The Role of Ontologies in a Linguistic Knowledge Acquisition Task

Steve Helmreich and Ron Zacharski
New Mexico State University

1. Introduction

This paper discusses the role of ontologies in a knowledge elicitation component of a natural language processing system. The system is intended to assist in the rapid development and deployment of a machine translation system from any so–called 'low–density' language (one lacking significant machine–tractable resources) into English. The elicitation component, called BOAS, is intended to guide non–expert informants to provide linguistic information in sufficient detail to automatically generate a machine translation system.[1] It elicits a variety of lexical, morphological, syntactic, and ecological data (letters, symbols, punctuation, etc.). The system also includes a module for converting this elicited information into a format useful for machine translation (e.g., converting paradigm tables into morphological rules, which are realized in a finite state morphological analyzer).

The linguistic challenge for the developers of Boas can be summarized as follows: how does one gather all the necessary information about all the phenomena that can occur in any natural language in a way that is both understandable to a non–expert informant and machine tractable without post–elicitation human intervention? How well this challenge is met is crucially dependent on the underlying ontology of the system.

After presenting a brief overview of Boas (§2), we describe the ontology of linguistic elements that is implicit in the Boas architecture and compare this ontology with the GOLD ontology (§3). In §4 we describe how the higher level linguistic ontology maps to acquisition modules. Finally, we describe some problems with this approach as well as our solutions to these problems (§5).

2. Description of Boas

Boas was designed to be used by informants who were not expected to be linguists much less, computational linguists. As such, all elicitation, training, help materials, and processing components had to be part of the Boas system. All this knowledge needed to be resident in the system with no opportunity to—at some later point–tailor the system to a particular language. The Boas system, in essence, needed to be a field linguist in a box. However, while a field linguist can

[1] We would like to acknowledge the architects and developers of the Boas system including Sergei Nirenburg and Marjorie McShane. See Nirenburg 1998, Nirenburg and Raskin 1998, and McShane et al. 2000 for more information. The project was supported by Department of Defense Contract MDA904-92-C-5189.

describe a language using any descriptive means possible, Boas needed to ground each description in a pre–determined set of concepts in order to automatically generate a machine translation system. Also, while field linguists often concentrate on language phenomena that are linguistically interesting, Boas concentrates on the most basic, the most processable, phenomena.

Boas is organized around a set of universal and non-universal parameters, their value sets, and possible realizations of the latter in the realms of ecology[2], morphology, and syntax (see, for example, Nirenburg 1998). The notion of parameters is a broad one and we use it here in its sense related to natural language processing. The notion of parameters in theoretical linguistics (e.g., Chomsky and colleagues' principles and parameters)—in particular, the inventory of parameters—is too limited for realistic large–scale applications.

The Boas interface is organized around two types of dynamically generated web pages. Navigational pages enable the informant to decide on what task to work on. An example of a navigation page, the one for paradigmatic morphology, is shown in Figure 1.

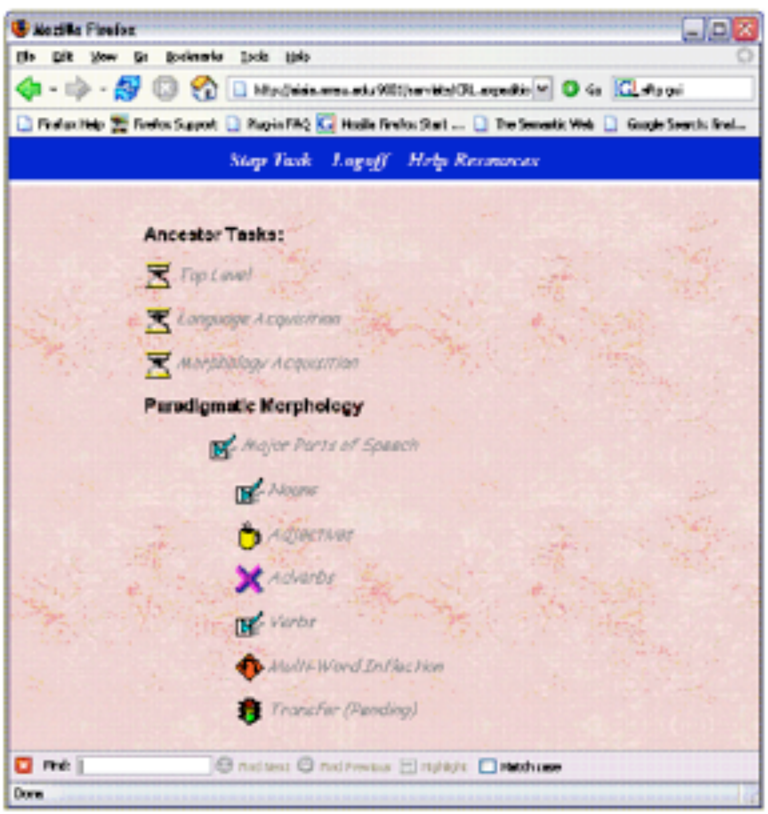


Figure 1: A Boas Navigation Page

---

[2] 'Ecology' is a term defined by Don Walker relating to issues connected with writing systems, text mark-up, punctuation, special symbols, dates, numbers, proper names, etc. Boas has a fairly comprehensive collection of ecological concepts.

Tasks in Boas are partially ordered. For example, the task of specifying the morphological features of a noun must be completed before the task of filling in a paradigm template for nouns. However, other tasks are independent—an informant can work on verb morphology before adjective morphology or vice versa. The navigation page enables the informant to select which task to work on from the set of available (all prerequisites met) tasks. The different icons on the navigation page represent different stages of completion (for example, the check mark indicates the task is completed).The ancestor tasks indicate where the informant is in the Boas task hierarchy. The top levels of this hierarchy are shown in Figure 2 (subtasks of BUILDING THE MT SYSTEM, and TUTORIALS are not shown).

