of 20% was reasonable.

6.3 Testing the design of the language and the implemented VIL application

Besides testing if the icons are universal, we also need to test the language and the application to see if VIL is easy to learn and use (learnability, encodability, and decodability). Shneiderman mentions five measurable quality criteria:[162]

1. Time to learn
2. Speed of performance
3. Rate of errors by users
4. Retention over time
5. Subjective satisfaction

Time to learn the system and retention over time are ways to test our principle of learnability. Speed of performance and rate of errors by users are ways of testing our principles of encodability and decodability. Subjective satisfaction is not easily tested, and the test subjects have to be asked.

---

[162] **Shneiderman, B.** *"Designing the User Interface: Strategies for Effective Human-Computer Interaction,"* Addison-Wesley, 1992.

We have conducted five separate tests to test the system:

1. A web test, in which subjects had to decode 10 sentences without help.
2. A web test, in which subjects had to decode 10 sentences with tooltip help.
3. A live test, in which subjects had to encode 10 sentences without training.
4. A live test, in which subjects had to encode 10 sentences with training.
5. A repeat of the live test a week later without instructions.

6.3.1 The Web Tests for Decodability

Besides testing decodability, the web-based tests gave us feedback as to what to change in the design of the language or the implemented VIL application. In both web tests there were some minimal instructions telling the subjects what the various visual elements in a VIL icon were. Since these visual elements are highly conventional (most of them representing grammatical entities), it was only fair to tell the subjects at least this and not have them guess at it. If we had not done this, then the test subjects taking the web test would have an unfair disadvantage of not knowing if they were on the right track. Consequently, learnability and even decodability would have been hard to test. Compare this to the live test were test subjects get feedback by the system after composing a message telling them if the sentence is correct and if not, which part is not correct.

The web tests could only test decodability. As with testing our icons over the web, we could not test speed of performance or time to learn. The main difference between the two web tests for the application was that one test had tooltip help and the other one did not. This way we were able to test what difference in decoding tooltip help makes.

The web tests served the following purposes:

1. To determine whether VIL is decodable enough
2. To determine whether tooltips increase performance

   Hypothesis: performance for the test with tooltip help is better than performance for the test without tooltip help

3. To learn what needs to be changed to the system

6.3.1.1 The Test Sentences

As with testing icons, the test subjects should ideally be people with different linguistic backgrounds. The two web tests mentioned are again a great means to reach as many multi-cultural people as possible. For the decodability test we created 10 sample sentences. Six of these sentences were simple declarative sentences:

1. I eat a hamburger.
2. I drink milk 3 times a day.
3. I put a pork chop in a frying pan.
4. I put (past tense) butter on my bread with a blunt knife.
5. I baked you a cake from red apples and yellow bananas.
6. In France, I ate 10 long French breads last summer.

Two sentences involved an imperative sentence:

7. Stir all eggs for a long time!
8. Mix 1 liter milk and 3 eggs in a bowl!

The last two sentences were interrogative sentences:

9. What are you eating?
10. When and for how long did you boil the water?

The following is a screenshot from the web application test showing the test page for "I eat a hamburger":
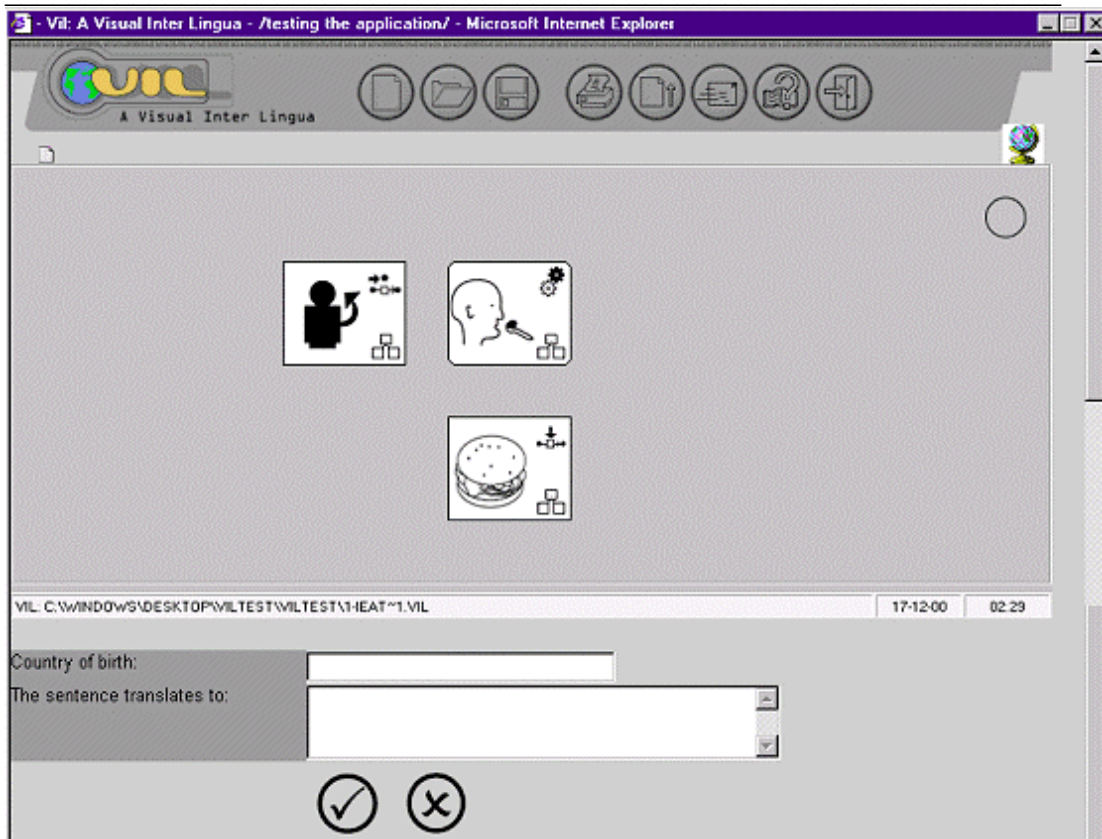
Subjects were asked to fill in their nationality and to give a translation of what they thought the sentence meant.

6.3.1.2 The Web Tests Results

The decodability test was conducted for about one month. This yielded the following feedback:

- *Number of sentences tested*: 98 sentences were tested by subjects
- *Nationality*: people from at least 9 countries took the test (10 people did not fill in their nationality). The following table shows the nationalities of the subjects along with the number of sentences tested by them:

Nationality	# sentences tested
Australia	4
England	8
Lebanon	3
Netherlands	27

Nationality	# sentences tested
Romania	2
Sweden	2
Turkey	24
United Kingdom	12
USA	6
Unknown	10

6.3.1.2.1 Interpreting the test results

From the total of 98 received translations for both tests, 25 answers were exactly correct. Exactly correct means that the translation corresponded exactly to the sentences as presented above. From these 98 answers, 75 were answers to the first test without help and 23 were answers to the test with tooltip help.

Type of test	# answers	# exactly correct answers	Percentage exactly correct
Test 1	75	18	24%
Test 2 with help	23	7	30%

As expected, the test with tooltip help scores somewhat higher than the one without. At a first glance, these numbers look pretty low. This does not mean that the rest of the answers were completely wrong. Some answers came close to the intended translation. For the remaining answers, we have to measure their partial correctness. Before we measure partial correctness, let us first look at some issues involving the web-based decoding test.

6.3.1.2.2 Issues

There was no way of knowing whether or not people had read the instructions before they started the test. From conversation with some subjects who took the web test (these were people I happened to know around the world), I learned that some had either not read the instructions or had not put much thought into decoding the sentences. The people who told me they had read the instructions, scored higher than the rest. If we could somehow have forced people to read the instructions before they took the test, we expect the result would have been better.

It is also hard to test the system in this way, since people had no idea if what they were doing was correct, since they did not receive any feedback after a sentence. For

instance, there was one person who translated some sentences right, but who kept putting in "the man" for "I". Had someone told him after the first mistake that "I" was meant, he/she would not have made this mistake in subsequent sentences.

Furthermore, a lot of people did not look further than the surface image of the sentence. These people either did not read the instructions or did not read the instructions about the asterisk, indicating noun-modifying properties. They did not interrogate any of the icons to find out about modifying properties. Again, had someone told them after the first mistake, they would not have made it a second time.

Finally, a lot of people did not understand the noun conjunction construction, like "red apples and yellow bananas". It turns out that people encoding the sentences (the last three live tests) have no problem understanding and using this construction, even without instructions.

These subjects certainly would have had a higher percentage of correct answers on latter sentences had they been given feedback after each sentence.

6.3.1.2.3 Partial correctness

We measured partial correctness by looking at a few different things:

1. Correct on top-level screen/frame, but did not include information that could be obtained from other screens through clicking on asterisked icon.

    The surface sentence was interpreted correctly, but the icons were not queried to see if they had modifying properties.
    - Example: "I baked you an apple pie" (instead of from "I baked you a pie from red apples and yellow bananas").

2. Correct, but didn't consider the noun conjunction.

    These subjects did obtain information by querying the asterisked icons. However, they did forget to interpret the noun conjunctions. Sentences were interpreted correctly, but the noun conjunction was missed.
    - Example: "I baked you an pie from red apples" (instead of "from red apples and yellow bananas").

3. Correct except some icons were recognized incorrectly.

    - Example: "I put the ham in the pan" (instead of "I put a porkchop in a frying pan").

4. Correct except sentence not grammatically correct, but is the meaning close to the one intended.

- Example: "I'm hungry and I'm going to eat a burger" or "I'd like to eat a hamburger" (instead of "I eat a hamburger").

5. Correct, but some grammatical element was ignored in the translation.

- Example: "I drink milk" (instead of "I drink milk 3 times a day").

The following table shows the number of answers that were partially correct for the web test without help. The table is divided according to the above mentioned criteria for partial correctness. At the end the remaining incorrect answers are shown:

Nr.	Partially correctness	# of answers
1	Correct but didn't consider underlying information	6
2	Correct, but didn't consider the noun conjunction	3
3	Correct except some icons were recognized incorrectly	6
4	Correct except sentence not grammatically correct, but is the meaning close to the one intended	8
5	Correct, but some grammatical element is not present in the translation	13
Remaining incorrect out of 75		21

The following table shows the number of answers that were partially correct for the web test with tooltip help:

Nr.	Partially correctness	# of answers
1	Correct but didn't consider underlying information	3
2	Correct, but didn't consider the noun conjunction	0
3	Correct except some icons were recognized incorrectly	2
4	Correct except sentence not grammatically correct, but is the meaning close to the one intended	1
5	Correct, but some grammatical element is not present in the translation	3
Remaining incorrect out of 23		3

As one can see, when partially correctness is measured, the number of incorrect translations is greatly reduced to 28 in total.

The following table shows the complete results:

Type of test	# answers	# exactly correct answers	# partially correct	# incorrect	% correct + partially correct
Test 1	75	18	36	21	72%
Test 2 with help	23	7	9	7	70%

The following table shows the types of partial correctness and suggests how these errors can be eliminated.

Partially correctness	Eliminate by
Correct but didn't consider underlying information	Covering in the instructions (explaining how the interface works)
Correct, but didn't consider the noun conjunction	Covering in the instructions (explaining how the interface works)
Correct except some icons were recognized incorrectly	Create more recognizable icons as pointed out by the icon test
Correct except sentence not grammatically correct, but is the meaning close to the one intended	Covering in the instructions (explaining how the grammar works)
Correct, but some grammatical element is not present in the translation	Covering in the instructions (explaining what a sentence in VIL consists of)

The table shows that these aspects need to be covered in the instructions and that it is important to read the instructions first, however small they may be.

For the incorrect answers that could not be classified under partial correctness, i.e. answers that were judged as completely wrong, the most common mistakes were people translating the sentences too generally. The following are some incorrect answers:

Intended meaning	Incorrect translation
Stir all eggs for a long time.	Cooking eggs
Mix three eggs and 1 liter of milk in a bowl.	Omelet Mix
I eat a hamburger.	Go and eat hamburger
I drink milk three times a day.	I am thirsty
I put a porkchop in a frying pan.	I'm preparing meat
In France, I ate ten long French breads last summer.	Last night I ate bread.
What are you eating?	What are you feeding me?

6.3.1.3 What Was Learned from the Decodability Tests: Summary

Let us look at the purposes of the decodability test again.

6.3.1.3.1 Determine whether VIL is decodable enough

It is hard to state a hypothesis for this. If we consider partial correctness we see a correctness percentage of around 72%. This is good performance considering the fact that people had never seen the program before. If we adjust the instructions and if people really read the instructions, we believe the results will be much better. The only hypothesis we can then make is that performance of subjects who read the instructions is better than performance of subjects that do not read the instructions. This hypothesis can be confirmed when we account most of the partial correctness errors to the fact that people did not read the instructions. In the next section on the live test, we can really measure the difference between test groups with and without some teaching. The difficult issue with this decodability test is that it was an Internet based test. We have no idea whether or not people had actually read the instructions. We can only conclude from the errors that a lot of them had not read them.

