The VERBOSE Project

Jerry T. Ball[1], *Andrea Heiberg*[2]
Air Force Research Laboratory[1]
L-3 Communications at AFRL[2]

Draft August 2, 2007

Project Overview

The main objective of the **Ver**balization **B**etween **O**perators and **S**ynthetic **E**ntities (VERBOSE) project is to develop language-enabled synthetic entities capable of being integrated into training simulations to reduce the need for human resources who are not being trained. To achieve this goal without detriment in training, the synthetic entities must be capable of closely matching human behavior, including human language behavior. The initial application for the VERBOSE research is the creation of a synthetic entity capable of performing the functions of a pilot for an Unmanned Aerial Vehicle (UAV) in the three-person Cognitive Engineering Research on Team Tasks (CERTT) (Cooke & Shope, 2005) testbed of the Cognitive Engineering Research Institute (CERI). The UAV pilot, or Air Vehicle Operator (AVO), controls the simulated UAV flight path and communicates with the other members of the UAV team—the sensor operator and the planning officer—in the performance of reconnaissance missions. The synthetic UAV pilot must be capable of performing these functions in a manner comparable to that of human participants.

Many of the components of the VERBOSE project are under development and this document is as much a planning document as it is a description of an existing system.

VERBOSE Architecture Overview

The VERBOSE project is intended to lead to development of an end-to-end language processing system. The core of the system is being implemented within the ACT-R cognitive architecture (Anderson and Lebiere, 1999; Anderson et al., 2004). The use of ACT-R reflects a focus on developing a cognitively plausible system capable of closely modeling human behavior. As argued in Ball (2006), given the inherently human nature of language, the use of a cognitive architecture to build a language comprehension system may actually facilitate, rather than hinder, development. The constraints imposed by the cognitive architecture push system development in cognitively plausible directions which are more likely to lead to human-like behavior than purely algorithmic solutions.

The major linguistic components of the system include language comprehension and generation components which are under the control of a discourse manager (see Figure 1). The linguistic subsystem will interact with a situation model (Zwann & Radvansky, 1998) that is a spatial-imaginal representation of the current state of affairs as encoded from visual and linguistic inputs. The situation model is an important component of the system, although the details of how to implement the situation model are yet to be worked out. The situation model functions to give meaning to the linguistic representations which contain words rather than word senses or abstract concepts. A motor behavior component is available to affect motor actions. Input to the system is mediated by ACT-R's perceptual module and motor actions are mediated by ACT-R's motor module. The perceptual and motor modules are ACT-R's interfaces to the external environment. All of the ACT-R internal components make use of ACT-R's declarative memory and production system. The situation model will make use of the spatial cognition capabilities that are currently being designed for inclusion in the ACT-R architecture by members of the Performance and Learning Models (PALM) team of the Air Force Research Laboratory (Gunzelmann & Lyon, in press).