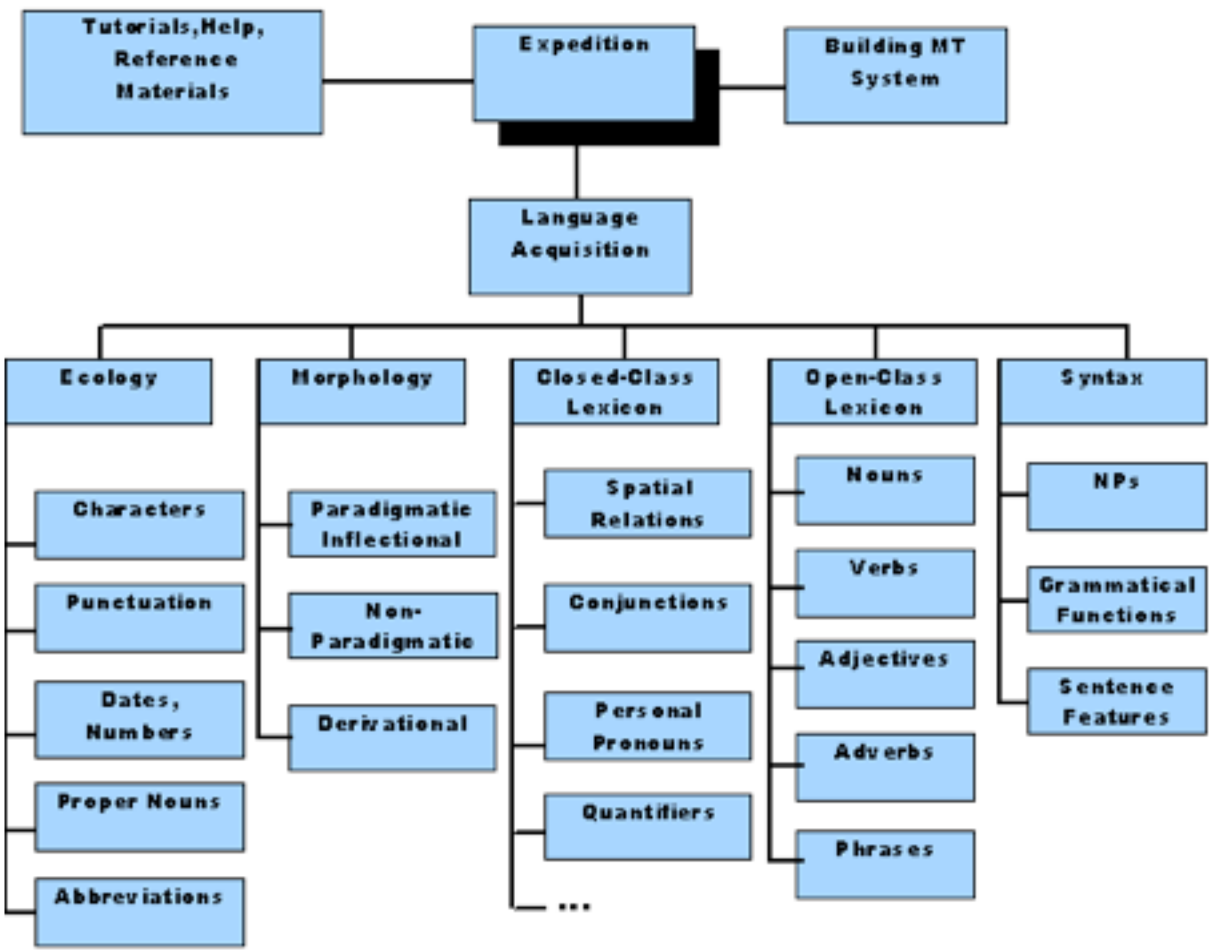


Figure 2: Top Levels of Boas

Knowledge acquisition pages either elicit information from the informant or provide pedagogical information. For example, Figure 3 shows the knowledge acquisition page for degrees of comparison. Both navigational and knowledge acquisition pages have a tool bar at the top with three options: stop task, which returns the user to the previous navigation page; logoff, which safely stops the current session; and help resources, which display a list of help options including alphabetical and hierarchical glossaries of linguistic terms.