6.3.1.3.2 Determine whether tooltips increase performance

The following hypotheses was made:

- **Hypothesis**: Performance for the test with tooltip help is better than performance for the test without tooltip help.

  Even though we see a difference in the level of correctness, results from the t-test indicate that there is no statistical significant difference between the test with tooltip help and the test without tooltip help. Consequently, we need to reject the hypothesis:

$$t(10) = 0.077395, \text{ with } \alpha = 0.05$$

NOTE: Even though t-test results suggest that we need to reject the hypothesis, we are still under the impression that initial learning is helped by tooltips. The amount of feedback we got on the decodability web test with tooltip help may not have been enough for this early conclusion.

6.3.1.3.3 Learn what needs to be changed to the system

The decodability test indicated couple of places where the system could possibly use some adjustment:

- In order to eliminate the issues mentioned in the section on partial correctness, we need to clearly address the issues in the instructions

  The following table shows the types of partial correctness and suggests how these errors can be eliminated.

Partially correctness	Eliminate by
Correct but didn't consider underlying information	Covering in the instructions (explaining how the interface works)
Correct, but didn't consider the noun conjunction	Covering in the instructions (explaining how the interface works)
Correct except some icons were recognized incorrectly	Create more recognizable icons as pointed out by the icon test
Correct except sentence not grammatically correct, but is the meaning close to the one intended	Covering in the instructions (explaining how the grammar works)
Correct, but some grammatical element is not present in the translation	Covering in the instructions (explaining what a sentence in VIL consists of)

- Clear instructions increase performance (this is concluded by talking to acquaintances, who took the test).
- The asterisk as a sole marker for modifying properties may have proven to be insufficient. We possibly need an additional marker, flashing or color or some other way to have the modified icons stand out more.
- We have learned which program icons need some redesigning, e.g. the "duration" case was misinterpreted sometimes because of the icon.

6.3.2 The Live Tests

The encoding tests served the following purposes:

1. To demonstrate learnability and retention
2. To determine whether a small teaching period increases performance
3. To determine whether VIL is encodable enough
4. To determine subjective satisfaction with the system
5. To learn what needs to be changed to the system

For the purpose of demonstrating learnability and retention we want to confirm the following hypotheses:

A) Subjects complete subsequent sentences of **smaller or equal grammatical complexity** faster than previous sentences.
B) Subjects make **fewer mistakes** on subsequent sentences of smaller or equal grammatical complexity.
C) Subjects select subsequent instances of the **same item** from the iconicon faster than initial instances of that same item.
D) Consequently, the time difference between the first and second instance (for the same item) is much bigger than the time difference between the second and subsequent selections.
E) Subjects select an item that belongs to the **same category** as a previously selected item faster than the ones previously selected from that category. For example "to drink" is selected faster than "to eat", since they belong to the same category "Ingest verbs" and "to drink" is selected after "to eat".
F) Performance in the retention test is better than performance in the initial test.

For the purpose of determining whether a small teaching period increases performance we want to confirm the following hypothesis:

G) Performance of the group with teaching is better than the performance of the group without teaching.

For the purpose of determining whether VIL is encodable enough we want to confirm the following hypothesis:

H) It is not harder to encode sentences in VIL than to translate sentences into another language with a dictionary.

6.3.2.1 Instructions and Teaching for the Live Test

The live tests consisted of three separate tests:
- An encoding test without teaching
- An encoding test with teaching
- A retention test

Instructions for the live test were conducted in the following way:

1. For the test with no training the subjects received short instructions for about 3 minutes. These instructions were the same as the instructions on the web test,

basically explaining the different grammatical elements and how the interface works.

2. The teaching period for the test with training was about 10 minutes. The following items were explained to the subjects:

- The grammar of the language with their visual elements
- The organization of the iconicon: the way the verb system and the nouns are organized. Explaining the organization of the iconicon is very helpful, because if a person knows the criteria for splitting up categories, e.g. verbs are divided up into verbs that focus on the result of the action and verbs that focus on the activity itself, then he/she will be able to find the words that he/she is looking for much faster. The fastest test subject did work exactly according to this. Upon seeing the verb "to boil", he asked himself "what kind of verb is this", then quickly answered himself "transformation" and went to the right location.
- The subjects were shown how to compose messages.

6.3.2.2 The Test Sentences

For the live test we also created 10 sample sentences. Eight of these sentences were simple declarative sentences, one sentence involved an imperative sentence, and one sentence was an interrogative sentence. The last two declarative sentences were pretty simple in grammatical complexity. They were included to check if people had learned from the system. We expect people to complete these last two sentences faster that the previous sentences.

1. I eat a hamburger.
2. I drink milk 3 times a day.
3. I put a pork chop in a frying pan.
4. I put (past tense) butter on my bread with a blunt knife.
5. I baked you a cake from red apples and yellow bananas.
6. In France, I ate 10 long French breads last summer.
7. Stir all eggs for a long time!
8. What are you eating?
9. He drinks milk with a glass.
10. You heated a sausage.

6.3.2.3 How the Live Tests Works

For the live test we changed the program so that it guides the test subjects through the sentences. The user has to type in his nationality and click the "Start Test" button. The first sentence appears. Whenever a subject feels he/she has completed a sentence, he/she clicks the "Check Sentence Correctness" button and the system checks the

composed message for correctness. If the sentence is not correct, the system tells the user which element is incorrect, e.g. a system message could be "The agent is incorrect". If the sentence is correct, the system congratulates the user and the second test sentence is displayed in the test frame.

Whenever the user has checked the same test sentence for correctness more than four times and the sentence is still incorrect, the system offers to stop this sentence and show the correct version. The user can decline this offer and continue to get the sentence right, if he/she wishes, or study the correct sentence offered. If the user has looked at the correct sentence, he/she must click on "CONTINUE TEST" to continue with the next sentence in the test.

When the user has completed all 10 sentences, the user is thanked for his/her participation and the system asks if the user wants the system to submit all test results by e-mail. If so, the system submits it. If not, the user has to tell the instructor that he/she is done.

6.3.2.3.1 What is measured by the system

The following data is recorded automatically by the system:
1. For each verb, noun, or adjective, the system records how long the user took to select the item in question plus the path traversed to the items, including backtracking.
2. For each sentence and for every attempt, the system records the time to completion. An attempt is when a user click on the "CHECK SENTENCE CORRECTNESS" button. It also records whether the attempt is correct and a copy of the completed attempt is saved.

At the end of a user test we are left with the following information
1. A directory with VIL messages. These are all attempts the user made.
2. A database with two tables. One with the time to each attempt, a link to the file in the directory, and their correctness. The other table gives us all data about the choices from the iconicon (path traversed + time).

6.3.2.4 Test Subjects

The encoding test without teaching was held at a local college, the "Grafisch Lyceum Rotterdam". We had a lab full of students, 23 in total. In the lab were 25 computers and a beamer. We used the beamer to give some short instructions, after which the test started.

The encoding test with teaching was held at a company in the Hague, KPMG. There were 8 people taking the test with teaching. Each of them got individual teaching.

6.3.2.5 The Live Tests Results

The live test yielded the following information:

- *Number of sentences encoded*: 310 sentences were encoded by subjects
- *Nationality*: 31 persons from 7 countries took the test. The following table shows the nationalities of the subjects along with the number of test subjects from each country:

Nationality	Total persons
Netherlands	25
Ireland	1
Surinam	1
Sweden	1
India	1
Bulgaria	1
Turkey	1

6.3.2.5.1 Interpreting the test results

The following table shows the average time (in minutes) test subjects needed to get each sentence correct for the test without teaching. It also shows the fastest and slowest time per sentence.

Nr	Sentence (without teaching)	Min	Max	Avg
1	I eat a hamburger.	0.8	22.7	5.6
2	I drink milk 3 times a day.	0.7	6.4	2.4
3	I put a pork chop in a frying pan.	1.6	10.2	5.4
4	I put (past tense) butter on my bread with a blunt knife.	2.9	10.7	7.4
5	I baked you a cake from red apples and yellow bananas.	2.5	11.9	6.1
6	In France, I ate 10 long French breads last summer.	1.0	8.9	6.1
7	Stir all eggs for a long time!	1.5	12.1	4.9
8	What are you eating?	0.7	4.0	1.8
9	He drinks milk with a glass.	0.4	4.2	1.7
10	You heated a sausage.	1.0	4.2	2.2

The following table shows the average time (in minutes) test subjects needed to get each sentence correct for the test with a 10 minutes teaching period:

Nr	Sentence (with 10 minutes teaching)	Min	Max	Avg
1	I eat a hamburger.	1.0	3.0	1.9
2	I drink milk 3 times a day.	1.0	2.0	1.6
3	I put a pork chop in a frying pan.	2.0	5.0	3.4
4	I put (past tense) butter on my bread with a blunt knife.	2.5	7.0	4.1
5	I baked you a cake from red apples and yellow bananas.	2.5	6.0	4.3
6	In France, I ate 10 long French breads last summer.	2.0	4.0	3.3
7	Stir all eggs for a long time!	1.0	5.0	2.4
8	What are you eating?	1.0	1.0	1.0
9	He drinks milk with a glass.	1.0	2.0	1.4
10	You heated a sausage.	1.0	2.0	1.6

If we compare the average times for each sentence for the two groups, we see the following (in minutes):

Nr	Sentence	Avg without teaching	Avg with teaching
1	I eat a hamburger.	5.6	1.9
2	I drink milk 3 times a day.	2.4	1.6
3	I put a pork chop in a frying pan.	5.4	3.4
4	I put (past tense) butter on my bread with a blunt knife.	7.4	4.1
5	I baked you a cake from red apples and yellow bananas.	6.1	4.3
6	In France, I ate 10 long French breads last summer.	6.1	3.3
7	Stir all eggs for a long time!	4.9	2.4
8	What are you eating?	1.8	1.0
9	He drinks milk with a glass.	1.7	1.4
10	You heated a sausage.	2.2	1.6

We can draw the following conclusions from the tables above:

- **Teaching (hypothesis G)**: A short teaching phase increases performance. This hypothesis G is confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

$$t(18) = 4.234375$$

This is as expected since these people do not have to find out for themselves how the program works and how the language is organized. Consequently, for each sentence the performance of test subjects was better in the test where a short teaching period preceded the test.

- **Learnability (hypothesis A)**: Learnability did occur. Sentence 9 and 10, which do not introduce new grammatical features, were encoded faster than the previous sentences. This already manifests itself in sentence 8, where the performance is already good even though a new sentence type is introduced, i.e. the interrogative sentence. The average times confirm the hypothesis A that subjects complete subsequent sentences of smaller or equal grammatical complexity faster than previous sentences.

- **New complexity**: The time to complete a sentence increased as the sentences got more complex and new elements were introduced. The time to complete sentences 3, 4, 5, and 6 took the longest time.
- **No new complexity**: When a sentence did not introduce new complexity or when a word was used that belonged to the same category as a word that occurred in a previous sentence, the time to complete did not increase anymore or even decreased. For example, sentences 8, 9, and 10 were completed fairly quickly.
- **Differences between min and max**: The differences between the minimum and maximum times per sentence are larger for the group without teaching (standard deviations for the min-max differences between the groups are respectively 5.48 and 1.49 for the groups without and with teaching). The reasons that account for this difference are:
  - In a group were people get individual teaching the differences between people are expected to be smaller.
  - In a small group the test can be controlled better. We could not constantly monitor the big group. People were sometimes distracted and started a conversation with their neighbor and after a while returning to the sentence they were working on. We could not measure **idle time**. In other words, in the case of the maximum time of 22,7 minutes for sentence 1 "I eat a hamburger", it is hard to determine whether all this time was spent trying to compose the sentence or that there was a large portion of idle time.
  - Also some people did not pay attention during instructions, whereas in the group with teaching each test subject received individual training.

6.3.2.5.2 Number of encoding chunks per sentence

So far, we have looked at learnability on basis of grammatical complexity of the sentences. We have seen that subsequent sentences of equal or less grammatical complexity are completed faster than previous ones. However, we can also do some normalization on the sentences themselves. If we look at the number of items that need to be encoded for each sentence, learnability should show us that each sentence should be completed faster even when new grammatical features are introduced. The number of items as a proxy for complexity is maybe a little too crude. Some items are harder to choose, for example they involve several mouse clicks or choices to make (like the noun modifiers). The following table shows the number of items to encode per sentence and the number of complexity points given to the sentence. If the complexity points are higher than the number of items to encode, then the last column describes what accounted for these extra points:

Nr	Sentence	Nr items	Nr chunks complex	Explanation of extra complexity
1	I eat a hamburger.	3	3	
2	I drink milk 3 times a day.	5	5	
3	I put a pork chop in a frying pan.	4	6	The verb "to put" is hard to select. Also, "frying pan" is a complex noun (see earlier remark about misinterpretation).
4	I put (past tense) butter on my bread with a blunt knife.	8	11	First time time when is used. User has to scan all options. Also, two modifiers.
5	I baked you a cake from red apples and yellow bananas.	9	11	Also, two modifiers.
6	In France, I ate 10 long French breads last summer.	7	11	First time place where is used. User has to scan all options. Also, two modifiers.
7	Stir all eggs for a long time!	4	6	One modifier. Also, the verb "to stir" to complex to select.
8	What are you eating?	3	3	
9	He drinks milk with a glass.	4	4	
10	You heated a sausage.	4	4	

The following table shows the time per encoding unit for our measure of complexity:

Sentence nr	Average time per encoding unit
1	1.9
2	0.5
3	0.9
4	0.7
5	0.6
6	0.6
7	0.8
8	0.6
9	0.4
10	0.5

The table shows that after the first sentence the encoding for each unit went down.