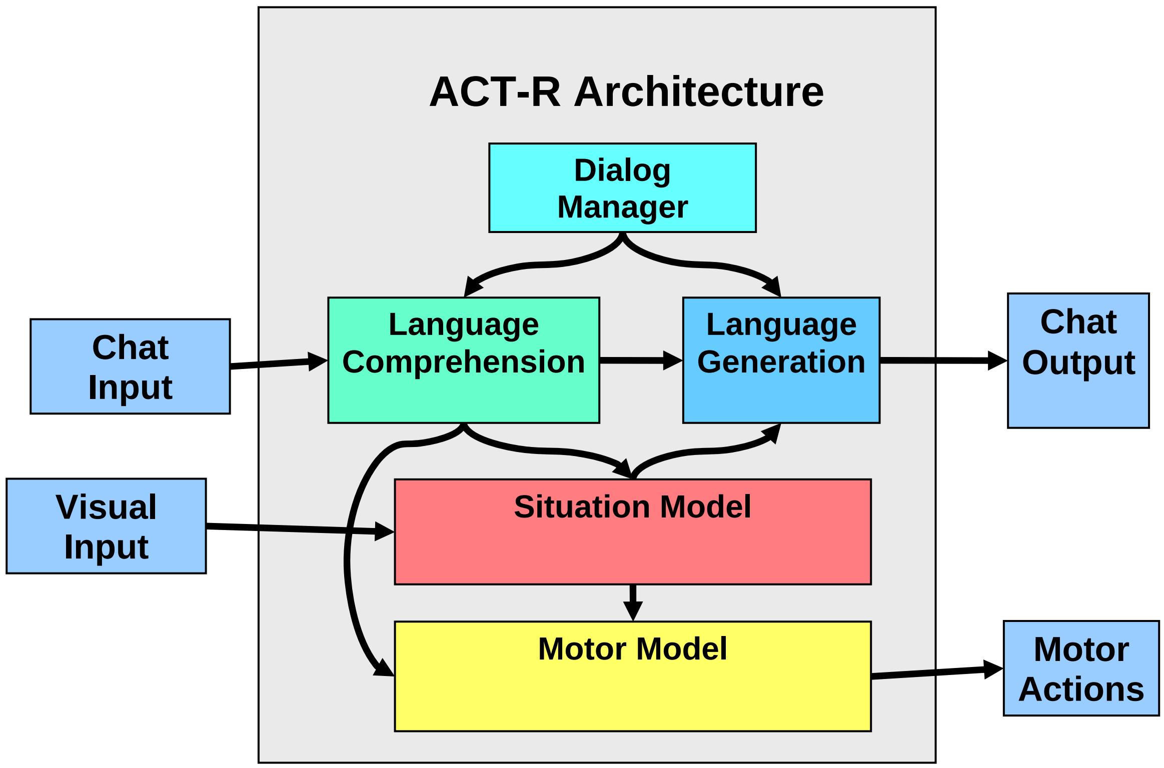

Figure 1: The VERBOSE Architecture

Language Comprehension

Most of the current research has been focused on the development of the language comprehension component of VERBOSE. Ball (2007) describes a construction-driven language comprehension system consistent with basic principles of cognitive linguistics and construction grammar which represents the theoretical commitments of the comprehension system. Ball (in press, 2005) describes the linguistic theory underlying system development. Ball, Heiberg & Silber (2007) describes the initial implementation of a prototype language comprehension system, called Double R Model (Referential and Relational Model), in ACT-R 6. An ACT-R 6 interface to WordNet has been developed by Bruno Emond (Emond, 2006) and is being integrated into Double R Model.

Double R Model at present demonstrates the possibility of implementing a cognitively plausible language comprehension system in ACT-R. The language comprehension model is currently capable of processing a range of grammatical constructions attested in the CERTT transcripts, including:

- Intransitive verb: “You can go.”
- Transitive verb: “We already hit [$_{\text{OBJ}}$ ROW].”
- Ditransitive verb: “You can give [$_{\text{IOBJ}}$ me] [$_{\text{OBJ}}$ R-STE].”
- Verb taking clausal complement: “You told [$_{\text{IOBJ}}$ me] [$_{\text{SITCOMP}}$ the altitude restriction was below 3000 feet].”
- Auxiliary verb: “I would have had a wrong picture.”
- Predicate nominal: “First waypoint is LVN.”
- Predicate adjective: “Altitude is stable.”
- Predicate preposition: “We are in those constraints.”
- Attributive adjective modifier: “It's a good picture.”
- Adverbial modifier: “Our altitude still should be fine.”
- Complex nominal: “The next photographic target point is M-STR.”
- Nominal conjunction: “We will maintain current airspeed and altitude.”
- Sentence conjunction: “The entry is KGM and the exit is FRT.”

The model creates a linguistic representation of the input, but doesn’t yet map that representation to the corresponding objects and situations in the situation model.