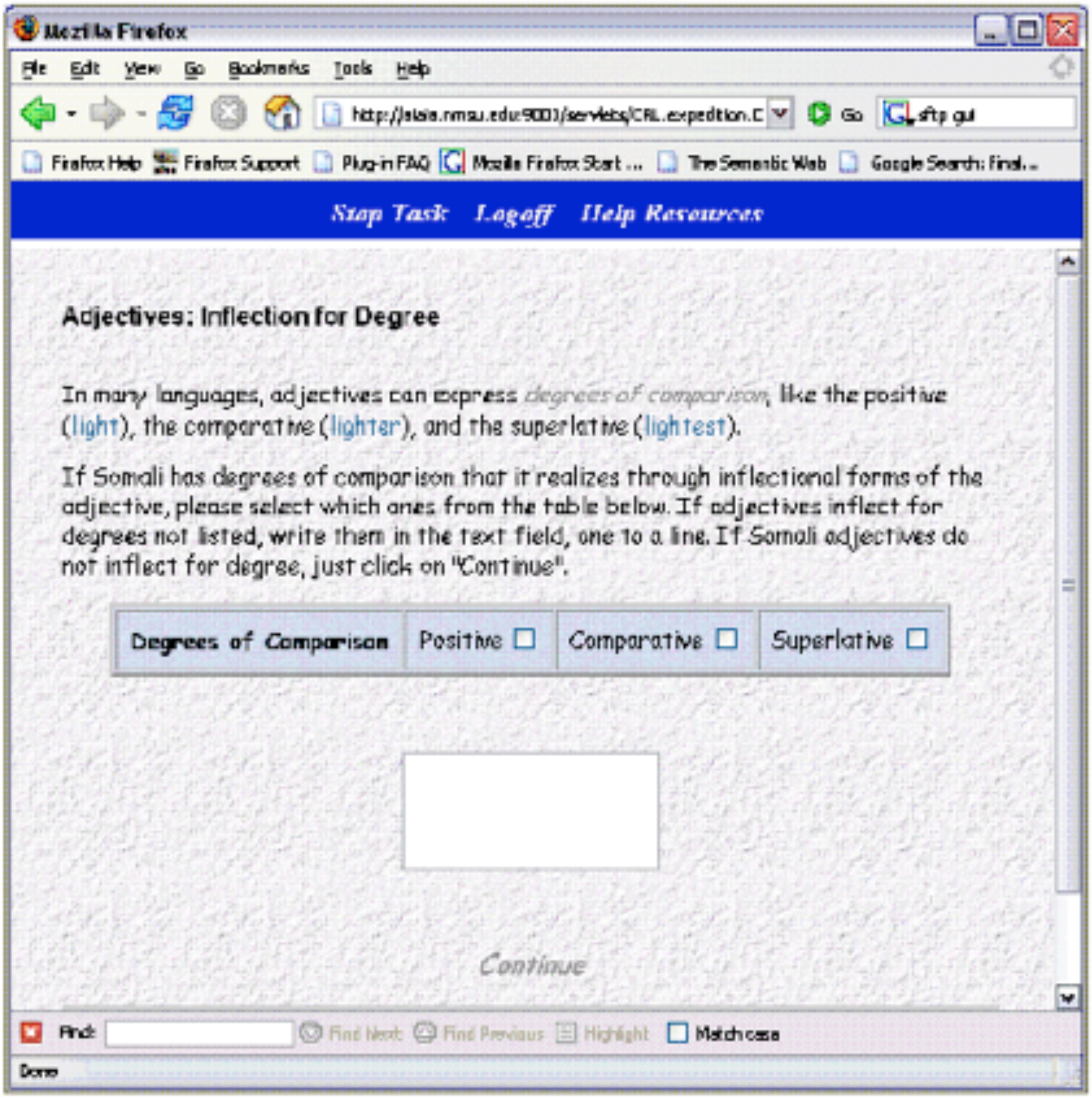
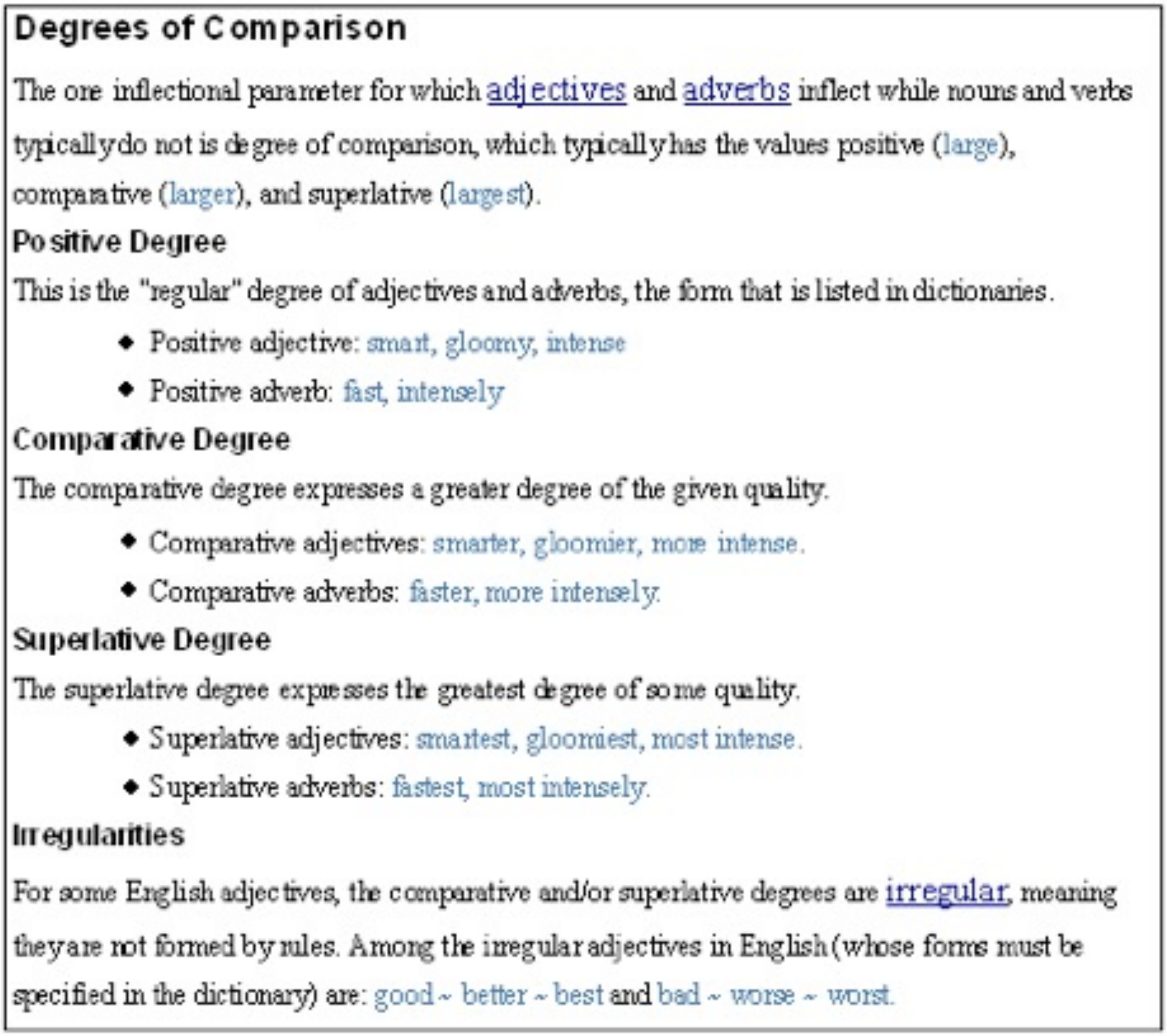