6.3.2.5.3 Measuring attempts for the group without teaching

For the largest test group (the group without teaching), we also measured how many attempts people needed to get to the correct answer. With each attempt the system

checked the sentence for correctness and gave feedback about the items that were still incorrect. The next table shows the average number of attempts for each sentence and the average time in minutes to the first attempt.

Nr	Sentence (without teaching)	Avg. attempts	Avg. time to first attempt
1	I eat a hamburger.	1.6	4.9
2	I drink milk 3 times a day.	2.1	1.9
3	I put a pork chop in a frying pan.	2.5	4.4
4	I put (past tense) butter on my bread with a blunt knife.	3.7	4.5
5	I baked you a cake from red apples and yellow bananas.	3.3	4.0
6	In France, I ate 10 long French breads last summer.	4.8	3.4
7	Stir all eggs for a long time!	2.0	4.0
8	What are you eating?	2.3	1.1
9	He drinks milk with a glass.	1.7	1.6
10	You heated a sausage.	2.6	1.6

From the table we can see that with increasing grammatical complexity, the average number of attempts also increases, along with the time to the first attempt. However, whenever the complexity is the same or decreases, then also the time to the first attempt. For example, sentence 1 and 10 are equal in grammatical complexity. The time to the first attempt decreased from almost 5 minutes to 1.6 minutes.

For each of the 25 subjects and for each sentence, we studied the errors in every attempt that was made. The following is a table with the most common errors made by subjects for each sentence. The second column shows, which mistakes were made most, and the last column shows how many subjects made this mistake.

Sentence	Most common mistakes	Nr
I eat a hamburger.	Hamburger as instrument	2
I drink milk 3 times a day.	Milk as instrument	1
	Duration instead of frequency of occurrence	4
I put a pork chop in a frying pan.	Wrong pan	6
	Porkchop as source	2
	Porkchop as instrument	2
I put (past tense) butter on my bread with a blunt knife.	Bread knife as knife	5
	Forgot "my" in bread	10
	Forgot "blunt" knife	2
	Knife as source	2
	Butter as source/bread as theme	2
	Butter instrument/bread source/knife theme	1
	Forgot time-when case (past tense)	10
I baked you a cake from red apples	Forgot noun conjunction	1

and yellow bananas.	Cake as goal	2
	Forgot goal	2
	You-instrument; cake-source; apples-theme; bananas-goal	1
	Forgot time-when case (past tense)	4
In France, I ate 10 long French breads last summer.	Try to fill in "last summer" in duration instead of time-when	7
	Property "10" of "French bread" as frequency of occurrence	5
	Forgot "long" or "10"	10
	Forgot time-when case	2
Stir all eggs for a long time!	Egg as source	2
	All as frequency of occurrence	2
What are you eating?	Forgot to select questioned field ("?")	10
He drinks milk with a glass.	-	
You heated a sausage.	-	

From studying the attempts we conclude the following:

- The above table about mistakes confirms the **hypothesis B** that subjects make fewer mistakes on subsequent sentences of smaller or equal grammatical complexity. This is argued by the fact that for the last two sentences, which introduce no new complexity, there are no mistakes made. The average time to a first attempt went down as well, indicating that subjects gained more confidence.
- The average number of attempts, however, did not decrease. The average number of attempts per sentence could have been lower since:
  - A lot of people checked the same sentence multiple times for correctness (50 times in total).
  - Some people checked for correctness every time they filled in a grammatical case (20 times in total).
- The number of attempts was also negatively influenced because the system very strictly checked the sentences for correctness. A sentence was either completely right, as we designed it, or otherwise the attempt was marked incorrect. For example:
  - In sentence 4, five subjects selected a "bread knife" instead of a regular "knife", which was logical considering we were talking about putting butter on bread with a knife. Yet, these attempts were rejected by the system.
  - For sentences were the tense was "present tense", the system expected the time-when slot to be left uninstantiated. However, four subjects selected "present" in the time-when "time pronoun". These, sentences should have been approved of course. This was an error in the grading program, which caused the sentences to be listed as errors in the data above.

- Some people had trouble with the fact that the test was in English. This also resulted in unnecessary attempts. For example,
  - 6 subjects did not know what a "frying pan" was and consequently chose the wrong pan.
- The time-when (tense) case posed most problems, accounting for 22 errors. This means that some additional work has to be done here. Two solutions that would solve a lot of problems are:
  - Remove the "present" icon from the possible selections. People should leave the time-when case uninstantiated when the tense of the sentence is "present tense".
  - Disable the "temporalizer", whenever the "time pronoun" is selected. A common mistake was that, whenever the tense was "past tense", subjects selected the "before" icon from the "temporalizer". They cannot both be selected at the same time.
- There was some confusion about the difference between duration and frequency of occurrence. This could be explained better in the instructions.
- Some subjects did not understand how to create a questioned field.
- Some subjects did not grasp the design of the language at first. They filled in the case slots in a linear fashion, from left to right and top to bottom. In other words, they did not look at the grammatical case for each slot. However, by looking at test results for individual people, we have seen that subjects, who put the icons in the wrong case slots in the first few sentences, did learn and did not make these mistakes in later sentences. From the table it can be seen that in the last five sentences people hardly made any errors determining which parts of the sentence correspond to which grammatical cases in VIL.

6.3.2.5.4 The organization of the iconicon

The encoding test was also a good means of testing our organization of the iconicon. Did the test subjects find the criteria, which we used in designing our hierarchies, logical? For the largest test group (the group without teaching), we also measured the time a test subject took to find an item in the iconicon (i.e. verbs, nouns, and adjectives), and also the path traversed to the item. We expected it would take test subjects longer to find the right verb, than it would take them to find the right noun. Likewise, we expected it would take test subjects longer to find a noun than it would take them to find a noun modifying property (adjective). Verbs are harder to find than nouns and adjectives, because test subjects have to understand the criteria for splitting up verbs into categories, e.g. verbs are divided up into verbs that focus on the result of the action and verbs that focus on the activity itself. For nouns this division is easier for subjects to grasp, since the test mostly dealt with concrete nouns. For instance, upon seeing the verb "to boil", people had recognize that this is actually a

"transformation" verb, and then recognize that "transformation" verbs are "creation" verbs, which fall under verbs of "existence". Compare this to looking for the noun "egg", which is a dairy product, which fall under food, which is part of the living world. Adjectives are easier to find than nouns, because the hierarchy is only one level deep.

To test the organization of the hierarchies and demonstrate learnability, we logged the time it took subjects to select a verb, noun, or adjective. The results should confirm the hypotheses C (subjects select subsequent instances of the **same item** from the iconicon faster than initial instances of that same item), hypothesis D (the time difference between the first and second instance for the same item is much bigger than the time difference between the second and subsequent selections, and hypothesis E (subjects select an item that belongs to the **same category** as a previously selected item faster than the ones previously selected from that category, for example "to drink" is selected faster than "to eat", since they belong to the same category "Ingest verbs" and "to drink" is selected after "to eat").

The next table shows the average time in seconds people needed to select the same item from the iconicon. Therefore, the table only shows times for words that occurred more than once in the test. For each word, the first row shows the time for the initial instance and the next rows show the average time in seconds on subsequent instances of the same item.

Type	Item	Sentence	Avg
Noun	I	1	11.7
Noun	I	2	3.0
Noun	I	3	3.8
Noun	I	4	3.9
Noun	I	5	3.0
Noun	I	6	3.6
Noun	milk	2	23.0
Noun	milk	9	12.0

Type	Item	Sentence	Avg
Verb	Drink	2	8.9
Verb	Drink	9	7.6
Verb	Eat	1	19.8
Verb	Eat	6	11.3
Verb	Eat	8	10.9
Verb	Put	3	33.5
Verb	Put	4	23.9

The table confirms **hypothesis C** and **hypothesis D**. Subsequent instances of the same item are indeed identified faster than initial instances. Also, the difference between the first and second instance is much bigger than the difference between the second and subsequent instances, if any. Compare the time for the verb "to eat" for the first sentence (19.8 seconds) to the times for sentence 6 and sentence 8 (both 11.3 seconds).

The next table shows the average time in seconds people needed to select an item from the iconicon that belonged to the same category as a previously selected item.

Therefore, the table only shows times of items for which another item from that same category occurred in the test. For each category, the first row shows the time for the initial item in the category and the next row shows the average time for subsequent new items belonging to that category.

Type	Category	Avg
Noun	Dairy products	23.7
Noun	Dairy products	12.5
Noun	Grain products	15.6
Noun	Grain products	12.5
Noun	Meats	44.8
Noun	Meats	31.6
Noun	Pronoun	11.7
Noun	Pronoun	3.8
Noun	Silverware	35.6
Noun	Silverware	34.2

Type	Category	Avg
Verb	Ingest	20.3
Verb	Ingest	8.9
Verb	Transformation	37.8
Verb	Transformation	27.5

The table confirms **hypothesis E**. Subsequent items from the same category are indeed identified faster than initial items. Once people knew where to look for an icon, they became more and more proficient in using the system. For example, the pronoun "I" took several seconds to find in the first sentence, but was found much faster in subsequent sentences. Also, people learned to find an icon, which belonged to the same category as an icon previously used, faster. They learned to grasp the organization of the hierarchies. For example, people learned that a verb like "to boil" (a transformation verb) was likely to occur in the same category as "to bake", and consequently found it faster.

The following table shows two more findings:
1. Subjects select adjectives faster than nouns and verbs.
2. Subjects select nouns faster than verbs.

Type	Avg
Adjective	6.3
Noun	15.6
Verb	23.3

For the whole test, adjectives are selected on average in 6.3 seconds, nouns in 15.6 seconds and verbs take the longest with 23.3 seconds. These hypotheses are also demonstrated in the next three more detailed tables. In these tables we show for each hierarchy, which elements took the longest to find and how many mistakes were made in finding that element. The following table shows the time in seconds it took a subject to find each *first* occurrence (in the test) of a verb. It also shows how many

wrong branches subjects took to get to each correct verb, i.e. how many times was backtracking necessary:

Verb	Avg	Number of times backtracking was needed									
		0	1	2	3	4	5	6	7	8	9
Eat	19.8	18	3	4	1						
Drink	8.9	20	2								
Put	33.5	9	5	3	2	2				1	
Bake	37.8	3	5	5	4	3	1			1	
Stir	48.1	8	2	5	1	3		2	1		1
Heat	27.5	5	6	2	4	1	1				

From the table we see that the hardest verbs to find were "to stir" and "to bake". Subjects had to realize the verbs belong to "connect verbs" and "transformation verbs", which are both part of "create verbs". The verbs "to eat" and "to drink" are the easiest to find, since they need the least backtracking, with resp. 18 and 20 subjects that needed zero backtracking. Zero backtracking means that the subject immediately went to the right location, i.e. took all correct branching decisions. We take the number backtracking as the measure for difficulty, since we can not assume that all time spent looking up a noun was actually spent on it and was not caused by distraction or something else. Furthermore, we see again the verb "to drink" is recognized faster than "to eat", since these verbs belong to the same category "Ingest verbs", also "to heat" is recognized faster than "to bake" ("Transformation verbs").

The following table shows the time in seconds it took a subject to find each *first* occurrence (in the test) of a noun, along with the number of times backtracking was necessary ("up" in the tree):

Noun	Avg	Number of times backtracking was needed									
		0	1	2	3	4	5	6	7	8	9
burger	23.3	17	4	1	2						
I	11.7	22	3	1	1		1				
milk	23.0	12	11	1	1		1				
porkchop	44.5	8	3	3	3	1					
frying pan	54.7	7	3	1	1	2	1	1			2
bread	16.8	16	3	2							
butter	13.8	16	4	1							
knife	34.6	7	2	3	2	2	2				
apple	9.5	15	2	1							
cake	19.5	9	8	2	1						
you	3.6	18									

Noun	Avg	Number of times backtracking was needed									
		0	1	2	3	4	5	6	7	8	9
French bread	12.0	17	4								
egg	10.0	15	5								
glass	33.5	5		3	6	3			1		
he	4.9	22									
sausage	31.6	10	1	2	1	1			1		

From the table we see that subjects had most difficulty with "frying pan", "knife" and "glass". Even though some other nouns took longer to find, less backtracking was necessary for these nouns. The fact that more backtracking was needed for some nouns can have several reasons:

- *Subjects did not understand the meaning of the word*. The test sentences were in English. For example, the noun "frying pan" was hard to find, since many (Dutch) subjects did not know the meaning of the word. They were looking for a "deep fryer", which is not in the system yet.
- *The icon representing the noun is not recognizable enough*. For example, the noun "knife" looked odd to some people, causing them to backtrack and look for it somewhere else (eventually returning to the correct icon).
- *The noun or category is classified illogically*. During beta testing of VIL we noticed that testers had problems finding the "food" category as a child of the "humans" category. It did not sound logical to them that to look for a "hamburger" one had to select the "humans" category first. Consequently we moved the category up one level, such that it is on the same level as "humans" together with categories like "animals", "fruit", "vegetables", "plants", and "evolution".
- *The tooltip describing the icon adds confusion*. For example, the category name "silverware" may have incorrectly suggested this is about knifes, forks, etc. However, what is meant is "objects are used for eating and drinking". Consequently, the noun "glass" could have been hard to find for some subjects, since it is classified under "silverware". This is an error in the program that may have contributed to worse results.

