[*First International Conference on Language Resources & Evaluation*, Granada, Spain, 28-30 May 1998]

Project Boas: "A Linguist in the Box" as a Multi-Purpose Language Resource

Sergei Nirenburg
Computing Research Laboratory
New Mexico State University
Las Cruces, NM, USA
sergei@crl.nmsu.edu

Abstract

This paper introduces Boas, a semi-automatic knowledge elicitation system that guides a team of two people through the process of developing the static knowledge sources for a moderate-quality, broad-coverage MT system from any "low-density" language into English in about six months. The paper focuses on the issues in the elicitation of descriptive knowledge in Boas—the grammatical and lexical parameters and their values—and then discusses some of the elicitation techniques developed for the acquisition of the realizations of the subset of parameters and values for each particular source language.

Introduction

Project Boas is a component of a larger project, called Expedition, which is devoted to facilitating fast "ramping up" of machine translation systems from less studied, so-called "low-density" languages into English. The main result of Expedition will be a human-computer interactive system which will allow a team of two people—neither of them a linguist or MT developer—to acquire the knowledge for and configure a moderate-quality, broad-coverage MT system from a low-density language into English in about six months. One of the team members will know both the source language and English well and the other will be a computer programmer.

Machine-aided linguistic knowledge acquisition is a complex task even for seasoned MT developers. Experience shows that at the beginning of any project involving sizeable knowledge acquisition, staff members must be trained in a) the nature of data to be acquired; b) the requirements and preferences of the system for which knowledge is acquired; and c) the use of the acquisition tools. Training is usually led by a senior project member and an expert in the source language. In Expedition, the rules of the game do not allow for this initial training period. One must, therefore, incorporate training into the knowledge acquisition tool, making it, in effect, a "linguist in a box."

Output from Boas

Boas must help produce the same knowledge that human grammar and lexicon acquirers produce for an MT system. For each SL, this will include, at the coarsest grain size of description:

- a morphological analysis grammar;
- a syntactic analysis grammar (determining, at least, phrase boundaries, dependency structures inside and between phrases, and grammatical functions);
- knowledge about a set of special "ecological" features (dates, acronyms, numbers, punctuation treatment, proper name recognition rules, orthographic conventions, etc.);
- a feature and structure transfer grammar between the SL and English; and
- a bilingual SL-English dictionary (including phrasals, proper names and closed-class items).

When people are acquiring the above information, they have at their disposal, in addition to linguistic knowledge, reference grammars, text corpora and mono- and multilingual dictionaries for the language(s) in question. People typically decide on the methodology of the work themselves or are assisted by their supervisors.

With Boas, the methodological initiative rests with the system: it is the system that must lead the acquirer, ordering the interactions (questions) and keeping in mind the coverage needs and the nature of the output. While the acquirers will still have access to the printed (or online) descriptive grammars, dictionaries and other reference materials, the responsibility for quality and coverage of the output now rests with Boas.

It is easy to perceive a similarity between the task of the Boas system and the work of a field linguist. Both in knowledge acquisition for an MT system and in describing new languages there is a special methodology, an inventory of lexical and grammat-

ical phenomena to be elicited and a collaborator: the user of Boas plays a role similar to that of a field linguist's informant. There are, however, important differences. The field linguist aims at eliciting descriptive knowledge about a language. Boas is geared at acquiring knowledge related to processing language, specifically, at supporting a translation system from the language under description into English. Descriptive knowledge is sought in Boas only inasmuch as it is a prerequisite for formulating processing knowledge. The Boas user is assumed to be able to use reference materials and answer open-ended questions.

Ecology of Boas

The overall development and runtime environment in the Expedition project includes, in addition to Boas, two other major system components: a configuration and control system (CCS) and the runtime MT system. The relationship among the three systems is illustrated in Figure 1. CCS helps Boas by rounding up (and presenting to Boas users) online resources available for a particular source language. CCS also compiles the knowledge recorded through Boas into a format suitable for processing by the resident MT system and manages the file systems and the databases to support the MT engine, whose operation does not change with new source languages. The MT engine itself has been developed under the Corelli project at NMSU CRL (e.g., Zajac, 1996).