Figure 3: An Acquisition Page

All this pedagogical material, which amounts to a targeted online introduction to descriptive linguistics, was developed expressly for this project.[3] The knowledge acquisition pages are devoted to elicitation tasks, with page–specific help links where warranted. These page–level

[3] This extensive resource was developed by Marjorie McShane.

help links are just one of Boas's methods of “progressive disclosure,” which permits users of various levels of experience to use the same elicitation interface. Another help method is linking important terms in the acquisition page to their respective glossary entries. For example, the italic gray reference to degrees of comparison in the first line of Figure 3 is a link to the glossary page shown in Figure 4.



Degrees of Comparison

The one inflectional parameter for which adjectives and adverbs inflect while nouns and verbs typically do not is degree of comparison, which typically has the values positive (large), comparative (larger), and superlative (largest).

Positive Degree

This is the "regular" degree of adjectives and adverbs, the form that is listed in dictionaries.

- Positive adjective: smart, gloomy, intense
- Positive adverb: fast, intensely

Comparative Degree

The comparative degree expresses a greater degree of the given quality.

- Comparative adjectives: smarter, gloomier, more intense.
- Comparative adverbs: faster, more intensely.

Superlative Degree

The superlative degree expresses the greatest degree of some quality.

- Superlative adjectives: smartest, gloomiest, most intense.
- Superlative adverbs: fastest, most intensely.

Irregularities

For some English adjectives, the comparative and/or superlative degrees are irregular, meaning they are not formed by rules. Among the irregular adjectives in English (whose forms must be specified in the dictionary) are: good ~ better ~ best and bad ~ worse ~ worst.

Figure 4: An example of a glossary page

3. The Underlying Ontology of Boas

The performance of the Boas system is crucially dependent on the underlying ontology. Boas uses a pre–existing ontology, Mikrokosmos, to represent language-independent semantic information. This ontology is described in §3.1. The relations among linguistic concepts are represented in an ontology that is implicit in the Boas architecture and this ontology is described in §3.2.

3.1 The Upper Ontology

The machine translation system that is automatically produced by Boas is a transfer system, but underlying this transfer approach is an interlingual understanding of language use, where formal and structural linguistic categories are mapped into real–world semantic ones. These semantic concepts and categories are represented in the Mikrokosmos ontology (Mahesh, 1996, Nirenburg and Raskin, 2004). In contrast to many other ontologies, Mikrokosmos was specifically designed for natural language processing applications. For example, Mikrokosmos is the main repository of selectional restrictions. Consider the Spanish sentence El grupo Roche adquirio Docteur Andreu. Is the instance of the Spanish verb adquirio linked to the ontological concept ACQUIRE or LEARN since the word can mean either of these. The ontology specifies that the theme of LEARN needs to be an ABSTRACT–OBJECT. Since Docteur Andreu is an ORGANIZATION, which is a SOCIAL–OBJECT and not an ABSTRACT–OBJECT, the meaning of the verb is likely to be ACQUIRE.

Boas offers the acquirer a number of methods of specifying words and phrases for the open–class lexicon (see McShane et al. 2004). One method is to provide translations for roughly 60,000 English word senses. These word senses are divided into manageable sized lists first by part of speech and then by word frequency. Each word sense is linked to one or more concepts in the ontology. For example, when an acquirer provides a translation of the English word sense for the verb write, not only can the machine translation system provide a translation for that word, but the system now has access to additional information that may help in processing the source language sentence (for example, that the agent of that verb must be an instance of the concept HUMAN).

Recently, the Mikrokosmos ontology has been merged with several other ontologies. The resulting superset ontology, Omega, offers expanded functionality for natural language processing applications. For example, Omega includes the Lexical Conceptual Structure database containing 492 classes of verbs (representing 4,452 total verbs) (see Dorr 1997, 2000 and Traum and Habash 2000). Work is underway to include the Suggested Upper Merged Ontology, SUMO, (Niles and Pease 2001a, 2001b) as a subset of Omega.[4]

3.2 The Linguistic Ontology

An ontology of linguistic concepts is crucial to the performance of the Boas system for several reasons. First, the ontology provides a strict organization for knowledge elicitation. This is especially important when the expected informant has little or no formal linguistic training. Second, it serves as an anchor for cross–linguistic research and generalization. Finally, the programs built to process language require narrowly defined types of input and output.

While each ontological concept in Boas typically has an associated definition in the glossary, the real meaning of a concept is determined by the relationship that concept has with other concepts in the ontology. That is, the meaning of a particular concept is solely determined by the relationships that concept has with other concepts (see, for example, Woods 1991 and Mac Gregor 1991). An ontology is more than a taxonomic hierarchy of concepts. For example, the concept GENDER is a subtype of the concept INHERENT FEATURE associated as a property with the concept NOUN. These concepts and the relationships among them attempt to encode what we know to be universally true about language as well as concepts that encode what is possible in any language. Our knowledge about what is universally true and what is potentially true is imperfect and, thus, the ontology is not perfect. In §5 we present some problem cases. For example, the ontology underlying Boas indicates that an inflectional affix encodes some feature associated with the word that affix is attached to. However, as shown in §5.2, in some cases affixes attached to one word indicate features associated with a different word. Another example is the constraint that the surface form of a word is completely dependent on its lexical form. (See, for example, the discussion of the clear distinction between morphology and syntax described in Di Sciullo and Williams 1987.) For example, fly+s —> flies, love+ing —> loving, and big+er —> bigger. However, as shown in §5.3 the surface form of a word is sometimes determined by word–external factors.