The following table shows the time in seconds it took a subject to find each *first* occurrence (in the test) of an adjective, along with the number of times backtracking was necessary:

Adjective	Avg	Number of times backtracking was needed									
		0	1	2	3	4	5	6	7	8	9
blunt	8.3	26									
red	3.0	28									

Adjective	Avg	Number of times backtracking was needed									
		0	1	2	3	4	5	6	7	8	9
yellow	3.1	26									
long	14.1	11	3	2							

From the table we see that adjectives are the easiest to find. Subjects knew immediately were to go for an adjective (almost no backtracking). This little backtracking is partly due to the hierarchy being only one level deep.

6.3.2.5.5 Issues

As mentioned before, one major problem with the test without teaching was that it was impossible to measure how long a person actually took to complete a sentence, i.e. we cannot measure idle time. We held the test at a school and many people were taking the test simultaneously. We could not keep track of who was doing what for everybody. We noticed that a lot of people were talking to each other during the test. For example, a person started a sentence, was distracted by something, and then continued the sentence. Meanwhile the timer continued running. This accounts for some major differences in performance for some sentences. For example, the fastest recorded time for the sentence "I eat a hamburger" was under a minute, while the slowest was well over 22 minutes.

Another problem was that not everybody paid attention during the instructions. Even though we asked people to pay attention, there were still people not paying attention and talking to each other during the instructions. This also helps to account for differences in performance.

Also, we need to keep in mind that there is only so much new information a person can remember. So, for the test group with teaching, there may have been things that were explained to the subjects, but were forgotten at the time of the test.

It turns out that subjects, who relied on tooltips the most, took the longest to complete the sentences. Some looked at the tooltips so much that they looked at every icon, waiting for the tooltip to appear, before making a selection. The fastest person was the one who did not look at tooltips at all. This person only looked at the tooltip just before selecting an icon, double-checking whether the icon he was going to select was indeed the right one.

Finally, there is the issue of sample size. What is an ideal sample size for our purpose? Since our target population is everybody in the world, we should really have used, for a 95% confidence, over 350 test subjects. This number of subjects proved to be too difficult to achieve. However, the test subjects that were used produced a large

number of answers (i.e. response items). Consequently, the t-test results have been based on the number of response items.

6.3.2.6 Testing retention over time

The 8 test subjects from the group with teaching were asked to take the test one week later. The subjects received no help anymore and had to rely on what they had learned and remembered from the first test. This second test was used to test the hypothesis that retention did occur and the performance is better.

The following table confirms the hypothesis F by showing that the subjects had really learned from their previous experience with the system. Almost all test subjects remembered how the system worked. The table shows the results of the second test for the group test subjects, who had received teaching in the first test. The last column is for comparison. It shows the average time to complete each sentence from the fist test (see above). As can be seen, the average time to complete a sentence went down for all sentences.

Nr	Sentence (with teaching)	Min	Max	Avg	Avg Test 1
1	I eat a hamburger.	0.5	1	0.8	1.9
2	I drink milk 3 times a day.	0.5	1.5	1.1	1.6
3	I put a pork chop in a frying pan.	1.5	2	1.6	3.4
4	I put (past tense) butter on my bread with a blunt knife.	2	3	2.4	4.1
5	I baked you a cake from red apples and yellow bananas.	2	4	2.6	4.3
6	In France, I ate 10 long French breads last summer.	1.5	3	2.2	3.3
7	Stir all eggs for a long time!	1	1.5	1.3	2.4
8	What are you eating?	0.5	1.5	0.7	1.0
9	He drinks milk with a glass.	0.5	1.5	1	1.4
10	You heated a sausage.	0.5	1.5	1.1	1.6

The **hypothesis F** that performance in the retention test is better than in the first test is confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

$$t(18) = -2.34996$$

6.3.2.7 Subjective Satisfaction

To test subjective satisfaction we developed a questionnaire. Test subjects were asked to fill out the questionnaire after the test. In 12 questions, test subjects were asked what they thought of the user interface, the language, and the test. We got a total of 25 questionnaires back. The following shows the questionnaire.

VIL Questionnaire

Thank you for taking the VIL test. We would like to know how you feel about the program and language. Please take some time to fill in this review. It is very important to us.

What is your nationality?	
What is your age?	
What is your gender?	☐ Male ☐ Female
What languages do you speak?	

Please rate the following questions with a number between 1 and 5. The numbers are explained below:

1 – Strongly Disagree/Very Bad
2 – Disagree/Bad
3 – No Opinion/Neither Good or Bad
4 – Agree/Good
5 – Strongly Agree/Very Good

User Interface	1	2	3	4	5
The program is intuitive.	☐	☐	☐	☐	☐
The interface is user friendly.	☐	☐	☐	☐	☐
The program icons are clear (e.g. the "save" icon).	☐	☐	☐	☐	☐
The grammar icons are clear (e.g. the "time when" or "agent" icon).	☐	☐	☐	☐	☐
The icons in the icon dictionary are clear (e.g. "apple").	☐	☐	☐	☐	☐
Language	**1**	**2**	**3**	**4**	**5**
The language is organized clearly.	☐	☐	☐	☐	☐
It is faster to encode a sentence in VIL than to translate it into a different language (e.g. looking up words to translate, etc.).	☐	☐	☐	☐	☐
An iconic communication language to allow speakers of different nations to communicate is useful.	☐	☐	☐	☐	☐
Test	**1**	**2**	**3**	**4**	**5**
I think the test was difficult.	☐	☐	☐	☐	☐
I got more confident on subsequent sentences (did you learn while taking the test).	☐	☐	☐	☐	☐
It was easy to analyze the sentences and map its elements to grammatical cases in the program (e.g. what is the source, what is	☐	☐	☐	☐	☐

the goal, what is the object, etc.).					

Do you have any other comments? Please fill them in below.

Thanks a lot!

6.3.2.7.1 Subjective Satisfaction – Questionnaire Results

The questions about age, gender, and languages that people speak were only used to find out more about the test subjects. We did not use them in the analysis of the questionnaire. The age of the test subjects ranged from 17 to 31 and the average age was 20,4. There were 22 male and 3 female subjects. Almost all test subjects spoke Dutch and English, with the foreign test subjects speaking their native language instead of or besides Dutch. Five subjects also spoke German and 3 also spoke French.

In the next table the average of the answers to the questions in shown. Anything above 3 is good, except for the question about the difficulty of the test, for which it is the other way around.

User Interface	Avg.
The program is intuitive.	3.8
The interface is user friendly.	4.2
The program icons are clear (e.g. the "save" icon).	4
The grammar icons are clear (e.g. the "time when" or "agent" icon).	3.6
The icons in the icon dictionary are clear (e.g. "apple").	4.2
Language	**Avg.**
The language is organized clearly.	3.8
It is faster to encode a sentence in VIL than to translate it into a different language (e.g. looking up words to translate, etc.).	3.4
An iconic communication language to allow speakers of different nations to communicate is useful.	3.7
Test	**Avg.**
I think the test was difficult.	3
I got more confident on subsequent sentences (did you learn while taking the test).	3.8
It was easy to analyze the sentences and map its elements to grammatical cases in the program (e.g. what is the source, what is the goal, what is the	3.8

object, etc.).	

The following conclusions can be drawn:

- **User Interface**: people found the user interface good. They thought it was user-friendly and that the icons were designed well. The grammar icons rank a bit lower and this can be accounted to the fact that almost all of them are conventional.
- **Language**: the language was organized clearly and also a lot of people could see the use for an iconic language. The reason why people rank lower the question that it is easier to encode a message in VIL than to translate it with a dictionary can be accounted by the fact that people had never used the system before. They were in the process of learning how to encode a message in VIL, whereas everybody knows how to use a dictionary. They did not take into account though, that the sentence produced when translating a sentence with a dictionary, is probably grammatically incorrect too.
- **Test**: people did not find the test hard or easy. However, they did feel they got more confident on subsequent sentences. Furthermore, they did not have too much trouble mapping the sentence elements to grammatical cases in the program.

6.3.2.8 Is VIL encodable enough?

To determine if VIL is encodable enough is hard to measure. We need to measure speed and accuracy and compare it against some other entity that we can measure. As a means for comparison we translated the ten sentences into Swedish by looking up the words in the dictionary, measured the time it took, and looked at the resulting translations. As a comparison, we had an experienced VIL user encoding the ten sentences. These results are reflected into the next table, which in the third column shows the average time in minutes to translate the ten sentences and in the fourth column the average time the experienced user took to complete the sentences. Besides comparing the time for translation to an experienced VIL user, we also compared the time to the averages times of one of the groups from the live tests described above (the last column):

Nr	Sentence	Dictionary lookup	Experienced user	Avg for new users with teaching
1	I eat a hamburger.	1.8	0.2	1.9
2	I drink milk 3 times a day.	3.1	0.4	1.6
3	I put a pork chop in a frying pan.	1.8	0.4	3.4
4	I put (past tense) butter on my bread with a blunt knife.	3.8	0.7	4.1
5	I baked you a cake from red apples and yellow bananas.	3.4	0.8	4.3
6	In France, I ate 10 long French breads last summer.	2.7	0.6	3.3
7	Stir all eggs for a long time!	1.6	0.4	2.4
8	What are you eating?	1.1	0.2	1.0
9	He drinks milk with a glass.	1.5	0.4	1.4
10	You heated a sausage.	1.0	0.4	1.6

From the table we see that an experienced VIL user creates the test sentences very quickly. We have already confirmed learnability and it is expected that users of VIL will eventually approach this performance. The t-test results when comparing dictionary lookup with the experienced user show a big difference:

$$t(18) = 5.404784, \text{ with } \alpha = 0.05$$

From the table we also see that the time needed to do the translation can be roughly compared to the test group who had received some teaching on how VIL works (it is more reasonable to compare to this group than to compare to the group that did not receive any training, since the dictionary lookup was also done by a person, who is very experienced in using dictionaries). Results from a t-test confirms the **hypothesis H** that there is no statistically significant difference between translating sentences with a dictionary and using VIL (unexperienced users):

$t(18) = -0.65356$, with $\alpha = 0.05$

The translation was done into Swedish only. Should we have wanted to communicate these sentences to more languages then the times would of course be much higher.

Furthermore, the translation was done from English to Swedish and checked for correctness by a Swedish person. Aside from inflectional errors there were few mistakes. This is also due to the fact that the word orders in both languages are much alike, so that translating could practically be done on a word-by-word basis. The Swedish language also does not contain any difficult letters. Should one have translated to Chinese, then not only would the sentences have been grammatically more incorrect, but it would also have taken much longer to write down the Chinese characters. Also, the topological relatedness, including grammatical, morphological,

and vocabulary similarities, is very relevant for accuracy or how much is lost in the communication process.

6.3.2.9 What about machine translation?

There now exists a variety of machine translation systems implemented for a variety of language pairs, but because of the size of the task of developing implementations for new language pairs, progress has been slow in consideration of the large number of languages, including some of considerable size and world importance, that are still not provided for. The best known MT systems only translate about 7 language pairs.

In VIL you do not need to worry about extending to a new language. In VIL we do not have the time consuming task of adding a new dictionary, taking into account the grammatical structure of the new language, and moreover dealing with difficult idiom issues to resolve ambiguity in translation (adding pragmatic/semantic knowledge), if at all possible[163]. "Current machine translation systems are consistently plagued with problems concerning extensibility to new languages or domains."[164]

"As human language consists of the single elements of word, phrase, sentence and text, the number of identification possibilities even explodes with elements combined. To resolve such ambiguity, additional pragmatic/semantic knowledge is needed, which is provided by human knowledge. Software, however, is not yet able to really understand text."[165]

In VIL we do need to translate the instructions, but these are only a few pages. If we stick with tooltips, then we need to provide these as well. However, the effort involved is much less complex than that for MT and consequently takes less time, since this translating only involves word-to-word translation. Also MT systems try to translate everything and have a hard time coping with words that have multiple meanings, especially when the context is not clear. Most systems we tried do not recognize May and June in the following sentence as proper names, and also interpret blue incorrectly as a color:

<p align="center">Why are May and June blue?</p>

---

[163] **Volk, M.** *"The Automatic Translation of Idioms. Machine Translation vs. Translation Memory Systems,"* in Nico Weber (ed.): Machine Translation: Theory, Applications, and Evaluation. An assessment of the state of the art. St. Augustin: gardez-Verlag. 1998.