The language comprehension model is approaching a scale and complexity atypical of most cognitive models. Verifying that the model generates theoretically motivated linguistic representations is an important on-going aspect of the project. Inputs to the model are comprised of actual utterances from the UAV-STE transcripts and a set of canonical phrases and sentences. The verification strategy includes running the model against this set of inputs, and testing that the model produces the expected output.

The model generates linguistic representations which include such information as phrase constituency, predicate/argument relations, head/modifier relations, and head/specifier relations. Linguistic representations are complex structures of DM chunks. For testing, the DM chunk structure is converted into a graphical representation (automatically generated with phpSyntaxTree, Eisenbach & Eisenbach, 2006) shown in Figure 2 (below).

At a gross level, testing is fully automated. The complex output structure (e.g., Figure 2) is traversed in left-to-right order, and the terminal symbols are reassembled into a string (e.g., “I increased the airspeed”). This output string is compared to the input string; any mismatches are flagged for further investigation. At a more detailed level of testing, the output representation is hand-checked to ensure its validity. Valid output representations are stored as the known-good baseline. A capability to dynamically visualize the evolving DM representation during the processing of each word in an input text also exists

(Heiberg, Harris & Ball, 2007). Any further changes to the model may be easily regression tested by regenerating the outputs, and comparing them to the known-good baseline with an automated file comparison tool. This set of methodologies taken together helps facilitate the development of a large scale and complex model.

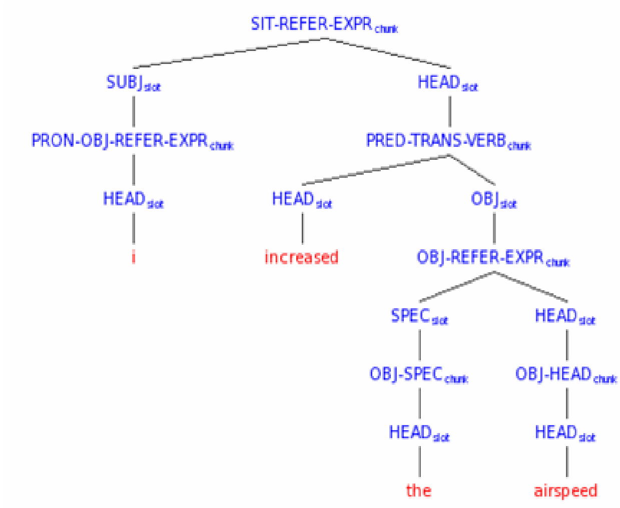

Figure 2 Graphical Representation

Despite these capabilities, the model falls short as a functional system and will need to be significantly expanded in the VERBOSE effort. The primary areas in which Double R Model needs to be expanded are listed below:

- Expand the lexical coverage
- Provide support for multi-word units and constructions with multiple lexical items interspersed with variable elements
- Expand the coverage of the grammar to encompass the full range of grammatical forms which occur in UAV reconnaissance mission comm. (e.g., questions, imperatives, relative clauses)
- Provide support for lexical and structural disambiguation
- Provide support for use of non-literal language
- Add a situation model to ground the referring expressions in the UAV comm.
- Add domain specific and general knowledge relevant to the UAV reconnaissance mission domain

Besides these additions and extensions to the language comprehension component, there are several other components that need to be developed and several additional requirements:

- Empirically validate the model
- Language generation
- Discourse management

- Develop model of UAV pilot performing reconnaissance missions
- Integrate the language-enabled synthetic entity into the CERTT Testbed

All of the above will require significant effort to accomplish. They are discussed in more detail below.