The General Ontology for Linguistic Description (GOLD) is an ontology specifically designed for linguistic data analysis (Farrar and Langendoen 2003a, 2003b). A major aim of GOLD is to develop a common ontology that will enable linguists to share data that has a uniform markup and foster collaboration among linguists. The knowledge acquired from informants using Boas would be of greater use to this community if Boas used the GOLD ontology or, minimally,

---

[4] Hovy, personal communication.

provided a mapping from the ontological concepts in Boas to concepts in GOLD. Part of our future work deals with this issue. Our goal is to gradually move Boas to the GOLD ontology.

The upper level of the current Boas linguistic ontology is described in §4, where we describe how this ontology maps to Boas acquisition modules. In this section, we illustrate the low–level coverage of the system by describing one section of the ontology in detail—the case system.

3.2.1 The Case System

In the Boas ontology there are 30 cases as shown in Table 1.

Abessive	Dative	Locative
Ablative	Destinative	Nominative
Absolutive	Elative	Objective
Accusative	Ergative	Partitive
Adessive	Essive	Purposive
Allative	Evitative	Subessive
Aversive	Genetive	Superessive
Benefactive	Ilative	Superlative
Causative	Inessive	Translative
Commitative	Instrumental	Vocative

Table 1: The case system of Boas

There were several design goals for the Boas ontology. First, the ontology had to have descriptive adequacy. In our case, this meant that any source language would need to be defined in sufficient detail to enable the development of a machine translation system from that language into English. In the case system this means that for each case we have a specific transfer rule that allows us to transfer that case feature into an English realization. Another design goal was to have the ontology understandable to a naïve informant. If the distinction between two cases was extremely nuanced, then it might prove difficult to automatically elicit accurate information from our informants. Balancing of the descriptive adequacy and use–of–use design criteria led to our system of 30 cases.

In contrast to the Boas ontology, there are 60 cases represented in GOLD arranged in a hierarchical structure as shown in Figure 5. (The classes of cases are shown; specific cases are elided.)

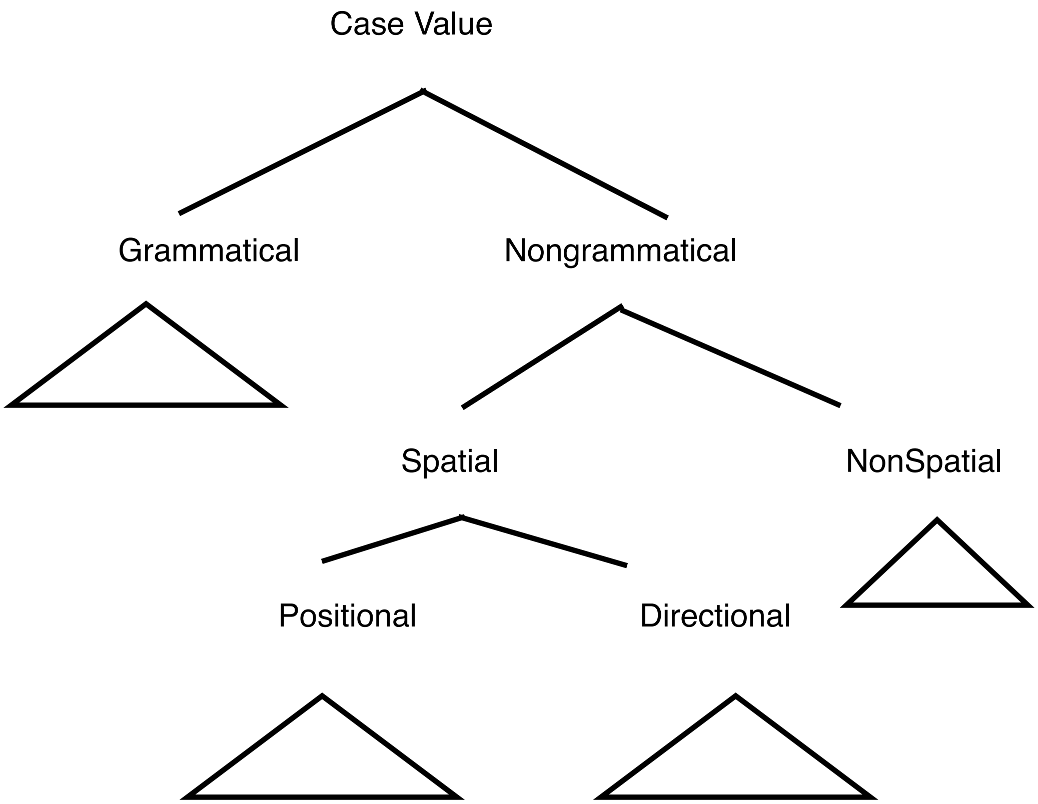


Figure 5: Case system of GOLD

This categorization of case classes offers distinct advantages over the flat Boas ontology. First, it provides improved descriptive adequacy. Not only is there an expanded number of cases, but the hierarchy itself allows for important generalizations to be made. For example, some selectional restrictions can be made at the level of the class of spatial cases, as opposed to repeating this information for each specific case. Secondly, ease-of-use is maintained by dividing the case system into well defined classes.