[164] **University of Maryland**, *"Research into automatic dictionary construction in machine translation,"* http://www.umiacs.umd.edu/labs/CLIP/mt.html.

[165] **Bauer, E.** *"Machine Translation - 2001 has already arrived,"* Technical Communications Forum, December 2000.

Many words have one or more meanings in English. These multiple meanings often do not translate to a foreign language with the same variety of meanings. As human users, we can identify the meaning of each sentence through context. Computers, at least for now, are incapable of consistently making distinctions at this level.[166]

"Real-time MT of the AltaVista/Systran type is free or very cheap (1 cent per word). It is very fast. There is no pre- or post-editing, the text is called a "suggestion", and AltaVista offers a needed disclaimer. The language is uncontrolled and there is no access to databases or online dictionaries. The system offers you a translation for EVERY word (often the wrong one). Words with typos are not translated. You can take the result and edit it conventionally, but that defeats the purpose?"[167]

"Whether we are talking about machine or human translation, various considerations will tend to make the position worse for the smaller languages in the, in contrast to a very few large communities whose languages are widely used. The more limited the resources of a given language community within the network, the greater the proportional effort required to achieve integration into a common multilingual system if the language community in question must make this effort on its own. Calculated in relation to the intensity of demand for translations between each language pair, also, the less widely a language is used in absolute terms, the more expensive, in terms of cost/benefit, it will be to provide the full range of translation services, whether by human or mechanical means. On the other hand, in all but the simplest multilingual networks, there are probably good organizational, not to mention political reasons why it is desirable for each language community to maintain some autonomous capacity and control over integration of its own language into the multilingual network, rather than depending on a top-down structure where decisions affecting all languages are centralized, or else subordinated to decisions in language communities other than one's own."[168]

Furthermore, VIL bundles all machine translation systems into one system, which is less complex and takes up fewer resources. The MT systems are much more complex and take up more resources. For most you need a live connection to a powerful server somewhere. The downloadable ones are less powerful, work mostly on a word-by-word basis and only work for a few languages on a few domains.

---

[166] **Dresen, K.** *"George Mason University,"* MQP, 1999.

[167] **Klein, F.** "Machine Translation - A New Dimension and What You Can Expect," Technical Communications Forum, July 1998.

[168] **STOA Study**, *"Linguistic Diversity on the Internet: Assessment of the Contribution of Machine Translation,"* European Parliament STOA Study, http://www.europarl.eu.int/stoa/publi/99-12-01/default_en.htm.

6.3.2.10 What Was Learned from the Live Tests: Summary

With the live tests we have tried to demonstrate some hypotheses about VIL. We stated these hypotheses as part of the purposes for which we created the test.

6.3.2.10.1 Demonstrate learnability and retention

The following hypotheses have been confirmed for learnability:

- **Hypothesis A**: Subjects complete subsequent sentences of smaller or equal grammatical complexity faster than previous sentences.

  Results of the average times to complete the sentences have shown that subjects completed these subsequent sentences faster. For example, sentence 9 and 10, which did not introduce new grammatical features, were encoded faster than for instance sentence 1:

Nr	Sentence	Avg without teaching	Avg with teaching
1	I eat a hamburger.	5.6	1.9
9	He drinks milk with a glass.	1.7	1.4
10	You heated a sausage.	2.2	1.6

- **Hypothesis B**: Subjects make fewer mistakes on subsequent sentences of smaller or equal grammatical complexity.

  This hypothesis was directly related to hypothesis A. From studying all the attempts subjects made and the mistakes made we were able to conclude that the subjects made less mistakes on these subsequent sentences. This was argued by the fact that for the last two sentences, which introduced no new complexity, there were no mistakes made.

- **Hypothesis C**: Subjects select subsequent instances of the same item from the iconicon faster than initial instances of that same item

  Learnability was also shown by the test data showing that subsequent instances of the same item were indeed identified faster than initial instances.

Type	Item	Sentence	Avg		Type	Item	Sentence	Avg
Noun	I	1	11.7		Verb	Drink	2	8.9
Noun	I	2	3.0		Verb	Drink	9	7.6
Noun	I	3	3.8		Verb	Eat	1	19.8
Noun	I	4	3.9		Verb	Eat	6	11.3

Type	Item	Sentence	Avg	Type	Item	Sentence	Avg
Noun	I	5	3.0	Verb	Eat	8	10.9
Noun	I	6	3.6	Verb	Put	3	33.5
Noun	milk	2	23.0	Verb	Put	4	23.9
Noun	milk	9	12.0				

- **Hypothesis D**: The time difference between the first and second instance (for the same item) is much bigger than the time difference between the second and subsequent selections.

  Subjects learned the system pretty fast and data showed that the differences between the first and second instance was indeed bigger than the time difference between the second and subsequent selections. For example, the time for the verb "to eat" for the first sentence was 19.8 seconds whereas for sentence 6 and sentence 8 is was 11.3 seconds (for both).

- **Hypothesis E**: Subjects select an item that belongs to the same category as a previously selected item faster than the ones previously selected from that category.

  Data has shown that subsequent items from the same category are indeed identified faster than initial items. Once people knew where to look for an icon, they became more and more proficient in using the system. They learned to grasp the organization of the hierarchies. For example, people learned that a verb like "to boil" (a transformation verb) was likely to occur in the same category as "to bake", and consequently found it faster. Also, the verb "to drink" was selected faster than "to eat", since they belong to the same category "Ingest verbs" and "to drink" was selected after "to eat".

  Furthermore, results showed that subjects selected adjectives faster than nouns and verbs, and nouns faster than verbs.

The following hypotheses have been confirmed for retention over time:
- **Hypothesis F**: Performance in the retention test is better than performance in the initial test.

  Test subjects showed retention over time and all performed better than in the initial test. This was confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

$$t(18) = -2.34996$$

6.3.2.10.2 Determine whether a small teaching period increases performance

The following hypothesis has been confirmed:

- **Hypothesis G**: Performance of the group with teaching is better than the performance of the group without teaching

  From the data we have seen that a short teaching phase increases performance. When comparing results from two groups, one with teaching and one without teaching, results showed a difference in performance. This was confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

  $$t(18) = 4.234375$$

6.3.2.10.3 Determine whether VIL is encodable enough

The following hypothesis has been confirmed:

- **Hypothesis H**: It is not harder to encode sentences in VIL than to translate sentences into another language with a dictionary.

We needed to setup some measurement to find out whether VIL is encodable enough. We compared the time it took a person to translate the test sentences to a foreign language (Swedish) by using the dictionary to the average times of the test subjects who had received some teaching. Results from a t-test confirmed that there was no statistically significant difference between translating sentences with a dictionary and using VIL (inexperienced users):

$$t(18) = -0.65356, \text{ with } \alpha = 0.05$$

We then had an experienced VIL user encode the sentences in VIL. It turned out that these times were remarkably faster than translating with a dictionary. The t-test results when comparing dictionary lookup with the experienced user showed a big difference:

$$t(18) = 5.404784, \text{ with } \alpha = 0.05$$

One thing to mention about translating from English to Swedish is that the word orders in both languages are much alike, so that translating could practically be done on a word-by-word basis. The Swedish language also does not contain any difficult letters. Should one have translated to Chinese, then not only would the sentences have been grammatically more incorrect, but it would also have taken much longer to write down the Chinese characters. Furthermore, the translation was done into Swedish only. Should we have wanted to communicate these sentences to more languages then the times would of course be much higher.

6.3.2.10.4 Determine subjective satisfaction with the system

Subjective satisfaction was determined through a questionnaire among all test subjects. From the results we could conclude that people found the interface user-friendly and that the icons were designed well. Most people felt they got more confident on subsequent sentences, as they familiarized themselves with the system and learned from using it.

6.3.2.10.5 Learn what needs to be changed to the system

The live test revealed a couple of places where the system could possibly use some adjustment:

- The time-when (tense) case posed some problems. We should remove the "present" tense icon from the time-when case, since this case should not be filled when the tense in the sentence is "present" (the default tense).
- Also in the time-when case, we need to disable the "temporalizer", whenever the "time pronoun" is selected. A common mistake was that, whenever the tense was "past tense", subjects selected the "before" icon from the "temporalizer".
- There was some confusion about the difference between duration and frequency of occurrence. This could be explained better in the instructions.
- Some subjects did not understand how to create a questioned field. This could be explained better in the instructions.
- Some icons may need redesigning. For example, the noun "knife" looked odd to some people, causing them to backtrack and look for it somewhere else (eventually returning to the correct icon).
- *The noun or category is classified illogically.* During beta testing of VIL we noticed that testers had problems finding the "food" category as a child of the "humans" category. It did not sound logical to them that to look for a "hamburger" one had to select the "humans" category first. Consequently we moved the category up one level, such that it is on the same level as "humans" together with categories like "animals", "fruit", "vegetables", "plants", and "evolution".
- Some tooltips describing the icon added confusion. For example, the category name "silverware" may have incorrectly suggested this is about knifes, forks, etc. However, what was meant was "objects are used for eating and drinking". Consequently, the noun "glass" could have been hard to find for some subjects, since it is classified under "silverware". This is an error in the program that may have contributed (very slightly) to worse results.

6.3.2.11 Deep Structure versus Surface Structure

One characteristic of VIL is that it represents the deep structure of a sentence. In many languages, sentences can have different surface structures for the same deep structure. The encoding test as we have conducted it does not test different types of surface structure. It asks the subjects to encode sentences, while showing these textual sentences to the tests subjects in the same surface structure each time.

While this is a good means to test the encodability of VIL, in the future we would like also to test if subjects would have been able to map different surface structures to the same deep structure. For example, the next sentences with different surface structures have the same deep structure and should consequently be represented in the same way in VIL:

<div align="center">

I baked a cake for you.
I baked you a cake.
A cake was baked for you by me.
You were baked a cake by me.

</div>

Here the agent is always "I", the object is always "cake", and the goal is always "You".

A future test for VIL would be to have test subjects encode sentences in VIL that are offered to them in different surface structures. In a test like this, we would test whether test subjects encode these English sentences to identical VIL messages.

7. Conclusions and Future Work

This dissertation was concerned with creating a computer-based iconic communication language that could serve as an international interlingua. The language, called VIL, is based on the notion of simplified speech. Our research into universal languages and iconic communication, together with our program that implements an iconic language, had several contributions. Throughout the dissertation we have proposed a set of principles, some of which are relevant to any universal language, other are specific to visual or iconic languages. These principles could be used by others in the field as guidelines or at least could serve as a starting point for discussion and debate. Besides the research, we have designed a language and an internationally recognizable set of icons and implemented an iconic communication language, VIL.

Even though not every aspect of VIL is perfect yet, we have demonstrated the viability of an international visual interlingua. As we have seen in the evaluation chapter, the results of the icon and system test have shown that a visual interlingua like VIL is a very possible way to try to solve the problems caused by language barriers. Furthermore, a simple restricted grammar and limited vocabulary, together with self-explanatory icons and an easy user interface, make the language and system easy to use and easy to learn. The icons designed for VIL are generally recognized across cultures.

Section 7.1 summarizes the work done in the dissertation. Section 7.2 identifies some possible enhancements that could make the system better and more attractive, for example intelligent aids to help the user.

7.1 Summary of Accomplishments and Results

In chapter one and two we researched previous attempts that had been made to create a universal or international language. Most of these languages were artificial languages, but also some existing languages had been proposed as a universal language. We divided these languages into three groups: artificial languages (written, non-visual), non-artificial languages, and the visual languages, and described each in turn. We stated the shortcomings of the first two approaches and the previous approaches to visual languages. Artificial languages like Esperanto have met little success, largely due to the initial effort required to learn them (by a significant number of people). Concerns about cultural and economic hegemony argue against the use of any natural language.

We argued for a universal computer-based iconic language. We showed that visual languages overcome most of the problems that written languages encounter. Furthermore, with the technological developments of the last decade, such as powerful computers, graphical interfaces, and the World Wide Web, an excellent opportunity had been created for a computer-mediated visual interlingua to meet the need for communication between speakers of different languages. This iconic language should be designed to take advantage of the technology. People would be able to communicate with an iconic language without the need to draw the pictures themselves, since they could choose these pictures from the screen.

The problem with previous approaches to visual communication languages is that the languages are too difficult to learn. The languages are either based on too complex theories (like Schank's conceptual dependency theory) or the symbols and icons are not intuitive and self-explanatory (conventional abstract icons should be avoided as much as possible). VIL is designed with the emphasis on simplicity and ease of use. It allows for fast learning and flexibility. To help VIL attain this simplicity, many principles were adopted from research on pidgins and the basic languages like Basic English. We have also designed a grammar for VIL, which reflects this simplicity.