Lexical Coverage

Double R Model currently contains a small vocabulary of content words along with a more complete coverage of function words. To scale up the lexical coverage of the system, the WordNet (http://wordnet.princeton.edu/) lexicon will be integrated into the system. Bruno Emond of the National Research Council of Canada has recently developed an ACT-R 6 interface to WordNet, WN-Lexical (Emond, 2006), that will integrated into Double R Model. The integration of WordNet introduces a number of challenges, including the importance of addressing the issue of lexical ambiguity. For example, the word "dog" has eight word senses listed in WordNet. Unfortunately, WordNet does not provide any frequency information for different word senses that could be used to set the word sense base levels in ACT-R's declarative memory. Further, although WordNet does not explicitly contain word senses, WordNet synsets are collections of synonymous words and words can be members of multiple synsets which indirectly raises the issue of whether or not word senses are psychologically real (Kilgariff, 1997).

Multi-Word Units and Constructions with Multiple Lexical Items

Double R Model currently relies on constructions associated with individual words to support language comprehension. For example, the word "kick" is associated with a transitive verb construction [Subj *kick* Obj] (e.g., [[Subj *I*] *kick* [Obj *the ball*]]). Processing of the word "kick" activates this construction, making the construction available to support language processing. In addition to individual word constructions, humans rely on the use of multi-word units and constructions containing multiple lexical items interspersed with variable elements. Double R Model needs to be expanded to support the processing of multi-word units and constructions with multiple lexical items. This means that a construction initially selected on the basis of a word in the input will often need to be replaced by a multi-word unit or a construction with multiple words. For example, "kicked" followed by "the" and "bucket" should activate the more specialized [Subj *kicked the bucket*] idiomatic construction which replaces the more general [Subj *kick* Obj] construction. A mechanism of context accommodation for doing this which does not involve algorithmic backtracking is currently being implemented. A benefit of having multi-word units and constructions with multiple lexical items is that they are less likely than single words to be ambiguous. In fact, multi-word units and constructions with multiple lexical items are among the primary mechanisms for dealing with ambiguity. A key principle of Double R is to recognize the largest units of meaning, rather than trying to derive meanings via composition of the meanings of individual words.

Grammatical Coverage

Double R Model currently covers an interesting range of grammatical constructions. However, it is not close to providing full grammatical coverage. Further, UAV comm. is likely to have its own set of specialized constructions. More generally, FrameNet (http://www.icsi.berkeley.edu/~framenet/) and VerbNet (http://verbs.colorado.edu/~mpalmer/projects/verbnet.html) are candidate systems for extending the grammatical coverage of Double R Model, but FrameNet frames and VerbNet templates will need to be converted into appropriate constructions for use in Double R Model.

Lexical and Grammatical Disambiguation

English is highly ambiguous, both lexically and grammatically. While it may be possible in a toy grammar to disallow such ambiguity, any real system will need to deal with it. The theory underlying Double R Model does not make use of separate word senses (or abstract concepts), and word sense disambiguation in Double R Model will be modeled in terms of what associated words and non-linguistic representations in the situation model are co-activated. Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) may be used for the empirical determination of the degree of association of lexical items (Levin, Sharifi & Ball, 2006) to support lexical disambiguation in conjunction with ACT-R's spreading activation and decay mechanism. Grammatical disambiguation will be modeled in terms of the objects and situations in the situation model to which linguistic expressions refer.

Situation Model

A situation model is the mechanism that Double R Model uses for grounding referring expressions. This is an active area of research and it has not yet been determined how to go about implementing this capability. The key issue is that the objects and situations represented in a situation model should incorporate spatial, temporal and visual features, but it is unclear how to provide a computational representation for such features short of solving the "vision" and "time" problems. There are several candidate systems being considered as the basis for developing a situation model:

- Situation Models a la van Dijk & Kintsch (1983; Kintsch, 1998) as extended by Zwann, Graesser and others (cf. Zwann, Langston & Graesser, 1995; Zwann & Radvansky, 1998)
- Mental Models a la Johnson-Laird (1983)
- Conceptual Spaces a la Gardenfors (2000)
- Mental Spaces a la Fauconnier (1985)

From a model theoretic semantics perspective, Kamp's Discourse Representation Theory (cf. Kamp & Reyle, 1993) is also a candidate for development of the situation model with extensions to support spatial-imaginal information and the representation of other minds.