One problem in common with both Gold and Boas is that an individual may have a different meaning for a case than that specified in the ontology. For example, in GOLD the essive case is a subclass of spatial case, and has the definition “EssiveCase expresses that the referent of the noun it marks is the location at which another referent exists.” However, the Summer Institute of Linguistic’s glossary (http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/) defines essive as “a case that expresses the temporary state of the referent specified by a noun.” An example of this definition can be found in Finnish. In Finnish essive case is marked by adding the suffix -na: lapsi: ‘child’ —> lapsena: ‘as a child’ (“When I was a child”). A danger is that an acquirer might associate the GOLD concept of EssiveCase to the essive case in Finnish. Such an inaccuracy would create problems for a machine translation system—the MT system would interpret the essive case as specifying a location and possibly translate lapsena as ‘at a child’ in parallel to kotona —> ‘at home’; Luen lehtiä kotona —> “I read newspapers at home.” We attempt to avoid some of these problems by augmenting the case–based acquisition task with one based on an ontology of semantic concepts. Because our application is a very specific one—to acquire information to automatically generate a machine translation system from a given language to English-we drive the acquisition task from English realizations; in particular, realizations of closed–class items. That is, the underlying question the informant needs to answer is How does my language encode the concept expressed in English as x? A portion of this part of the ontology is shown in figure 6. The realizations of these concepts can take the form of a feature, a word, a phrase, or an affix

spatial relations

- below *The rock was below the surface of the water*
- beside *The man was sitting beside the driver*
- between *The hammock hung between the trees*
- *...*
- from (origin) *The boat from Stockholm should arrive shortly*

temporal relations

- ...
- from *We hope to go on holiday a month from today*

Figure 6: Semantic Concepts realized in English as Closed–Class Items

4. The Ontology Defining a Modular Architecture

Modularity is the notion that complex systems are partitioned into a set of special purpose, autonomous modules. One important aspect of modularity is that the input for each module is restricted—limited to the necessary and sufficient information required for the module's task. In addition, any particular module is minimally affected by the operation (and output) of other modules. These two aspects together are known as information encapsulation. For example, a person's auditory system is informationally encapsulated because it has restricted input—only information from the chochlea structures—and has limited, if any, access to other modules, like the visual perception system. One widely known account of modularity within cognitive science is presented in Fodor 1983, which proposes that modules are

I. **Informationally encapsulated**. Modules have strictly limited input, minimal interaction with other modules, and are not driven by central cognitive processes.

II. **Fast**. Fast processing is the result of encapsulation, since a given system need only consider specific information in a specific way: not all information need be interpreted by every possible cognitive system.

III. **Hard wired**. Fodor makes the conjecture that modules are hard-wired (not derived from induction or experience) and are localized in a particular area of the brain.[5]

IV. **Domain–specific**. Modules are often described as either horizontal—i.e., deriving from general reasoning ability, or vertical—i.e., domain–specific. Fodor considers language ability to be accounted for by vertica modules.

The assumption of modularity has been a driving force in descriptive linguistics, theoretical linguistics (e.g., Government and Binding Theory, Minimalism), and psycholinguistics (see, for example, Chomsky 1986, and Osherson and Lasnik 1990). The general view is that there is an innate language faculty that is distinct from that part of the mind responsible for general cognitive processing. The language faculty in turn consists of specialized modules for each language subtask (syntactic processing, lexical processing, etc.); each module has a well–defined task, specific inputs, and limited types of interaction with other modules. It is assumed, for example, that the syntactic processor does not have access to the speech waveform or even to the phonetic representation of that waveform. It is further assumed that there are no input loops

[5] For an alternative view regarding hard-wiring, see Karmiloff–Smith 1994.

between modules. Syntax, for example, cannot affect morphology since it receives input from morphology.[6]

The modules of Boas correspond to the underlying ontology. At the highest level there are 6 modules in the system, which are described briefly below.

a. Ecology: This module collects information about the writing conventions of a language, including the inventory of letters and punctuation marks, the treatment of numbers, dates, etc.

b. Inflectional Morphology: This module "learns" rules of inflection based solely on sample inflectional paradigms. The informant is guided through a process of building a paradigm template for each inflecting part of speech by answering questions about the parameters and values for which words inflect. For example, nouns might inflect for the parameter CASE using the values NOMINATIVE, GENITIVE, DATIVE, and for the parameter NUMBER using the values SINGULAR, PLURAL, DUAL. Once a paradigm template is established for a given part of speech, the informant provides all the inflectional forms of an inventory of examples that he or she selects. This inventory of examples should reflect all the productive patterns of inflection in the language. A morphological learner (Oflazer and Nirenburg 1999) creates rules of inflection based on the inflectional paradigms.

c. Productive Affixation: This module represents Boas's minimal treatment of derivational morphology. It collects two types of derivational affixes: (i) those that correspond to a small inventory of productive derivational affixes in English-e.g., affixes expressing negation (un–, non–, in–, anti–, counter–) and lesser degree (mini–, sub–, under–); (ii) affixes that only change the part of speech of the word with no significant shift in meaning—e.g., English –ly (joyful ~ joyfully). Considering that derivational morphology—particularly compounding and reduplication—is extremely widespread in natural language, one would think that Boas should handle it productively. However, these word–formation processes are often semantically ambiguous or non–compositional.

d. Syntax: This module learns phrase structure rules based primarily on translations of a graduated set of English phrases and clauses. This set has been designed such that the values of basic syntactic parameters like word order and agreement constraints can be determined.

e. Closed-Class lexicon: The closed–class lexicon contains an inventory of English closed–class items (pronouns, propositions, conjunctions, etc.) organized semantically. Informants are asked to provide as many equivalents for each English sense as are employed in their language. The

---

[6] For convincing counterevidence to the no-loop hypothesis, see Levelt and Maassen 1991, Dell 1986, and Bock 1987.