We studied Basic English and Pidgins in order to show that it is very possible to communicate using a restricted grammar and a limited vocabulary. Similar to pidgins, languages arising from the prolonged contact of people speaking two or more languages, VIL utilizes features that are in the 'greatest common denominator' of features in different languages. This allows its complexity to be significantly reduced; for example, it has no inflection, no number, gender, or tense markers, and no articles. VIL has no linear order. This is possible because it was designed as a visual language, in contrast to written languages which are the result of a transfer to visual modality of spoken language, which evolved in the context of auditory modality where sequencing and ordering is critical.

In chapter 3 we discussed the semantics of the language. Various principles for VIL were stated and VIL's grammar was described in detail. The classification of verbs, nouns, and adjectives were researched and the resulting hierarchies were given. VIL's grammar was based on the subcategorization of verbs presented in section 3.2.2.5. As in case grammar, a sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship. Although there can be compound instances of a single case (through noun phrase conjunction), each case relationship occurs only once in a simple sentence. The verbal elements of the sentence are the major source of the structure: the main verb in the proposition is the focus around which the other phrases, or cases, revolve.

The following lists all principles from chapter 3. For a description of them we refer to chapter 3:

Principle 3.1: *Greatest Common Denominator or Minimal Grammar and Vocabulary*

Principle 3.2: *Directive and Referential Function*

Principle 3.3: *No Linear Word Order*

Principle 3.4: *Number: no synthetic plurals*

Principle 3.5: *No gender distinctions*

Principle 3.6: *No inflection*

Principle 3.7: *Negation in invariant way*

Principle 3.8: *Verbs and tense*

Principle 3.9: *No auxiliary verbs*

Principle 3.10: *No article*

Principle 3.11: *Distinguishing Roles/Cases*

Principle 3.12: *Distinguishing the Major Categories*

Principle 3.13: *Derivational Morphology: Changing a Category*

Principle 3.14: *Modality: Imperative*

Principle 3.15: *Modality: Passive*

Principle 3.17: *Yes/No Questions*

Principle 3.18: *Wh-Questions*

The following principles apply to the classifications, i.e. the organization of nouns, adjectives and verbs:

Principle 3.19: *Depth of Hierarchies*

Principle 3.20: *Width of Hierarchies*

Principle 3.21: *Using items to predict subsequent items*

Principle 3.22: *Attributes: Adjectives as Features*

Principle 3.23: *Lattice in Hierarchies or Cross-Classification*

Principle 3.24: *Non-Predicative Adjectives*

Principle 3.25: *Contrary Adjectives*

Principle 3.26: *Contradictory Adjectives*

Principle 3.27: *Decomposition of Verbs*

The following seven categories are the top-level categories for the noun classification:
1. Physical World
2. Beliefs, Customs, and Society
3. Arts & Entertainment
4. Sports
5. Communication
6. Science & Technology
7. Transportation

The following three categories are the top-level categories for the adjective classification:

1. Lower Level Perceptual
   The lower level perceptual adjectives are primarily comprised of adjectives that involve our 5 senses: sight, hearing, touch, smell, and taste.
2. Higher Level Evaluative
   The higher level evaluative adjectives are adjectives that involve our ability to judge and evaluate situations and events.
3. Comparative/Relational
   The Comparative/Relational adjectives are adjectives that involve our ability to compare and relate things.

The following seven categories are the top-level categories for the adjective classification. Verbs were classified according to their shared meaning. The verbs are not only grouped according to their shared meaning, but also according to their case predictive features.

1. Mental
2. Alienable Possession
3. Physical Transfer and Location
4. Existence
5. Identificational
6. Involuntary Activity
7. Voluntary Activity or Verbs Involving the Body w/o Changing Location

In chapter four, we focused on the representation of the system. We discussed general principles in icon design and gave various examples of good and bad icons along with principles to improve icon design. After that we gave icon design principles for VIL, and stated rules on how to combine visual elements to produce the icons and grammatical entities that we need.

Principle 4.1: *Iconic vs. Visual Representation in VIL*
Principle 4.2: *Visual elements as case distinguishers*
Principle 4.3: *Visual elements to distinguish instantiated icons from uninstantiated icons*
Principle 4.4: *Visual elements as category distinguishers*
Principle 4.5: *Visual elements to represent derivational morphology*
Principle 4.6: *Visual elements to indicate degree of gradable concepts*
Principle 4.7: *Visual elements for negation*
Principle 4.8: *Visual elements for direction, state and choice*
Principle 4.8.1: *Visual elements for direction/motion*
Principle 4.8.2: *Visual elements for change of state*
Principle 4.8.3: *Visual elements for choice*

Principle 4.9:   *Visual elements to indicate a selected icon*
Principle 4.10: *Visual elements to indicate modification*
Principle 4.11: *Visual elements to allow the user to find where an item occurs in the iconicon*

Icon representation principles for verbs

Principle 4.12: *Showing the tool or activity to represent verbs of creation and destruction*
Principle 4.13: *Showing before and after states to represent verbs of change of state*
Principle 4.14: *Showing concrete action to represent manipulation verbs*
Principle 4.15: *Showing arrows to represent verbs of motion or conceptual movement*
Principle 4.16: *Showing the activity itself to represent verbs of activity*

Icons Representation principles for Nouns

Principle 4.17: *Showing the object itself to represent concrete nouns*
Principle 4.18: *Showing an associated concrete object to represent abstract concepts*
Principle 4.19: *Using symbols to represent abstract concepts*
Principle 4.20: *Showing several children of a category to represent noun categories*

Designing Icons Representing Noun Modifiers or Adjectives

Principle 4.21: *Showing a slider in combination with two antonyms to represent gradable adjectives*
Principle 4.22: *Showing one of our senses to represent adjective categories*

Then, in section 4.4, we discussed the domain of cooking, eating and drinking, which is the domain that has been implemented for the first release of VIL. In order to show the viability of a computer-based iconic communication language such as VIL, we decided to implement one particular domain. Our initial goal was not to implement a whole system, but just to show that it is feasible to communicate with icons. Another reason for a limited domain was to allow testing and evaluation. The domain we had chosen to implement was that of cooking, eating and drinking. We showed all icons that had been designed for this domain. Chapter 5 explained in detail the computer program that was designed for VIL, along with the sample sentence "I put (past tense) a porkchop in a frying pan."
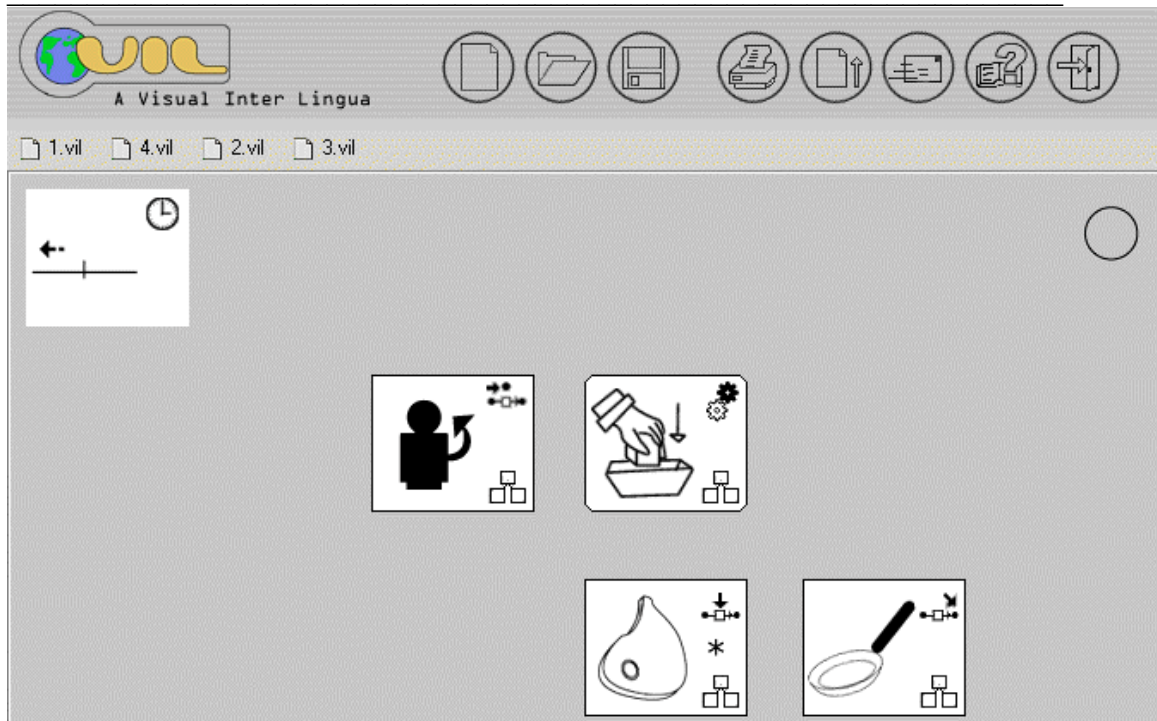
7.1.1 Evaluating VIL

Chapter six contained an evaluation of the system. It described three major tests (icon test, decodability test, and encodability test) to evaluate the following principles:
- Learnability
- Encodability
- Decodability

We argued that the system was **easy to learn** because:
- VIL uses a drastically reduced grammar and vocabulary
- Icons in VIL are easy to recognize
- VIL uses an easy interface
- VIL has numerous help functions that help explaining the meaning of an icon, like animation, category help, and tooltips

We argued that it was fairly **easy to generate messages** with the language and its delivery system because:
- For a visual or iconic language messages are generated faster and with greater ease if the language is computer-based, since the user is not required to draw icons by hand. The icons can simply be chosen from the screen.

- The simple grammar, the design of the hierarchies for our grammatical categories, and the design of the program itself were all designed to facilitate encoding.

We argued that messages composed with the system were **easy to comprehend** because:
- The simple grammar helped in messages being easier to comprehend
- VIL's self-explanatory icons and consistent visual combining features facilitate decoding of messages.
- VIL supports animated explanation of verbs
- VIL supports category help: showing where an icon stands in the hierarchy and showing its siblings.
- VIL supports explanation of the different visual combining features of an icon by tooltips.

Three major tests were used to determine learnability, encodability, and decodability:
- Icon test
- Decodability test
- Encodability test

For each test we stated several purposes and hypotheses. These hypotheses were then confirmed or rejected by the test results. The first two tests were web-based tests and the encodability test was a live test.

7.1.1.1 The Icon Test

For the icon test we set up a survey for testing the icons. The icon test served the following purposes:
1. Determine how recognizable the icons are
2. Determine whether there are cross-cultural differences in recognition of icons

   Hypothesis: There is no difference in recognition of icons for people of western and non-western culture.

3. Find out which icons need to be redesigned

Determine how recognizable the icons are
From the results of the test we have seen that the icons were recognized very well, with about 79% correct answers for all questions.

Summarizing, the following observations were made:
- Category icons were generally harder to recognize than terminal icons.

- Adjectives were recognized better than verbs.
- Noun category icons are recognized better than noun terminal icons.

Determine whether there are cross-cultural differences in recognition of icons
The following hypothesis was made:

- **Hypothesis**: There is no difference in recognition of icons for people of western and non-western culture.

The hypothesis was confirmed. We saw that the results were close together for all types of questions. This meant that the icons for terminals are generally recognized equally well across cultures (no statistically significant differences). These findings were confirmed by statistically results from t-tests with $\alpha = 0.05$ (95%):

Type of question	Category vs. terminals	t-test result
Image-meaning	Category	t(16) = 0.789793
Image-meaning	Terminal	t(8) = 1.502955
Open ended	Category	t(8) = 0.863064
Open ended	Terminal	t(42) = -0,30565

Type of question	Category vs. terminals	t-test result
Verb	Category	t(8) = 0.315254
Verb	Terminal	t(10) = -0.10464
Noun	Category	t(12) = -0.610185
Noun	Terminal	t(28) = -1.504452
Adjective	Category	t(2) = 1.006349
Adjective	Terminal	t(10) = -0.60311

Find out which icons need to be redesigned
We determined a cutoff point for when an icon needs to be redesigned. We actually compared the number of incorrect answers for an icon against the number of correct answers for that icon. By carefully studying all the errors we believed that a cutoff percentage of 20% was reasonable. We identified a number of icons to be redesigned.

7.1.1.2 The decodability Test

The web test for decodability served the following purposes:
1. To determine whether VIL is decodable enough
2. To determine whether tooltips increase performance

Hypothesis: performance for the test with tooltip help is better than performance for the test without tooltip help

3.   To learn what needs to be changed to the system

Determine whether VIL is decodable enough
It was hard to state a hypothesis for this. If we consider partial correctness we saw a correctness percentage of around 72%. This was good performance considering the fact that people had never seen the program before. If we had adjusted the instructions and if people really had read the instructions, the results would have been better. The only hypothesis we could then make was that performance of subjects who read the instructions was better than performance of subjects who had not read the instructions. This hypothesis was confirmed when we accounted most of the partial correctness errors to the fact that people had not read the instructions. From the errors made in decoding, we could conclude that a lot of them had not read the instructions.