Put in practical terms, in order to be able to lead the user through the acquisition of the required static knowledge sources for a given language, the "linguist in the box" system must know at least the inventory of the SL's parts of speech, their orthographical (e.g., variant spellings), morphological (e.g., word formation), and syntactic (e.g., subcategorization) properties: the existence of features and their values relating to gender, animacy, number, noun class, and the manner of indicating such things as grammatical function, tense, person, etc., basic phrase and sentence structure patterns, word order preferences, and, last but not least, ways in which features required for English generation are realized in the SL (to support, at a minimum, feature transfer). More source language-specific information, such as the existence, number, and form of conjugation or declension classes, syntactic agreement between categories and in various features, must also be included.

The above knowledge is elicited from the user and stored in a language profile—a collection of key formal features (parameters?) of the language, and their values. But for the system to be able to elicit a language profile from the user, it must, in turn, have more knowledge: specifically, knowledge about the formal features and parameters in a variety of languages as well as the ways in which languages can express values of these parameters.

This inventory of grammatical categories (parameters), together with the legal values of such categories, attested in at least one language is known as the **parameter/value inventory.** This inventory must be compiled by system developers and somehow incorporated in

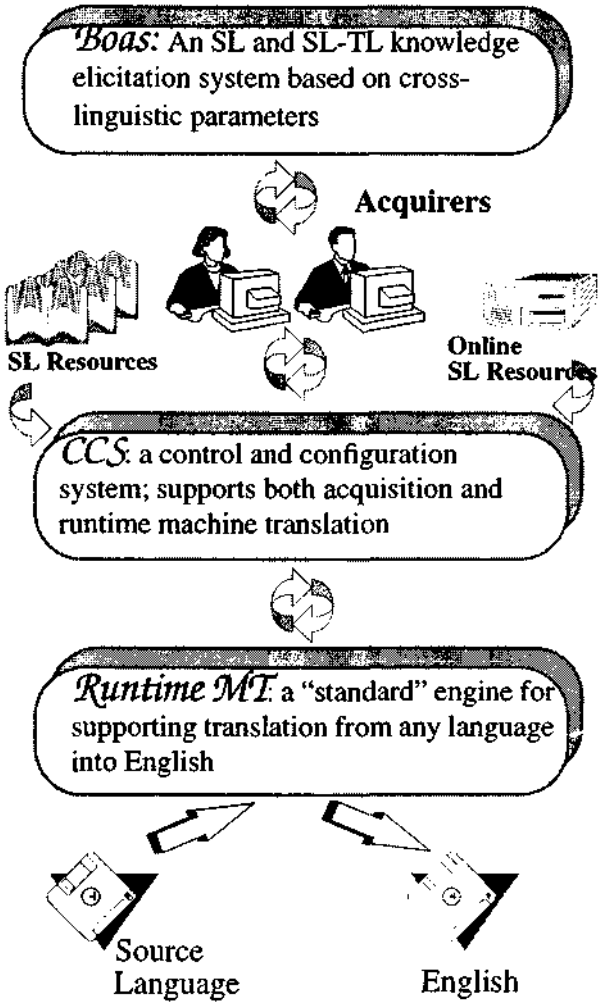

Figure 1. Boas as Part of the Expedition System. The acquirers obtain access to reference sources through the CCS. The SL operational knowledge elicited from the acquirers through Boas is fed, through the CCS, to the runtime MT system.

the elicitation system. The knowledge from the parameter/value inventory is used by the elicitation system developers to formulate the elicitation sessions which will subsequently pace the user of the elicitation system through the massive primary (lexical, syntactic, morphological and other) knowledge acquisition tasks.

A Glimpse into Diversity of Languages

There are many difficulties associated with building the parameter/value inventory describing a sufficient set of parameters for any language. Language variety is truly staggering. The following is just a tiny sample of such difficulties:

- Reduplication is the main means of forming plurals in Vietnamese and Bahasa Indonesia, while in Hausa there are up to forty different pluralization kinds, making use of affixes as well as reduplication.
- Syllable-final devoicing of occlusive consonants and vowel harmony are reflected in spelling in Turkish, making it more difficult to identify the citation form of a word.
- Hausa verbs are not inflected for tense, person or number. Tense is indicated by a marker attached to a preverbal pronoun; however sometimes this marker is only in the tone or vowel length change of the basic pronoun.