equivalents can take the form of a word, a phrase, an affix, or a feature. Examples (1) and (2) show cross–linguistic examples of affixal and feature realizations of closed class items respectively.[7]

(1)	BULGARIAN definite article:	*more* ~ *more**to*** ~ 'sea ~ the sea'
	RUSSIAN reflexive/reciprocal affix:	myt' ~ *myt'**sja*** 'wash ~ *wash oneself*'
	PERSIAN possessive pronoun:	*kt|b* ~ *kt|b**t*** 'book ~ your book'
	ARABIC preposition:	byt ~ ***b**byt* 'house ~ in a house'
	CREE possessive pronoun:	*astotin* ~ ***nit**astotin* 'cap ~ my cap'

(2) RUSSIAN

a. *On* *šel* ***lesom***
He-NOM walked woods-INSTR
He walked through the woods

b. *On* *ubil* *vora* ***nožom***.
He-NOM killed thief-ACC with-a-knife
He killed the thief with a knife.

f. Open-Class Lexicon: The open–class lexicon collects translations of nouns, verbs, adjectives, adverbs, phrases, collocations, and idioms.

These modules (a-f) based on the ontology cover most language phenomena. The adherence to modularity and the underlying ontology allows an informant to focus on providing just the knowledge associated with a particular aspect of language (inflectional morphology, syntax, etc.) rather than face the daunting task of interacting with an undifferentiated knowledge elicitation system that places the burden of organization on the informant. As concerns processing the

[7] The Cree example is from Wolfart 1981. All examples, here and elsewhere, that are not attributed to a source were elicited from informants or created by the authors.

information acquired from the informant, modularity allows us to create efficient, specialized programs to handle different aspects of language. For example, finite state machines can handle morphological analysis and chart–based parsing algorithms can handle syntactic analysis. In sum, the simplifying assumption of modularity provides numerous advantages for the architecture of the Boas system. However, not all facts about language fall neatly into the above mentioned modules; some crucial language phenomena fall between the cracks. These cross–modular phenomena, and Boas's treatment of them, are the subject of the next section.

5. Cross-modular Phenomena

For language phenomena that do not neatly fall into one of Boas's major modules, we developed tailor-made micro–components. A sample of these, described in terms of their expected functionality, is presented below.

5.1 Micro–component for Multi–word Inflection

Multi–word inflectional forms, like would have been going, straddle the line between morphology and syntax and are not acceptable input for Boas's morphological learner. Therefore, the task of establishing inflectional paradigms must be split into single–word and multi–word subtasks. Once the informant establishes a paradigm template, he or she is presented with that template and asked to indicate whether each combination of features is realized as a single word, multiple words, or either.[8] All single–word and 'either' entities remain in the main paradigm and are processed as by the morphological learner. All multi–word and 'either' entities are extracted and are sent to the multi–word micro–component.

The multi–word micro–component asks the informant to describe multi–word inflectional forms as the combination of auxiliaries and head words. The inventory of auxiliaries is collected in a prerequisite task. As concerns processing, multi–word inflectional forms present the same complexities as phrasals and idioms: often they can be scrambled and/or split by intervening words (I would definitely have gone).

5.2 Movement of Inflectional Affixes

Another phenomenon spanning morphology and syntax is the movement of inflectional affixes from their head words to another place in the sentence. Sometimes a moved affix cliticizes onto another word, sometimes not. A case in point is certain Polish person markers, which can move

[8] An example of 'either' is the Ukranian future tense: robitimu and budu robiti are both valid ways of expressing the 1st person singular 'will work'.

from their head verb to virtually any pre–verbal position. For example, the 1st person plural suffix šmy jas the legal placements shown in (3a–d).

(3) a. *My**śmy** znowu wczoraj poszli do parku*

we-1PL again yesterday went to park

b. *My znowu**śmy** wczoraj poszli do parku*

c. *My znowu wczoraj**śmy** poszli do parku*

d. *My znowu wczoraj poszli**śmy** do parku*

e. * *My znowu wczoraj poszli do**śmy** parku*

f. * *My znowu wczoraj poszli do parku**śmy***

"We went to the park again yesterday." (Franks and Bański 1999:125)

The processing problems for sentences like (3a–d) are obvious: the morphological analyzer will not find lexical matches for words like mysmy 'we–1PL' znowusmy 'again–1PL' or wczorajsmy 'yesterday–1PL'. In addition, the left–over verb forms in (3a)—(3c) will be incorrectly analyzed as 3rd person plural (plural forms have no person suffix).

The movement of inflectional affixes is handled in Boas using the micro-component Affix Movement. After the inflectional paradigms for a given part of speech are created, the informant is asked if affix movement occurs in the language. If so, he/she selects one paradigm to serve as a test case and highlights all affixes that can move. If different affixes from different paradigms can move the process is repeated for more paradigms. In the end, Boas will contain an inventory of mobile affixes similar to the inventories of affixes collected through the productive affixation and closed-class modules.

5.3 Spelling Changes Induced Word-externally

Yet another phenomenon that straddles morphology and syntax is spelling changes induced by word-external factors. For example, lenition and eclipsis in Irish are word–initial mutations triggered by certain types of preceding words. Table 2 presents a small sampling of such alternations.

Basic consonant	Lenited consonant	Eclisped consonant
c	ch	gc
b	bh	mb
g	gh	ng

Table 2: Irish alternations

Lenition can occur, for example, after the preposition ar ‘on’: bad ‘boat’ ar bhad ‘on (the) boat’; eclipsis can occur after the positive interrogative particle an: bris ‘break’ An mbriseann se...? 'Does he break ...?' (These processes occur in many other context as well and affect many other letters.)