Domain Specific and General Knowledge Base

The Situation Model is constructed from the linguistic and visual input within the context of the model's (or team member's) background knowledge. UAV reconnaissance mission knowledge is crucial for representing the knowledge that UAV team members use in their communication and for generating situation models. Implementation of the knowledge base will require a detailed investigation of the knowledge needed and will require a person familiar with the task and knowledge engineer several months to develop an initial knowledge base. Besides this domain specific knowledge, ConceptNet (http://web.media.mit.edu/~hugo/conceptnet/), Open CYC (http://www.opencyc.org/doc) and MikroKosmos (http://crl.nmsu.edu/Research/Projects/mikro/htmls/oldstuff-htmls/training-intro.html) are candidate systems for supporting the creation of a knowledge base of general knowledge, but they will need to be integrated with Double R Model. Ball, Gluck & Rodgers (2004) describes an initial attempt to integrate CYC into Double R Model. Some mechanism for loading the relevant portions of the knowledge base at runtime is likely to be needed to avoid overloading the system.

Empirical Validation

The claim that Double R Model is a cognitively plausible model of language comprehension (Ball, 2004) needs to be empirically validated. While there are empirical results which motivate Double R Model (cf. Ball, 2004), it is less clear how to validate the processing mechanism given the lack of a firm connection to real-time performance. The basic reality is that the more complex a cognitive model is, the less directly tied to specific psychological measures it is reasonable to expect it to be. For example, to compute reaction times in text comprehension, one needs to closely model eye movements. But since moving the eye is effectively on the critical path for processing, what you wind up measuring is eye movements (David Kieras, personal communication). Most of the text comprehension processes are going on behind the scenes of eye movements and cannot be effectively measured. It is only at the end of a clause or sentence that extended eye fixations suggest that there is something going on other than eye movements and lexical encoding. The processing of garden path sentences (cf. Bever, 1970; Townsend & Bever, 2001) is an exception in that regressive eye movements and extended fixations before the end of a sentence provide an indirect measure of language comprehension difficulties which is not in evidence during the processing of normal sentences. It is in part for this reason that garden path sentences have been studied extensively by the psycholinguistic community. The visual world paradigm (cf. Tanenhaus et al., 1995) is a newer technique for getting at language comprehension processes by tying linguistic expressions to visual scenes that are described by these expressions. The results of this paradigm provide evidence for a close connection between linguistic and non-linguistic visual processing which argues against any strict separation of syntactic and semantic processing—a theoretical position which has long dominated psycholinguistic research, but which is rejected in Double R Theory.

Although it is difficult to validate complex language comprehension processes given the perceptual processes which mask them, the representations that are generated during processing are open to analysis. It is entirely reasonable to ask subjects to reflect on the representations they create during language comprehension, despite the long held belief that this is inappropriate (Ball, in preparation). A new line of research is currently getting underway to empirically study linguistic representations as a way of validating the representations proposed in Double R Grammar.

Language Generation

The ACT-R model of the language generation component of the VERBOSE project selects an optimal utterance from declarative memory, given a set of attributes for the utterances and a set of ranked constraints about the attributes.

The specific task to be modeled is the selection of an utterance, based on pragmatic constraints (ala Varges 2006). In the system as a whole, the synthetic AVO comprehends an utterance, and decides on the basic content of a response. Then the language generation component chooses from a set of response utterances, based on pragmatic constraints. This principled variation in response gives human-like behavior to the synthetic AVO.

The data set comes from transcripts of the dialogue of all-human runs of the simulation in the CERTT testbed. Human AVO utterances from these transcripts have been catalogued into a set of types; the variation within these types will be further classified according to pragmatic factors. The model will be evaluated on how well it corresponds to utterance selection as evidenced in the transcripts.