An Ostensible Redundancy

While Boas needs a complete list of all parameters in all languages, it is clear that each individual language uses only a subset of that list. For instance, the parameter of gender used by many languages including French, Ukrainian, Hebrew and many others is not utilized by English. Indeed, there is no inflection for gender in any word of English (see, e.g., Quirk *et al.,* 1985, pp. 314ff). The pronominal forms *he, she, his, her* and *hers,* do stand for the English nouns which incorporate the meaning of maleness or femaleness, but this agreement is due to deictic anaphoric rules pertaining also to person and number parameter values. As to the forms similar to *lion/lioness,* the relationship is derivational, relating to word- not form-formation.

So, it may appear that if English is a target language of an MT system, the gender parameter should be activated only for translating pronouns. Is it true then that in any particular MT situation only the parameter subset utilized by the target language is important? Why bother with determining the gender of a source language noun if there is no trace of this parameter value in the translation?

In fact, if the parameter of gender is active in a source language, it can be used not only for the purpose of transfer in MT but also to support the analysis of the source text before the bilingual step in machine translation. Thus, the gender of the antecedent of the Russian pronoun *kotoryy* in (l a,b) determines not only the form of the pronoun itself (which would not be reflected in an English translation) but, importantly, guides the dependency structure of the relative clause, resulting in quite different translations (2a, b, respectively).

(la) Ruka cheloveka, kotoraya podderzhivala polku, soskol'znula.
(1b) Ruka cheloveka, kotoryy podderzhival polku, soskol'znula.
(2a) The hand of the man which propped up the shelf slid down.
(2b) The hand of the man who propped up the shelf slid down.

In other words, it is still necessary in the framework of a Russian-English translation system to activate the parameter of gender. Thus, the set of parameters activated for an MT system is not determined by the target language alone. Rather, one should revive the thirty-year-old CETA hypothesis (see, for instance, Vauquois 1969; Veillon 1968) that the syntactic interlingua ("pivot language") is determined by a specific SL-TL pair, extending it to cover the set of activated parameters in an MT system.

We do not expect to end up being able to cover all the languages of the world. Instead, we strive to compile as many of the parameter sets as possible, with as many as possible attested realization options for a realistically large number of languages. A straightforward methodology for such an effort requires a parametrical exploration of a large number of potential target and source languages. Having only one target language and reusing available computational resources developed simplifies this methodology and, in fact, makes the whole enterprise feasible.

The Elicitation Process

Elicitation of descriptive and processing (operational) knowledge about a language consists in Boas of a set of elicitation "episodes." Descriptive knowledge is elicited about the source language open- and closed-class lexical items, about the grammatical (morphological and syntactic) categories in the source language and their values and context-related information including morphotactics and syntactic constituent structures. Operational information relates to bilingual issues—morphosyntactic feature transfer, lexical transfer, syntactic dependency transfer, etc. As the target language in Expedition is fixed, no information is elicited for the support of target text generation. All such information is resident in the runtime environment. The nature of the knowledge elicitation process practically requires that all knowledge, whether descriptive or operational be

recorded declaratively because it is much more difficult to set up an elicitation session for procedural knowledge acquisition. Conversely, all the underlying processors in the Expedition runtime environment are designed to operate with declarative knowledge sources.

The content which needs to be elicited through Boas ultimately leads to the acquisition of a necessary set of lexical meanings and semantic structure dependencies characterizing the source language. The difference from the interlingua (knowledge-based) translation approaches is in the way the elicited knowledge is recorded. In Expedition, it will be recorded as bilingual, SL-English correspondences rather than in an abstract metalanguage. Thus, instead of representing data for computer programs, we concentrate on preparation of knowledge necessary to elicit information about a specific source language from a human user. The knowledge which needs to be prepared is largely parametric.

Our work on knowledge acquisition in Boas is divided into three related parts: a) compiling as complete an inventory of possible parameters and their values as the resources permit, b) developing the elicitation episodes corresponding to this list, and c) implementing the elicitation techniques to acquire all realizations of parameter values in a given SL and all mappings of SL realizations into TL realizations.

A Taxonomy of Parameters

A significant portion of the inventory of parameters has already been acquired in the PROPERTY subtree of the Mikrokosmos ontology (e.g., Mahesh, 1996): most of the "grammatical" meanings (as in grammatical semantics—cf. Frawley 1993, Raskin 1994) are already recorded and systematized there. This inventory is in the process of being expanded. Sources for this expansion include information about languages not used in the development of the Mikrokosmos ontology and literature on field linguistics. This expanded information includes, for instance, the HONORIFIC mode, as in Korean or Japanese, itself a value of the MODE parameter, but having a range of values of its own that require considerable modifications in the English translations.