Determine whether tooltips increase performance
The following hypotheses was made:
- **Hypothesis**: Performance for the test with tooltip help is better than performance for the test without tooltip help.

  Even though we saw a difference in the level of correctness, results from the t-test indicated that there was no statistically significant difference between the test with tooltip help and the test without tooltip help. Consequently, we had to reject the hypothesis:
  $$t(10) = 0.077395, \text{ with } \alpha = 0.05$$

Learn what needs to be changed to the system
The decodability test revealed a couple of places where the system could possibly use some adjustment:
- In order to eliminate the issues mentioned in the section on partial correctness, we need to clearly address these issues in the instructions.
- Clear instructions increase performance (this is concluded by talking to acquaintances, who took the test).
- The asterisk as a sole marker for modifying properties may have proven to be insufficient. We possibly need an additional marker or color or some other way to have the modified icons stand out more.
- We have learned which program icons need some redesigning, e.g. the "duration" case was misinterpreted sometimes because of the icon.

7.1.1.3 The Encoding Test

The live tests served the following purposes:
1. To demonstrate learnability and retention.
2. To determine whether a small teaching period increases performance.
3. To determine whether VIL is encodable enough.
4. To determine subjective satisfaction with the system.
5. To learn what needs to be changed to the system.

For the purpose of demonstrating learnability and retention we wanted to confirm the following hypotheses:
A. Subjects complete subsequent sentences of **smaller or equal grammatical complexity** faster than previous sentences.
B. Subjects make **less mistakes** on subsequent sentences of smaller or equal grammatical complexity.
C. Subjects select subsequent instances of the **same item** from the iconicon faster than initial instances of that same item
D. Consequently, the time difference between the first and second instance (for the same item) is much bigger than the time difference between the second and subsequent selections.
E. Subjects select an item that belongs to the **same category** as a previously selected item faster than the ones previously selected from that category. For example "to drink" is selected faster than "to eat", since they belong to the same category "Ingest verbs" and "to drink" is selected after "to eat".
F. Performance in the retention test is better than performance in the initial test.

For the purpose of determining whether a small teaching period increases performance we wanted to confirm the following hypothesis:
G. Performance of the group with teaching is better than the performance of the group without teaching

For the purpose of determining whether VIL is encodable enough we wanted to confirm the following hypothesis:
H. It is not harder to encode sentences in VIL than to translate sentences into another language with a dictionary.

Demonstrate learnability and retention
The following hypotheses were confirmed for learnability:
*   **Hypothesis A**: Subjects complete subsequent sentences of smaller or equal grammatical complexity faster than previous sentences.

Results of the average times to complete the sentences have shown that subjects completed these subsequent sentences faster. For example, sentence 9 and 10, which did not introduce new grammatical features, were encoded faster than for instance sentence 1:

Nr	Sentence	Avg without teaching	Avg with teaching
1	I eat a hamburger.	5.6	1.9
9	He drinks milk with a glass.	1.7	1.4
10	You heated a sausage.	2.2	1.6

- **Hypothesis B**: Subjects make fewer mistakes on subsequent sentences of smaller or equal grammatical complexity.

  This hypothesis was directly related to hypothesis A. From studying all the attempts subjects made and the mistakes made we were able to conclude that the subjects made less mistakes on these subsequent sentences. This was argued by the fact that for the last two sentences, which introduced no new complexity, there were no mistakes made.

- **Hypothesis C**: Subjects select subsequent instances of the same item from the iconicon faster than initial instances of that same item

  Learnability was also shown by the test data showing that subsequent instances of the same item were indeed identified faster than initial instances.

Type	Item	Sentence	Avg
Noun	I	1	11.7
Noun	I	2	3.0
Noun	I	3	3.8
Noun	I	4	3.9
Noun	I	5	3.0
Noun	I	6	3.6
Noun	milk	2	23.0
Noun	milk	9	12.0

Type	Item	Sentence	Avg
Verb	Drink	2	8.9
Verb	Drink	9	7.6
Verb	Eat	1	19.8
Verb	Eat	6	11.3
Verb	Eat	8	10.9
Verb	Put	3	33.5
Verb	Put	4	23.9

- **Hypothesis D**: The time difference between the first and second instance (for the same item) is much bigger than the time difference between the second and subsequent selections.

Subjects learned the system reasonably fast and data showed that the differences between the first and second instance was indeed bigger than the time difference between the second and subsequent selections. For example, the time for the verb "to eat" for the first sentence was 19.8 seconds whereas for sentence 6 and sentence 8 is was 11.3 seconds (for both).

- **Hypothesis E**: Subjects select an item that belongs to the same category as a previously selected item faster than the ones previously selected from that category.

  Data has shown that subsequent items from the same category are indeed identified faster than initial items. Once people knew where to look for an icon, they became more and more proficient in using the system. They learned to grasp the organization of the hierarchies. For example, people learned that a verb like "to boil" (a transformation verb) was likely to occur in the same category as "to bake", and consequently found it faster. Also, the verb "to drink" was selected faster than "to eat", since they belong to the same category "Ingest verbs" and "to drink" was selected after "to eat".

  Furthermore, results showed that subjects selected adjectives faster than nouns and verbs, and nouns faster than verbs.

The following hypotheses was confirmed for retention over time:
- **Hypothesis F**: Performance in the retention test is better than performance in the initial test.

  Test subjects showed retention over time and all performed better than in the initial test. This was confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

$$t(18) = -2.34996$$

Determine whether a small teaching period increases performance
The following hypothesis was confirmed:
- **Hypothesis G**: Performance of the group with teaching is better than the performance of the group without teaching

  From the data we have seen that a short teaching phase increases performance. When comparing results from two groups, one with teaching and one without teaching, results showed a difference in performance. This was confirmed by statistically significant results from t-tests with $\alpha = 0.05$ (95%):

$$t(18) = 4.234375$$

Determine whether VIL is encodable enough

The following hypothesis was confirmed:

- **Hypothesis H**: It is not harder to encode sentences in VIL than to translate sentences into another language with a dictionary.

We needed to setup some measurement to find out whether VIL is encodable enough. We compared the time it took a person to translate the test sentences to a foreign language (Swedish) by using the dictionary to the average times of the test subjects who had received some teaching. Results from a t-test confirmed that there was no statistically significant difference between translating sentences with a dictionary and using VIL (unexperienced users):

$$t(18) = -0.65356, \text{ with } \alpha = 0.05$$

We then had an experienced VIL user encode the sentences in VIL. It turned out that these times were remarkably faster than translating with a dictionary. The t-test results when comparing dictionary lookup with the experienced user showed a big difference:

$$t(18) = 5.404784, \text{ with } \alpha = 0.05$$

Determine subjective satisfaction with the system

Subjective satisfaction was determined through a questionnaire among all test subjects. From the results we could conclude that people found the interface user-friendly and that the icons were designed well. Most people felt they got more confident on subsequent sentences, as they familiarized themselves with the system and learned from using it.

Learn what needs to be changed to the system

The live test revealed a couple of places where the system could possibly use some adjustment:

- The time-when (tense) case posed some problems. We had to remove the "present" tense icon from the time-when case, since this case should not be filled when the tense in the sentence is "present" (the default tense).
- Also in the time-when case, we had to disable the "temporalizer", whenever the "time pronoun" was selected. A common mistake was that, whenever the tense was "past tense", subjects selected the "before" icon from the "temporalizer".
- There was some confusion about the difference between duration and frequency of occurrence. This could have been be explained better in the instructions.
- Some subjects did not understand how to create a questioned field. This could have been be explained better in the instructions.
- Some icons needed redesigning. For example, the noun "knife" looked odd to some people, causing them to backtrack and look for it somewhere else (eventually returning to the correct icon).

- Some tooltips describing the icon added confusion. For example, the category name "silverware" may have incorrectly suggested this is about knifes, forks, etc. However, what was meant was "objects are used for eating and drinking". Consequently, the noun "glass" could have been hard to find for some subjects, since it is classified under "silverware". This was an error in the program that may have contributed to worse results.

The live test showed that it is very possible to encode sentences in VIL with little to no instruction. Users need a small learning period for themselves to get acquainted with the system, the language, and the organization of the iconicon. However, test results have shown that people learn this reasonably fast and that performance increases as one gets more familiar with the system. Later sentences that were grammatical equal in complexity as earlier questions were completed faster. Furthermore, the retention over time test also shows that people learn and remember how to use the system.

Previous work on iconic communication and languages created (see chapter 2) offered interesting proof of concept of our ideas, but did not have the potential to be scaled up to commercial applications. They were either too symbolic (e.g. Blissymbolics or Elephants memory) or were too restricted by theory (e.g. CD-Icon, based on conceptual dependency theory). Nevertheless, examples of the use of icons as replacement of text have begun to appear on the market on restricted domains, for example *The Wordless Travel Book* by Meader.[169] VIL, mainly because of its ease of use, also has the potential to be scaled up to a commercial system if more domains would be implemented. A web-based version of the program looks even more promising, because then people can communicate with each other over the Internet.

The next section identifies some possible enhancements that could make the system better and more attractive.

7.2 Identify Possible Enhancements and Future Work

In this last section we identify some areas in which the language and the system could be extended. The evaluation from chapter 6 indicated areas in which work needs to be done, e.g. redesigning of some icons. The following is a list of possible expansion of and enhancements to the system:

---

[169] **Meader, J.** *"The Wordless Travel Book,"* Ten Speed Press, 1995.

1. *Implement more domains:* Now that we have implemented the domain of cooking, eating and drinking, the next step is to implement new domains. A domain like traveling is a good one to start with.

2. *Implement all functionality of the language:* With the first version of VIL, it is, for instance, only possible to compose simple sentences. In the next version of VIL, it should be possible to compose complex sentences using subordinators (e.g. because/although) and coordinators (and/or), e.g. "Before I went to the market, I visited by my mother" or "I cannot come, because I have to go to the dentist".

3. *Create an Internet version of VIL:* As stated before, an Internet version of VIL could really add more value to the usability of the language, and is currently under construction.

4. *Translating tooltips into more languages:* The current version of VIL only supports tooltip help in English, Dutch, and Spanish. Although this covers a lot of countries, at least several other major languages need to be added in order to justify language specific tooltip help. Languages like German, French, Portuguese, Chinese, Arabic, and Persian need to be there.

5. *Adding intelligent capabilities:* The system could have precompiled knowledge. It could fill in cases with default icons. It could anticipate likely choices of the user and take him/her to specific pages of the iconicon. For example, with location, it is not likely that people do cookouts in the office or plant seeds there either. Vice versa, you do not generally copy or fax from your garden. So, depending on the background, the user is taken to a specific page in the iconicon. As another example, if the user picks the verb "to run", then the system could assume that the subject is animate.

6. *User specific system:* The system could be designed to keep more track of a particular user's history of using the system, and according to this come up with defaults to fill cases (specific for this particular user).

7. *Emphasis and Showing Parts:* It will be really helpful to build in a possibility to emphasize part of an icon. This can be done by highlighting part of an icon to direct the user's attention, to eliminate unnecessary ambiguity, or to stress the content of the message itself. Examples could be "the tail of a dog" or a sentence like "**I** baked the pie" (emphasizing I). Parts of an icon could also be referred to by using an arrow.

8. *Adding earcons as help:* earcons are sounds that can help explain the meaning of an icon, like barking for a dog or the sound of boiling for the verb to boil.

8. Bibliography

AIGA, *"Symbol Signs,"* The American Institute of Graphic Arts, 1993.

Aiken, J. R. *"Little English,"* Journal of Adult Education, 1:11-14, June 1936.

Andrews, A.D. *"Lexical Structure,"* in Linguistics: the Cambridge Survey, Vol1 Linguistic Theory: Foundations, F.J. Newmeyer, Cambridge University, 1988, pp. 60-88.

Ballmer, Th. and **Brennenstuhl, W.** "Speech Act Classification: A study in the Lexical Analysis of English Speech Activity Verbs," Ed. W. Levelt, Spring-Verlag, 1981.

Barker, P. *"Human Communication Processes,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

Barker, P. and **Schaik, P. van** *"Designing and Evaluating Icons,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

Bauer, E. *"Machine Translation - 2001 has already arrived,"* Technical Communications Forum, December 2000.

Beardon, C. *"CD-Icon: an iconic language based on conceptual dependency,"* Intelligent Tutoring Media, Vol. 3, No. 4, November, 1992.

Beardon, C., Dormann, C., Mealing, S. and **Yazdani, M.** *"Talking with pictures: exploring the possibilities of iconic communication,"* Association for Learning Technology Journal, 1, 1, pp. 26-32, 1993.

Beardon, C. *"IconText: An exploration of the Limitations of Iconic Languages,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

Beardon, C. *"Iconic Communication,"* University of Brighton Grand Parade, http://www.intellectbooks.com/iconic/iconcomm/iconcomm.htm, 1993.

Berlin, B. and **Kay, P.** *"Basic Color Terms,"* 1960s.

Bertin, J. *"Semiology of Graphics,"* translated by P. Berg, Madison: University of Wisconsin Press, 1983.

Bliss, C. K. *"Semantography,"* Semantography publications, Australia, 1965.