Optimality Theory (OT) (Prince & Smolensky 1993) is modeled. The three major components of OT are:

- Con: the hierarchy of violable constraints
- Gen: generates potential candidates
- Eval: evaluates candidates against the constraint hierarchy to find the optimal candidate(s)

The optimal utterance is the one that least violates the constraint hierarchy. The optimal utterance likely incurs some constraint violations; crucially, these violations are less important than the violations of the other utterances. Utterances may tie on any constraint, and there may be more than one optimal utterance.

To capture the ways in which the AVO’s utterances change over time, it is proposed that the constraint ranking changes over time. External factors influence the constraint hierarchy. This is akin to the dynamic reranking of optimality theoretic constraints for dialogue systems of Ericsson (2004), and the adaptive content selection of Walker et al. (2004).

The different constraint rankings produce the *factorial typology* of possible systems. For OT, the factorial typology typically describes different language systems (Prince &

Smolensky, 1993). However, for the pragmatic domain here, it is proposed that the complete set of utterances predicted by the constraint system is found in a single language. In fact, the complete range of utterances is produced by a single speaker, in a single task domain.

Discourse Management

…

Piloting a UAV

A cognitive model of a UAV predator pilot flying a reconnaissance mission is currently under development as part of an AFOSR grant (cf. Gluck, Ball & Krusmark, 2007). The original intent was to use this model as the basis for the language-enabled UAV pilot to be integrated into the CERTT testbed. However, the piloting requirements of the cognitive model which is under development and the requirements for the pilot in the CERTT testbed are significantly different. At this point, it is unclear to what extent the UAV predator pilot model will transfer to the CERTT testbed.

…

Integration into the CERTT Testbed

The language-enabled UAV pilot will ultimately need to be integrated into the CERTT testbed, replacing the human operator. The software (and hardware) infrastructure needed to support this integration is likely to be significant. The functionality of the pilot GUI will also need to be implemented within the Lisp environment in which ACT-R runs, since the ACT-R model cannot directly encode information from the pilot GUI.

References

Anderson, J. & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: LEA.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S, Lebiere, C, and Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review* 111, (4). 1036-1060

Ball, J. (in preparation). Empirical study of the form and function of linguistic expressions.

Ball, J. (in press). A Bi-Polar Theory of Nominal and Clause Structure and Function (journal article). *Annual Review of Cognitive Linguistics*.

Ball, J. (2007). Construction-Driven Language Processing. *Proceedings of the 2nd European Cognitive Science Conference*, pp. 722-727. Edited by S. Vosniadou, D. Kayser & A. Protopapas. NY: LEA.

Ball, J. (2006). Can NLP Systems be a Cognitive Black Box? In *Papers from the AAAI Spring Symposium, Between a Rock and a Hard Place, Cognitive Science Principles Meet AI Hard Problems.* Technical Report SS-06-02: AAAI Press.

Ball, J. (2005). A Bi-Polar Theory of Nominal and Clause Structure and Function. In *Proceedings of the 2005 Cognitive Science Conference*. Sheridan Printing.

Ball, J. (2004). A Cognitively Plausible Model of Language Comprehension. In *Proceedings of the 13th Conference on Behavior Representation in Modeling and Simulation*, pp. 305-316. ISBN: 1-930638-35-3

Ball, J., Heiberg, A. & Silber, R. (2007). Toward a Large-Scale Model of Language Comprehension in ACT-R 6. *Proceedings of the 8th International Conference on Cognitive Modeling*, pp. 163-168. Edited by R. Lewis, T. Polk & J. Laird.

Ball, J., Rodgers, S., & Gluck, K. (2004). Integrating ACT-R and Cyc in a Large-Scale Model of Language Comprehension for Use in Intelligent Agents. In *Papers from the AAAI Workshop, Technical Report WS-04-07*, pp. 19-25. Menlo Park, CA: AAAI Press

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (ed.) *Cognition and the Development of Language*. NY: Wiley.