A taxonomy of parameters must include several orthogonal features. We suggest four such features. We expect that the composition and the number of features will both change over time. In principle, the assignment of a certain feature to a parameter affects how this parameter is treated in our work. This holds true for all the features below with the surprising exception of syntagmatic / paradigmatic feature.

Universal vs. Non-universal Parameters

Parameters can be universal (language-independent) or non-universal (language-specific) ones. There is a set of universal parameters, common for all languages in the sense that each parameter can and must be expressible in each language, as per the effability principle.[1] The importance of the universal parameters is that they determine the level at which transfer should be carried out if the translation is to be based on meaning. In other words, a parameter is universal if its value must be preserved in translation. This is because at that level all parameters are realizable in all languages, and the correspondence is in a single locus, the value of a universal parameter. Thus understood, transfer is, in fact, a two-step operation: a transition from a SL realization of a parameter into a universal parameter value and a transition from the latter to its realization in TL. It is not difficult to see that this description applies also to the interlingual MT scheme, and the universal parameters can be thought of as elements of an interlingua.

For a simple transfer to occur at the level of universal parameter representation, a complete syntactic identity of the sentences in SL and TL has to be obtained. This is plainly unattainable in the general case, and a considerable number of rules describing discrepancies in the syntactic structures of SL and TL put in correspondence in transfer will have to be elicited from the acquirer and formulated (see

1. The principle of effability of natural languages, asserts that "[e]ach proposition can be expressed by some sentence in any natural language" (Katz 1978: 209; see also Katz 1972/1974: 18-24, Frege 1963: 1, Tarski 1956: 19-21, and Searle 1969: 19-21—a view which is opposite to that formulated by Quine 1960: 26-30). In our work, both in Mikrokosmos and, especially, in Expedition, we are actually proceeding on a practical hypothesis formulated below (both in general and for computational purposes) which is a stronger form of this principle.

   Hypothesis of Practical Effability: Each sentence can be translated into another natural language on the basis of a lexicon with equally limited polysemy.

   Hypothesis of Practical Effability for Computational Microtheories: Any text in the source language can be translated into the target language in an acceptable way on the basis of a lexicon for the source language and a lexicon for the target language, such that their respective entries are limited in exactly the same fashion with regard to polysemy.

below).

Each universal parameter may be realized in a specific language using a set of non-universal parameters that the language in question selects from the "superset" of non-universal parameters. For instance, actualization is a universal parameter. Its values, also universal parameters, include time, modality, speech act, etc. Time is expressed in a specific language through a combination of non-universal parameters, such as tense, some special closed-class lexical items (*during, after, prior to, etc.*) and a small group of open-class lexical items, such as *yesterday* or *next June.*

Similarly, the universal parameter governance has as its values agreement, government and parataxis. Agreement, in turn, is a parameter which has as its values agreement in the subject-verb phrase, agreement in the noun-adjective phrase and so on. Each of these specific types of agreement is in itself a non-universal parameter which takes its values from the power set of values of GENDER, NUMBER, CASE and other possible properties which realize this kind of agreement in various languages.

GENDER, NUMBER and CASE are non-universal parameters, with their own values, e.g., {M, F, N, ...} for gender. M ("masculine") is not itself a parameter. It is a non-universal parameter value which has, associated with it, a set of language-dependent realizations, that is, inflectional paradigms. In some languages, realizations of values of several parameters are conflated into a single inflectional paradigm (for instance, in Russian, nominal declension paradigms are realized using endings (flective suffixes) which jointly express the values of gender, number and case). In some other languages (e.g., agglutinating languages, such as Turkish) suffixes typically realize a single parameter value.

Each language profile contains the entire set of universal parameters and a subset of the non-universal parameters applicable to this language. The specific set of such non-universal parameters is selected from the parameter inventory. It is clear that, say, the non-universal parameter of gender in, say, Ukrainian, may be different in its value set, function and assignment of realizations from the parameter which we also call gender in, say, Hebrew or German, to say nothing about the classifiers in Swahili or Lak which will also be termed gender. It is useful, we believe, to retain the same parameter name for all languages which feature it because they all share the invariant property of using the parameter mainly as an expression means for agreement.[2]

How can we differentiate between gender, as an expression of agreement, and number or case, which also express agreement, and thus not to bunch them together in a single parameter or confuse them with each other, especially since the actual realizations of these three parameters in a number of languages (e.g., German, Russian, French, English, Spanish) are fused into one suffix? While a complete differentiation procedure will be necessary at some step in the new language acquisition process, we will constrain our discussion here to an example. In the case of case, it will, in addition to agreement, serve the parameter of government (e.g., requiring the value of accusative on the direct object of a verb). In this it will be different from gender which serves only agreement.