6. Summary

In this paper we described Boas, a knowledge elicitation component designed to guide informants into providing linguistic information in sufficient detail to automatically construct a machine translation system. The quality of this acquisition system, and ultimately the quality of the resulting machine translation systems, directly depends on the quality and coverage of the underlying ontology. We have shown that adopting both a general ontology, Mikrokosmos, and an ontology of linguistic concepts helps us organize the acquisition of knowledge and to structure the resulting machine translation system. However, our current ontology fails because some language phenomena appear to violate constraints specified in our ontology. These failures have led to a refinement of our ontology and the construction of micro–components that handle these phenomena. Continued work on Boas will concentrate on making Boas more useful to a broader group of people and to allow the data that is collected by Boas to be of greater value to the linguistic community. One part of this effort is to make Boas more compatible with the GOLD ontology, and to make greater use of the Omega ontology.

References

BOCK, J. K. 1987. An effect of the accessibility of word forms on sentence structures. Journal of Memory and Language 26(2).119–137.

CHOMSKY, NOAM. 1986. Knowledge of Language: Its Nature, Origin, and Use. New York: Praeger.

DELL, G.S. 1986. A spreading activation theory of retrieval in sentence production. Psychological Review 93.283–321.

DORR, BONNIE J. 1997. Large–scale dictionary construction for foreign language tutoring and interlingual machine translation. Machine Translation 12(4).271–322.

DORR, BONNIE J. 2000. LCS database documentation. ms. http://www.umiacs.umd.edu/~bonnie/LCS_Database_Documentation.html

DI SCULIO, A. M.. and E. WILLIAMS. 1987. On the definition of word. Cambridge: MIT Press.

FARRAR, SCOTT, and D. TERENCE LANGENDOEN. 2003a. Markup and the GOLD ontology. Proceedings of EMELD–03.

FARRAR, SCOTT, and D. TERENCE LANGENDOEN. 2003b. A linguistic ontology for the semantic web. GLOT International 7(3).97–100.

FODOR, JERRY A. 1983. The Modularity of Mind: An Essay on Faculty Psychology. Cambridge, MA: MIT Press.

FRANKS, S. and P. BANSKI. 1999. Approaches to 'schizophrenic' Polish person agreement. Annual Workshop on Formal Approaches to Slavic Linguistics: the Seattle Meeting, 1998, ed. by K. Dziwirek and C. M. Vakareliyska. Ann Arbor: Michigan Slavic Publications, 123–143.

KARMILOFF–SMITH, A. 1994. Precis of: Beyond Modularity: a developmental perspective on cognitive science. Behavioral and Brain Sciences 17(4).693–745.

LEVELT, W.J.M., and B. MAASSEN. 1981. Lexical search and order of mention in sentence production. Crossing the Boundaries in Linguistics: Studies Presented to Manfred Bierwisch, ed. by W. Klein and W.J.M. Levelt. Dordrecht: Reidel, 221–252.

MCSHANE, MARJORIE, STEVE HELMREICH, SERGEI NIRENBURG, and VICTOR RASKIN. 2000. Slavic as a testing ground for a linguistic knowledge elicitation system. Annual

Workshop on Formal Approaches to Formal Slavic Linguistics: the Philadelphia Meeting, edited by T. H. King and I. A. Serkerina. 279–295.

MCSHANE, MARJORIE, SERGEI NIRENBURG, JAMES COWIE, and RON ZACHARSKI. 2002. Embedding knowledge elicitation and MT systems within a single architecture. Machine Translation 17(4).271–305.

MAC GREGOR, ROBERT. 1991. The evolving technology of classification–based knowledge representation systems. Principles of Semantic Networks: Explorations in the Representation of Knowledge, ed. By John F. Sowa. San Mateo, CA: Morgan Kaufmann, 385–400.

MAHESH, KAVI. 1996. Ontology development for machine translation: ideology and methodology. Memoranda in Computer and Cognitive Science MCCS–96–292, New Mexico State University.

NILES, I AND PEASE, A. 2001a. Towards a standard upper ontology. Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS–2001, ed. By Chris Welty and Barry Smith, 17–19. Ogunquit, Maine, Octover 17–19, 2001.

NILES, I AND PEASE, A. 2001b. Origins of the standard upper merged ontology: A proposal for the IEEE standard upper ontology. Working Notes of the IJCAI Workshop on the IEEE standard upper ontology, Seattle, Washington, August 6, 2001.

NIRENBURG, SERGEI. 1998. Project Boas: ‘A linguist in a box’ as a multi-purpose language resource. Proceedings of CoLing.

NIRENBURG, SERGEI, and VICTOR RASKIN. 2004. Ontological Semantics. Cambridge, MA: MIT Press.

NIRENBURG, SERGEI, and VICTOR RASKIN. 1998. Universal grammar and lexis for quick ramp–up of MT systems. Proceedings of CoLing 2.975–979.

OFLAZER, KEMAL, and SERGEI NIRENBURG. 1999. Practical bootstrapping of morphological analyzers. Proceedings of the Workshop on Computational Natural Language Learning at EACC, Bergen, Norway.

OSHERSON, D.N. and H. LASNIK. 1990. An invitation to cognitive science: Language. Cambridge: MIT Press. TRAUM, DAVID, and NIZAR HABASH. 2000. Generation from Lexical Conceptual Structures. Workshop on Applied Interlinguas. ANLP–2000. Seattle, WA.

WOLFART, H.C. 1981. Meet Cree: A guide to the Cree language. Lincoln: University of Nebraska Press.

WOODS, W. A. 1991. Understanding subsumption and taxonomy: A framework for progress. Principles of Semantic Networks: Explorations in the Representation of Knowledge, ed. By John F. Sowa. San Mateo, CA: Morgan Kaufmann, 95–108.