Cooke, N. & Shope, S. (2005). Synthetic Task Environments for Teams: CERTT's UAV-STE. *Handbook on Human Factors and Ergonomics Methods*. 46-1-46-6. Boca Raton, FL: CLC Press, LLC

Eisenbach, A. & Eisenbach, M. (2006). phpSyntaxTree tool, http://ironcreek.net/phpsyntaxtree.

Emond, B. (2006). WN-LEXICAL: An ACT-R module built from the WordNet lexical database. In *Proceedings of the Seventh International Conference on Cognitive Modeling*, Trieste, Italy.

Ericsson, S. (2004). Dynamic optimisation of information enrichment in dialogue. In *Proceedings of the 8th International Workshop on Formal Semantics and Pragmatics of Dialogue*, Catalog, Barcelona, Spain.

Fauconnier, G. (1985). *Mental Spaces*. Cambridge, MA: The MIT Press.

FrameNet. Downloaded from: http://www.icsi.berkeley.edu/~framenet/

Gaerdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: The MIT Press.

Gluck, K. A., Ball, J. T., & Krusmark, M. A. (2007). Cognitive control in a computational model of the Predator UAV operator. *Integrated Models of Cognitive Systems* pp. 13-28. Edited by W. Gray. NY: Oxford University Press.

Gunzelmann, G. & Lyon, D. R. (in press). Mechanisms of human spatial competence. In T. Barkowsky, C. Freksa, M. Knauff, B. Kried-Bruckner, and B. Nevel (Eds.) *Lecture notes in artificial intelligence: Proceedings of Spatial Cognition (V) 2006*. Berlin, Germany: Springer-Verlag.

Heiberg, A., Harris, J. & Ball, J. (2007). Dynamic Visualization of ACT-R Declarative Memory Structure. In *Proceedings of the 8th International Conference on Cognitive Modeling*.

Johnson-Laird, P. N. (1983). *Mental Models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kamp, H. and Reyle, U. (1993). *From discourse to logic: Introduction to the model theoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer Academic Publishers.

Kilgariff, A (1997). I Don't Believe in Word Senses. In *Computers and the Humanities* 32(2), pp. 91-113.

Kintsch, W. (1998). *Comprehension, a Paradigm for Cognition*. New York: Cambridge University Press.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Levin, E., Sharifi, M. & Ball, J. (2006). Evaluation of Utility of LSA for Word Sense Discrimination. In *Proceedings of the NAACL-HLT Conference.*

Mavridis, N. & Roy, D. (2006) Grounded Situation Models for Robots: Where words and percepts meet. Downloaded from CiteSeer: http://citeseer.ist.psu.edu/mavridis06grounded.html

Mikrokosmos. Downloaded from: http://crl.nmsu.edu/Research/Projects/mikro/htmls/oldstuff-htmls/training-intro.html

Prince, A. & Smolensky, P. (1993). Optimality Theory: Constraint interaction in Generative Grammar. Rutgers University and University of Colorado, Boulder, manuscript.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K. & Sedivy, J. (1995). Integration of visual and linguistic information in spoken-language comprehension. *Science*, 268, 1632-1634.

Townsend, D. & Bever, T. (2001). *Sentence Comprehension, The Integration of Habits and Rules*. Cambridge, MA: The MIT Press.

Zwann, R., Langston, M. & Graesser, A. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, pp. 292-297.

Zwann, R., and Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Van Dijk, T. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.

Varges, S. (2006). Overgeneration and ranking for spoken dialogue systems. In *Proceedings of the Fourth International Natural Language Generation Conference*, 20–22, Sydney, Australia, July 2006. Association for Computational Linguistics.

VerbNet. Downloaded from http://verbs.colorado.edu/~mpalmer/projects/verbnet.html

Walker, M. A., S. J. Whittaker, A. Stent, P. Maloor, J. Moore, M. Johnston, & G. Vasireddy. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science* 28:5, 811-840.

WordNet. Downloaded from: http://wordnet.princeton.edu/