Paradigmatic vs. Syntagmatic Parameters

Another way of taxonomizing parameters is along the syntagmatic/paradigmatic dimension. Such parameters as governance, predication and others pertaining to the actual text are syntagmatic. Other parameters, such as lexical categories (parts of speech), are paradigmatic. More specifically, paradigmatic parameters characterize individual words as members of lexical categories and their respective inflectional paradigms, if any. While there is no hard and fast criterion for distinguishing these types of parameters, we can offer several diagnostics. Paradigmatic parameter values are listed in dictionaries and/or grammars as properties of words. Paradigmatic parameters tend to be morphological rather than syntactic. Syntagmatic parameters tend to have conjunctive values while paradigmatic parameters tend to have disjunctive values: thus, the parameter of agreement is syntagmatic because its values are expressed as a combination of realizations of several (paradigmatic!) parameters such as gender, case and number. Each of these, however, is disjunctive because it requires the selection of exactly one of the values from its inflection paradigm, e.g., a noun cannot appear in a phrase in both genitive and ablative cases. The same unit of language can be seen as syntagmatic when it is presented as a sum of smaller compo-

2. Moreover, gender is typically made manifest through agreement between nouns and adjectives and/or possessive pronouns, which is why the classifier systems in languages like Swahili and Lak or, for that matter, the grammaticalized distinctions among English nouns along the dimensions of ± common, ±count, ±mass, ±animate, ±human and ±abstract (cf. Chomsky 1965) are not seen as gender.

nents and as paradigmatic when seen as a component of a larger unit. Thus the word as an aggregate of morphemes is syntagmatic, and as an element of a phrase is paradigmatic.

Lexical vs. Morphosyntactic Parameters

It seems that when parameters are discussed as a phenomenon, it is primarily the morphosyntactic ones that get attention. In the environment of translation, it is beneficial to include lexical parameters into consideration, too because lexical meanings must also be rendered across languages. Lexical parameters are viewed as language-independent lexical meanings (ontological concepts), such as $\text{table}_{\text{furniture}}$. The values of this parameter are the word senses corresponding to this ontological concept across the inventory of languages. The realizations for these values are the words or phrases that express this meaning in each language, with a possibility of a lexical gap (a null value) included. Lexical meanings qualify for parameterness exactly because, in our approach, they are linked across languages through the universal lexical meaning, realized in each language through word senses or phrase senses.

Terminal vs. Non-Terminal Parameters

Some parameters have other parameters as values, and other do not. The former are referred to as non-terminal, and the latter, as terminal. This means that the list of parameters includes several type hierarchies (e.g., that of governance, of sentence grammar, etc.). The semantics of the range property of non-terminal parameters is type inheritance.

Parameter Values

The inventory of parameter values is at this point much less complete. This inventory must include every grammatical meaning, for instance, each nominal case meaning which would include phenomena such as ergativity or the French partitive case, with *de l'eau* translating as "some water" and *l'eau* as "water" or "the water."

Boas and Methods in Field Linguistics

As far as elicitation techniques are concerned, some methodology has been adapted from field linguistics (see, for instance, Samarin 1967, Bouquiaux and Thomas 1992, Payne 1997). As the native speaker's input must be interpreted by a computational system, not a human, the field linguistics methodology is not applicable directly. Thus, in Comrie and Smith (1977), which is essentially a checklist of parameters for a field linguist, the existence and actual listing of the lexical categories in a SL is taken for granted, and the membership criteria never explored, a luxury that Boas cannot afford. Some recent attempts at automatic language knowledge acquisition (see, for instance, Knight 1996, Knight *et al.* 1995) are also of some relevance to our task. However, in the cited work the source of language information is not a native speaker, the scope of inquiry is more constrained, and the response range is more limited.

Boas experiments with new elicitation techniques, including asking the acquirers to consult a descriptive grammar of a language (if it is available) to derive comprehensive lists of phenomena, for instance, all the forms of the noun declension paradigm on a typical example with its translations into English, thus avoiding the need in a lengthy and tediously repetitive elicitation episode.