Bohne, S., Bunar, M., Gardella, S. and Hynes, P. *"Cross-Cultural Icon Test"* An Interactive Qualifying Project Report, WPI, May, 1996.

Brown, R. W. and Bellugi, U. *"Three processes in the child's acquisition of syntax."* Harvard Educ. Rev. 34, pp. 133-151 (Repr. in "New directions in the study of language," Lenneberg, E. (ed.), Cambridge, Mass., 1964).

Cazden, C. B. *"The Acquisition of Noun and Verb Inflections,"* in Ferguson, C.A. and Slobin, D. I. (eds), Studies of Child Language Development, Publ: Holt, Rinehart and Winston, 1973, pp. 226-240.

Cliff, N. *"Adverbs as Multipliers,"* Psychological Review, 66, pp. 27-44, 1959.

Corbeil, J. *"Visual Dictionary,"* ISBN: 0-8160-1542-9, 1986.

Corder, S. P. *"Simple codes and the source of the second language learner's initial heuristic hypothesis."* Paper presented at the Colloque 'Theoretical Models in Applied Linguistics' IV, Universite de Neuchatel, 1975.

D'Andrade, R. *"The development of Cognitive Anthropology,"* chapter 5: Folk Taxonomies, Cambridge University Press, ISBN: 0-521-45976-1, 1995.

Dorling Kindersley, *"Visual Encyclopedia,"* ISBN: 1-56458-985-4, 1995.

Dormann, C. *"Self-Explaining Icons,"* University of Brighton Grand Parade, http://www.intellectbooks.com/iconic/self/self.htm, 1993.

Dresen, K. *"Is There a Universal Translator in YOUR Future?"* George Mason University MQP, 1999.

Dutton, T. *"Police Motu: iena sivarai (its story)."* Port Moresby: University of Papua New Guinea Press, 1985.

Elsevier, Grote Winkler Prins Encyclopedia, Elsevier Publishers, Part 8, pp. 143-145, Amsterdam, 1984.

STOA Study, *"Linguistic Diversity on the Internet: Assessment of the Contribution of Machine Translation,"* European Parliament STOA Study, http://www.europarl.eu.int/stoa/publi/99-12-01/default_en.htm.

Feierbaugh, W. C. *"Computer Graphics: Tools for Visualization,"* Publisher: William C. Brown, 1993.

Ferguson, C. A. *"Absence of Copula and the Notion of Simplicity: A study of Normal Speech, Baby Talk, Foreigner Talk, and Pidgins,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 277-292.

Ferguson, C. A. *"Baby Talk in Six Languages,"* in Language Structure and Language Use, Stanford University Press, 1971, pp. 113-133.

Ferguson, C. A. *"Simplified registers and linguistic theory,"* in L. Obler & L. Wise (eds.) Exceptional language and linguistics, New York: Academic Press, 1982.

Foley, W. A. *"Language birth: the processes of pidginization and creolization,"* Chapter 9 in Linguistics: The Cambridge Survey, Volume IV, Language: The Socio-cultural Context, Edited by Frederick J. Newmeyer, Cambridge University Press, Cambridge, 1988.

Frutiger, A. *"Signs & Symbols: Their Design and Meaning,"* New York: Van Nostrand Reinholt, 1989.

Gittins, D. *"Icon-based human-computer interaction,"* International Journal for Man-Machine Studies 24, 1986, pp. 519-543.

Gode, A. and **Blair, H. E.** *"Interlingua: a grammar of the international language,"* Storm Publishers, New York, 1955.

Hall, R. *"Pidgin and Creole Languages,"* Cornell University Press, 1966.

Head, W. D. *"Needed: A Spare Tire Language,"* Rotarian, 50:28-30, March 1937.

Honeywill, P. *"Visual Language for the World Wide Web,"* Intellect, Exeter, UK, 1999.

Horton, W. *"The Icon Book: Visual Symbols for Computer Systems and Documentation,"* John Wiley & Sons, ISBN: 0-471-59900-X, 1994.

Housz, T.I. "The Elephant's Memory,"
http://www.khm.de/~timot/PageElephant.html ,1994.

Huggins, W. H. and **Entwisle, D. R.** *"Iconic Communication: an annotated biography,"* The John Hopkins University Press, 1974.

Hymes, D. Commentary. In A. Valdman and A. Highfiled (eds.) *"Theoretical orientations in language studies."* New York: Academic Press, 1980.

International Encyclopedia of Linguistics, William Bright (ed.), Oxford University Press, 1992.

Jackendoff, R. *"Nonspatial Semantic Fields and the Thematic Relations Hypothesis," in* Semantics and Cognition, Chapter 10, The MIT Press, 1983.

Johnson, J. E. *"Excerpts,"* from Basic English, The Reference Shelf. Vol. 17, No. 1, The H. W. Wilson Company, N. Y., 1944, p. 58-68.

Justeson, J. and **Katz, S.** *"Co-occurrences of antonymous adjectives and their contexts,"* Yorktown Heights, NY: IBM Research Report, 1989.

Kamman, W. F. *"The Problem of a Universal Language,"* Modern Language Journal, 26:177-182, March 1942.

Kiger, J. *"The depth/breath trade-off in the design of menu-driven user interfaces,"* International Journal of Man-Machine Studies 20, 1984, pp. 201-213.

King, A. J. *"On the possibility and impossibility of a universal Iconic Communication System,"* in Iconic Communication, Eds. M. Yazdani and P. Barker, Intellect, 2000.

Klein, F. *"Machine Translation - A New Dimension and What You Can Expect,"* Technical Communications Forum, July 1998.

Kummel, P. *"Formalization of Natural Language,"* Springer-Verlag, 1979, Berlin.

Levi, J.N. *"The Syntax and Semantics of Complex Nominals,"* New York: Academic Press, 1978.

Levin, B. *"English Verb Classes and Alternations: A Preliminary Investigation,"* The University of Chicago Press, 1993.

Liebelt, L., McDonald, J., and **Stone, J.** *"Searching for items in menus: the effects of organization and type of target,"* Proc. Human Factors Society, Twenty-Seventh Annual Meeting, 1983, pp. 834-837.

MacMillian, *"The MacMillian Visual Desk Reference,"* MacMillian Publishing Group, ISBN: 0-02-531310-X, 1993.

Marcus, A. *"Graphic Design for Electronic Documents and User Interfaces,"* ACM Press, New York, 1992.

Matsukawa, T. and **Yokota, E.** *"Development of the Concept Dictionary -- Implementation of Lexical Knowledge,"* in Pustejovsky, J. and Bergler, S. (Eds.) "Lexical Semantics and Knowledge Representation," Proc. of the First SIGLEX Workshop, Berkeley, CA, June 17, 1991, Springer-Verlag, Chapter 21.

McCawley, J.D. *"Lexical insertion in a transformational grammar without deep structure,"* in CLS 4, 1968 reprinted in McCawley *"Grammar and Meaning,"* 1973.

McCloskey, J. *"Syntactic Theory,"* in Linguistics: the Cambridge Survey, Vol1 Linguistic Theory: Foundations, F.J. Newmeyer, Cambridge University, 1988, pp. 18-59.

Meader, J. *"The Wordless Travel Book,"* Ten Speed Press, 1995.

Mealing, S and **Yazdani, M.** *"A Computer-based Iconic Language,"* Intelligent Tutoring Media, 1(3), pp. 133-136, 1992.

Mealing, S and **Yazdani, M.** *"Communicating Through Pictures,"* http://www.media.uwe.ac.uk/masoud/author/icon-3/siena.htm, Department of Computer Science, University of Exeter, England, 1992.

Mealing, S. *"Do you see what I'm Saying,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

Mencken, H. L. *"The American Language,"* Copyright, Alfred A. Knopf, New York, 1936, p. 603-607.

Meunier, J. *"The Categorical Structure of Iconic Languages,"* Université du Quebec a Montré, http://pluton.lanci.uqam.ca/membres/meunier/iconlang/iconlag.htm.

Miller, G.A. and **Fellbaum, C.** *"Semantic Networks of English,"* in Levin and Pinker (Eds.), Lexical & Conceptual Semantics, Blackwell Publishers Ltd, ISBN: 1-55786-354-7, 1995.

Morris, A. V. *"A Language Cut to Fit the World,"* Rotarian, 53:36-39, November 1938.

Mullet, K. and **Sano, D.** *"Designing Visual Interfaces,"* SunSoft Press, 1995.

Murray, J. *"IBM, Reaching the global community,"* June 2000.

Nadin, M. *"Interface Design and Evaluation-Semiotic Implications,"* in Hartson, R. and Hix, D. (eds.), Advances in Human-Computer Interaction, Volume 2, 1989, pp. 45-100.

Neurath, O. *"International Picture Language,"* Department of Typography and Graphic Communication, University of Reading, England, 1978.

Ogden, C. K. *"Basic English: A General Introduction with rules and grammar,"* London: Kegan Paul, Trench, Trubner & Co. Ltd., Broadway House, Carter Lanes E.C., 1930.

Ogden, C. K. *"Basic for Science,"* London: Kegan Paul, Trench, Trubner & Co. Ltd., Broadway House, Carter Lanes E.C., 1942, p. 514.

Peirce, C. *"Collected Papers,"* Cambridge, MA: Harvard University Press, 1931 (excerpted in Buckler, Justus, ed., Philosophical Writings of Peirce, New York: Dover, 1955).

Richards, I. A. *"Basic English and Its Applications,"* Royal Society of Arts. Journal (London), 87:735-755, June 2, 1939.

Richards, I. A. *"Basic English and Its Uses,"* W. W. Norton & Company Inc., New York, 1943.

Rosch, E., Mervis, C.B., Gray, W., Johnson, D. and **Boyes-Braem, P.** *"Basic Objects in Natural Categories,"* Cognitive Psychology, 9, pp. 382-439, 1976.

Rogers, Y. *"Icons at the Interface,"* Interacting with Computers, 1(1), pp. 105-117, 1989.

Roscoe, J. *"The limits Of Iconic Communication,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.

Sassoon, R. and **Gaur, A.** *"Signs, symbols, and Icons: Pre-History to the Computer Age,"* Intellect, Exeter, UK, 1997.

Saussure, F. de *"Course in General Linguistics,"* trans. Wade Baskin, London, 1960.

Schank, R. *"Computer Models of Thought and Language,"* W. H. Freeman, San Francisco, 1973.

Schank, R. *"Conceptual Information Processing,"* North-Holland, New York, 1975.

Schumann, J. *"The Pidginization Process: A Model for Second Language Acquisition,"* Newbury House Publishers, 1978.

Shneiderman, B. *"Designing the User Interface: Strategies for Effective Human-Computer Interaction,"* Addison-Wesley, 1992.

Smith, D. M. *"Pidginization and language socialization: the role of marking,"* Unpublished paper, Georgetown University, 1973.

Smith, E.E. *"Theories of Semantic Memory,"* in W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes (vol. 5), Hillsdale, NJ: Erlbaum, 1978.

The National Commission on the Enrichment of Adult Life, *"A General View of the System of Basic English,"* Committee Report on Basic English for Adult Education, Washington, D. C. mineo p. 2-19.

The Orthological Committee, *"Notes on Basic English,"* Number One, A short Account of the System, October 1940.

Todd, L. *"Pidgins and Creoles,"* Language and Society Series, Publ: Routledge & Kegan Paul Ltd, 1974.

Tomkins, W. *"Indian Sign Language,"* Dover Publishers, ISBN 048622029X, 1969.

Tversky, B. and **Hemenway, K.** *"Objects, Parts, and Categories,"* Journal of Experimental Psychology: General, 113, pp. 169-193, 1984.

University of Maryland, *"Research into automatic dictionary construction in machine translation,"* http://www.umiacs.umd.edu/labs/CLIP/mt.html.

Volk, M. *"The Automatic Translation of Idioms. Machine Translation vs. Translation Memory Systems,"* in Nico Weber (ed.): Machine Translation: Theory, Applications,

and Evaluation. An assessment of the state of the art. St. Augustin: gardez-Verlag. 1998.

Wallace, D., Anderson, N., and Shneiderman, B. *"Time stress effects on two menu selection systems,"* Proc. Human Factors Society, Thirty-First Annual Meeting, 1987, pp. 727-731.

Walsh, C. *"The Verb System in Basic English,"* American Speech, 8:137-143, December 1933.

West, **M. P., Swenson, E. et al.** *"Vocabulary Limitation,"* Journal of Education, London, 71:582-584, September 1939.

West, **M. P., Swenson, E. et al.** *"A Critical Examination of Basic English,"* Toronto 1934.

Winston, M.E., Chaffin, R. and **Herrmann, D.** *"A Taxonomy of Part-Whole Relations,"* Cognitive Science, 11, pp. 417-444, 1987.

Yazdani, M. and **Goring, D.** *"Iconic Communication,"* Internal Report, Department of Computer Science, University of Exeter, England, 1990.

Yazdani, M. *"Communication through icons,"* in Iconic Communication, eds. M. Yazdani and P. Barker, Intellect, 2000.