The inventory of parameters and values and the elicitation techniques in Boas are used and put to a test in the process of actual acquisition of the realizations of each parameter value in the SL. Thus, to return to the example of nominal case values, one has to a) elicit the noun inflection paradigms (if any); b) elicit prepositions (if any); c) combine prepositions and cases; d) elicit prepositional meanings; e) elicit meanings of preposition-case combinations (e.g., the Russian *s dereva* "from the tree," *s derevom* "with the tree," *s derevo* "the size of a tree"—see Nirenburg 1980); f) juxtapose these combinations with their parameter values. In the process of knowledge elicitation, the meanings can be expressed by the native speaker in a number of ways—ontologically, as English phrases, using pictures, diagrams, examples, etc. Multimodal representation, if made possible, improves the quality of acquisition by, among other things, breaking the tedium of the long sessions.

Conclusion

This paper presented an initial sketch of the goals, issues and methods in the Boas knowledge elicitation project. Many more concrete and general issues will become clear as the work progresses. One rather unfortunate circumstance that was detected early on is that neither the vast literature on principles and parameters in theoretical linguistics nor the publications in the area of field linguistics nor the contributions on the methodology of building knowledge elicitation systems is of any real help in our concrete task. While this state of affairs makes the work somewhat more risky, it is, at the same time, a pleasure to recognize the trailblazing nature of Boas and Expedition in their treatment of the interface between static and

dynamic knowledge sources, in the descriptive work on the generic parameter/value inventory and in the realization of the elicitation environment. An initial prototype of Boas already exists and is being tested by users. Of course, the ultimate judgment of Boas utility will be in its ability to elicit sufficiently fine-grain distinctions in the source language specification and the quality of the resulting MT system.

References

Ageno, A., I. Castellon, G. Rigau, H. Rodriguez, M. F. Verdejo, M. A. Marti, and M. Taule 1992. SEISD: An environment for extraction of semantic information from online dictionaries. In: **Proceedings of the 3rd Conference on Applied Natural Language Processing—ANLP '92.** Trento, Italy, pp. 253-255.

Ahlswede, Thomas E., Martha Evans, K. Rossi, and Judith Markowitz 1985. Building a lexical database by parsing Webster's Seventh Collegiate Dictionary. In: **Advances in Lexicology. Proceedings.** Waterloo, Ontario, Canada, pp. 65-78.

Amsler, Robert A. 1982. Computational lexicology: A Research Program. In: **AFIPS. Proceedings of the National Computer Conference, Vol. V, pp.** 657-663.

Amsler, Robert A. 1984a. Lexical knowledge bases. In: **Proceedings of COLING** '84. Stanford, CA: Stanford University, pp. 458-459.

Amsler, Robert A. 1984b. Machine-readable dictionaries. In: M. E. Williams (ed.), **Annual Review of Information Science and Technology,** Vol. 19. White Plains, N.Y.: Knowledge Industry Publications, pp. 161-209.

Apresyan Yury D., Igor' A. Mel'čuk, and Alexander K. Zholkovsky 1969. Semantics and lexicography: Towards a new type of unilingual dictionary. In: F. Kiefer (ed.), **Studies in Syntax and Semantics.** Dordrecht: Reidel, pp. 1-33.

Apresyan Yury D., Igor' A. Mel'čuk, and Alexander K. Zholkovsky 1973. Materials for an explanatory combinatory dictionary of Modern Russian. In: F. Kiefer (ed.), **Trends in Soviet Theoretical Linguistics.** Dordrecht: D. Reidel, pp. 411-438.

Asher, Nicholas, and Alex Lascarides 1995. Metaphor in discourse. In: Klavans *et al.,* pp. 3-7.

Atkins, B. T. *S.* 1991. Building a lexicon: The contribution of lexicography. In: Boguraev, pp. 167-204.

Attardo, Donalee Hughes 1996. Lexicographic Acquisition: A Theoretical and Computational Study of Linguistic Heuristics. An unpublished Ph.D. thesis, Graduate Interdepartmental Program in Linguistics, Purdue University, West Lafayette, Indiana.

Bartsch, Renate, and Theo Venemann 1972. **Semantic Structures.** Frankfort: Athenaum.

Barwise, Jon, and John Perry 1983. **Situations and Attitudes.** Cambridge, MA: M.I.T. Press.

Beale, Stephen, Sergei Nirenburg, and Kavi Mahesh 1995. Semantic Analysis in the Mikrokosmos Machine Translation Project. In **Proceedings of the Second Symposium on Natural Language Processing** (SNLP-95), August 2-4. Bangkok, Thailand.

Bendix, Edward H. 1966. **Componential Analysis of General Vocabulary.** Indiana University Research Center in Anthropology, Folklore, and Linguistics, Publication 41. Bloomington, IN: Indiana University.

Benson, Morton, Evelyn Benson, and Robert Ilson 1986. **Lexicographic Description of English.** Amsterdam: John Benjamins

Bergler, Sabine 1995. Generative lexicon principles for machine translation: A case for meta-lexical structure. In: Dorr and Klavans, pp. 155-182.

Berry, Margaret 1975. **Introduction to Systemic Linguistics 1: Structures and Systems.** New York: Saint Martin's Press.

Berry, Margaret 1977. **Introduction to Systemic Linguistics** 2: Levels and Links. London: Batsford.

Bloomfield, Leonard 1933. **Language.** New York: Holt.

Boguraev, Branimir (ed.) 1991. Special Issue: **Building a Lexicon.** *International Journal of Lexicography 4:3.*

Boguraev, Bran, and Ted Briscoe 1987. Large lexicons for natural language processing: Exploring the grammar coding system for LDOCE. *Computational Linguistics* 13, pp. 203-218.

Boguraev, Bran, and Edward J. Briscoe 1989. **Computational Lexicography for Natural Language Processing.** London: Longman.

Boguraev, Bran, Ted Briscoe, J. Carroll, D. Carter, and C. Grover 1987. The derivation of a grammatically indexed lexicon from the Longman Dictionary of Contemporary English. In: **Proceedings of ACL '87.** Stanford, CA: Stanford University, pp. 193-200.

Bolinger, Dwight 1965. Atomization of meaning. *Language* 41:4, pp. 555-573.

Bouquiaux, L. and J.M.C.Thomas. 1992. **Studying and Describing Unwritten Languages.** Dallas, TX: Summer Institute of Linguistics Press.

Bresnan, Joan (ed.) 1982. The Mental Representation of Grammatical Relations. Cambridge, MA: MIT Press.

Briscoe, T. 1993. Introduction. In: Briscoe *et al.,* pp.l-12.

Briscoe. Edward J., and Ann Copestake 1991. Sense extensions as lexical rules. In: **Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language,** Sydney, Australia, pp. 12-20.

Briscoe, Ted, and Ann Copestake 1996. Controlling the application of lexical rules. In: Viegas *et al.,* pp. 7- 19.

Briscoe, Edward J., Ann Copestake, and Bran Boguraev 1990. Enjoy the paper: Lexical semantics via lexicology. In. **Proceedings of COLING '90.** Helsinki, Finland, pp. 42-47.

Briscoe, Edward J., Ann Copestake, and Alex Lascarides 1995. Blocking. In: Patrick Saint-Dizier and Evelyne Viegas (eds.), **Computational Lexical Semantics.** Cambridge: Cambridge University Press, pp. 273-302.

Briscoe, Ted, Valeria de Paiva, and Ann Copestake (eds.) 1993. **Inheritance, Defaults, and the Lexicon.** Cambridge: Cambridge University Press.

Bruce, B. 1975. Case system for natural language. *Artificial Intelligence* 6, pp. 327-360.

Calzolari, Nicoletta 1984. Machine-readable dictionaries, lexical database and the lexical system. In: **Proceedings of COLING '84.** Stanford, CA: Stanford University, p. 460.

Cann, Ronnie 1991. **Formal Semantics: An Introduction.** Cambridge: Cambridge University Press.

Carlson, Lynn, and Sergei Nirenburg 1990. World modeling for NLP. Technical Report CMU-CMT-90-121, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA.

Charniak, Eugene, and Yorick Wilks (eds.) 1975. **Computational Semantics: An Introduction to Artificial Intelligence and Natural Language Comprehension.** Amsterdam: North-Holland.

Chierchia, Gennaro 1995. **Dynamics of Meaning: Anaphora, Presupposition, and the Theory of Grammar.** Chicago-London: University of Chicago Press.

Chierchia, Gennaro, and Sally McConnell-Ginet 1990. **Meaning and Grammar: An Introduction to Semantics.** Cambridge, MA-London: M.I.T. Press.

Chodorow, Martin S., Roy J. Byrd, and George E. Heidorn 1985. Extracting semantic hierarchies from a large on-line dictionary. In: **Proceedings of ACL** '85. Chicago: University of Chicago, pp. 299-304.

Chomsky, Noam 1957. **Syntactic Structures.** The Hague: Mouton.

Chomsky, Noam 1965. **Aspects of the Theory of Syntax.** Cambridge, MA: M.I.T. Press. *Computational Linguistics* 1987. Special Issue on the Lexicon. 13: 3-4.

Chomsky, N. 1981. **Lectures on Government and Binding.** Dordrecht: Foris.

Chomsky, N. 1986. **Knowledge of Language: Its Nature, Origin, and Use.** New York: Praeger.

Chomsky, N. 1995. **The Minimalist Program.** Cambridge, MA: MIT Press.

Comrie, B., and N. Smith 1977. Lingua Descriptive Studies: Questionnaire. Lingua 42:1, pp. 1-72.

Copestake, Ann 1990. An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary. In: **Proceedings of the First International Workshop on Inheritance in Natural Language Processing.** Toulouse, France, pp. 19-29.

Copestake, Ann 1992. The ACQUILEX LKB: Representation issues in semi-automatic acquisition of large lexicons. In: **Proceedings of the 3rd Conference on Applied Natural Language Processing--ANLP '92.** Trento, Italy, pp. 88-96.

Copestake, Ann 1995. Representing lexical polysemy. In: Klavans et al., pp. 21-26.

Copestake, Ann, and Ted Briscoe 1992. Lexical operations in a unification-based framework. In: Pustejovsky and Bergler, pp. 101-119.

Copestake, Ann, Ted Briscoe, Piek Vossen, Alicia Ageno, Irene Castellon, Francesc Ribas, German Rigau, Horacio Rodriguez, and Anna Samiotou 1994/1995. Acquisition of lexical translation relations from MRD's. In: Dorr and Klavans pp. 183-219.

Cruse, D. A. 1986. **Lexical Semantics.** Cambridge: Cambridge University Press.

Culikover, P. W. 1997. **Principles and Parameters. An Introduction to Syntactic Theory.** Oxford University Press.

Delbrück, B. 1919. **Einleitung in das Studium der indogermanischen Sprachen,** 6th ed. Leipzig.

Dillon, George L. 1977. **Introduction to Contemporary Linguistic Semantics.** Englewood Cliffs, N.J.: Prentice-Hall.

Dolgopol'skiy, Aron B. 1962. Izuchenie leksiki s tochki zreniya transformatsionnogo analiza plana soderzhaniya yazyka /A study of lexics from the point of view of the transformational analysis of the content plane of language. In: **Leksikograficheskiy sbornik** 5. Moscow: Nauka.

Dorr, B. 1993. Interlingual Machine Translation: A Parametrized Approach. *Artificial Intelligence* 63, 429-92.

Dorr, Bonnie J. 1993. **Machine Translation: A View from the Lexicon.** Cambridge, MA: M.I.T. Press.

Dorr, Bonnie J., Joseph Garman, and Amy Weinberg 1994/1995. From syntactic encodings to thematic roles: Building lexical entries for interlingual MT. In: Dorr and Klavans, pp. 221-250.

Dorr, Bonnie J., and Judith Klavans (eds.) 1994/1995. Special Issue: **Building Lexicons for Machine Translation I.** *Machine Translation* 9: 3-4.

Dorr, Bonnie J., and Judith Klavans (eds.) 1995. Special Issue: **Building Lexicons for Machine Translation II.** *Machine Translation* 10: 1-2.

Dowty, David 1979. **Word Meaning and Montague Grammar.** Dordrecht: D. Reidel.

Enç, Mürvet 1988. The syntax-semantics interface. In: Frederick J. Newmeyer (ed.), Linguistics: **The Cambridge Survey. Vol. I. Linguistic Theory: Foundations.** Cambridge: Cambridge University Press, pp. 239-254.

Farwell, David, Louise Guthrie, and Yorick A. Wilks 1993. Automatically creating lexical entries for ULTRA, a multilingual MT system. *Machine Translation* 8:3, pp.127-146.

Fauconnier, Gilles 1985. **Mental Spaces.** Cambridge, MA: M.I.T. Press.

Fillmore, Charles J. 1968. The case for case. In: Emmon Bach and Robert T. Harms (eds.), **Universals in Linguistic Theory.** New York: Holt, Rinehart, and Winston, pp. 1-88.

Fillmore, Charles J. 1971. Types of lexical information,